Identifying Noise Shocks*

Luca Benati
University of Bern†

Joshua Chan
Purdue University‡

Eric Eisenstat
University of Queensland§

Gary Koop
University of Strathclyde¶

Abstract

We study identifying restrictions that allow news and noise shocks to be recovered empirically within a Bayesian structural VARMA framework. In population, the identification scheme we consider exactly recovers news and noise shocks. Monte Carlo evidence further demonstrates its excellent performance, as it recovers the key features of the postulated data-generation process—the real-business cycle model of Barsky and Sims (2011) augmented with noise shocks about future total factor productivity—with great precision. We provide several empirical applications of our identification scheme. Evidence uniformly support the conclusion that noise shocks play a minor role in macroeconomic fluctuations.

*We wish to thank Harris Dellas, Carlo Favero, Francois Gourio, Luca Sala and Eric Sims for useful discussions and/or suggestions. The usual disclaimers apply.

†Department of Economics, University of Bern, Schanzeneckstrasse 1, CH-3001, Bern, Switzerland. Email: luca.benati@vwi.unibe.ch

‡Department of Economics, 100 Grant St, West Lafayette, IN 47907, USA. Email: joshuacc.chan@gmail.com

§School of Economics, University of Queensland, Brisbane St Lucia, QLD 4072, Australia. Email: e.eisenstat@uq.edu.au

¶Department of Economics, University of Strathclyde, 199 Cathedral Street, Glasgow G4 0QU, United Kingdom. Email: gary.koop@strath.ac.uk

1 Introduction

Noise shocks—defined as announcements about future fundamentals which, in fact, never materialize—offer the possibility of generating macroeconomic fluctuations without any variation in the economy's fundamentals. They are therefore radically different from news shocks, i.e. announcements about future fundamentals which do indeed materialize at some future date.

In this paper we investigate theory-driven restrictions which motivate an identification scheme we use in order to recover the impact of news and noise shocks from the data. The restrictions we propose extend those employed by Barsky and Sims (2011) and Forni *et al.* (2017) in ways which are both justified from a theoretical standpoint, and effective at recovering the two shocks. In particular, the proposed identification scheme exactly recovers news and noise shocks in population.

Although our identification scheme cannot be imposed within a structural Vector Autoregressive (SVAR) framework, it can indeed be imposed using structural Vector Autoregressive Moving Average (SVARMA) models. We implement it empirically within a Bayesian framework, and we demonstrate, in a Monte Carlo study, its excellent performance. We then use our identification scheme in an empirical application investigating the role of news and noise shocks to Total Factor Productivity (TFP). We compute impulse response functions (IRFs) to news and noise shocks, and forecast error variance decompositions (FEVs). We find that noise shocks play a minor role in macroeconomic fluctuations.

Our identification strategy is based on the general principle that agents' inability to distinguish news and noise shocks on impact—which is the essence of the entire 'news *versus* noise' problem—implies that

1. the responses of the economy to the two shocks at t=0 will be the same, whereas

2. they will progressively diverge further out.

As time goes by, all else held constant, agents gradually learn whether the announcement was news (i.e. that the fundamentals will indeed change) or simply noise, leaving fundamentals unchanged.

To provide some intuition, consider an announcement that purports to promise a future positive permanent shock to TFP. Is this announcement news or noise? Initially, it is impossible to tell. As we illustrate in Section 3 using the real business cycle (RBC) model of Barsky and Sims (2011) augmented with noise shocks about TFP, such an announcement indeed produces, on impact, identical impulse vectors regardless of whether it is news or noise.

The *crux* of our identification strategy is that, although news and noise shocks are observationally equivalent on impact, they are not at all subsequent horizons. The news shock will ultimately cause a permanent increase in TFP, whereas the noise shock will not. This implies that there is no way that the two shocks might be

confused once the entire set of their properties is taken into account. Quite simply, within the present example, a permanent shock cannot be observationally equivalent to a transitory one. Working with an RBC model with noise shocks about TFP, we then provide a straightforward illustration of how the IRFs to news and noise shocks, and the fractions of FEV of the variables they explain, can be exactly recovered in population by imposing the model-implied restrictions that (*i*) news and noise shocks produce identical impulse vectors on impact, and (*ii*) news and non-news shocks are the only disturbances having a permanent impact on TFP.

We employ an econometric methodology based on Bayesian estimation of structural VARMA models, developed in Chan *et al.* (2016, 2019).[1] The need for using VARMAs instead of VARs originates from the fact that the 'news *versus* noise' problem automatically produces a reduced-rank structure for the matrix of the shocks' impact responses at t=0.[2] The details of this and related methodological issues are discussed in Chan *et al.* (2019). Building on this work, we implement our identification scheme within the SVARMA methodology. In our Monte Carlo study, we find our algorithm to work extremely well.

We consider an empirical application where the 'news *versus* noise' problem pertains to TFP. We find that noise shocks play a minor role in macroeconomic fluctuations. Our results for the system featuring TFP contrast with those of both Blanchard et al. (2013) and Forni et al. (2017), who found a significant role for noise shocks pertaining to productivity. Likewise, results for the system featuring dividends and stock prices contrast with the corresponding findings of Forni et al. (2017).

The paper is organized as follows. The next section discusses several theoretical implications of the 'news *versus* noise' problem. In Section 3 we illustrate the main features of the RBC model of Barsky and Sims (2011) augmented with noise shocks about TFP. We show, working in population, that low-order SVARMAs can approximate very well the model's theoretical IRFs and fractions of FEV explained by individual shocks. We also discuss our identifying restrictions. In Section 4 we then show, working in population, that our identifying restrictions allow to exactly recover the RBC model's IRFs and FEVs. Section 5 briefly outlines our econometric methodology, and discusses the Monte Carlo evidence about its performance. Section 6 contains our empirical application. Section 7 concludes, and discusses possible directions for future research.

[1] The general point that SVARMAs can be used to estimate (nonfundamental) semi-structural models goes back to Hansen and Sargent (1981) and Ito and Quah (1989).

[2] Intuitively, this has to do with the fact that the very essence of this problem causes the impulse vectors to news and noise shocks at t=0 to be identical. In turn, this automatically implies that, with two identical columns, the matrix of the shocks' structural impacts at t=0 cannot have full rank.

2 Theoretical Implications of the 'News *versus* Noise' Problem

In this section we discuss several theoretical implications of the news *versus* noise problem and illustrate them with reference to a standard present-value model for dividends and stock prices. Three implications which play a crucial role in our identification strategy have already been mentioned:

(1) On impact (i.e., at t=0) news and noise shocks generate identical IRFs.

(2) After impact (i.e., for all t>0) IRFs to news and noise shocks progressively diverge.

(3) The impact matrix of the structural shocks at t=0 has reduced rank.

Two further theoretical implications of the news versus noise problem are the following:

(4) The structural model possesses a non-fundamental representation.

(5) The absolute magnitude of the economy's response to noise shocks at t=0 is monotonically decreasing in their standard deviation.

As already discussed, (1) is a direct logical implication of agents' inability to distinguish news and noise shocks on impact—which is the essence of the news *versus* noise problem—whereas (2) simply originates from the fact that, as time goes by, the true nature of the shock is progressively revealed to agents. In particular, in the RBC model of Section 3, the news shock ultimately causes a permanent increase in TFP, whereas the noise shock does not cause any permanent change in any variable. One small qualification to (1) is that exact equality between the IRFs to news and noise shocks at t=0 holds when the standard deviations of the two shocks are identical. In the general case, the two impulse vectors at t=0 will be proportional, with the coefficient of proportionality being equal to the ratio between the two shocks' standard deviations. Our identification strategy will therefore impose proportionality, rather than strict equality, between the two impulse vectors at t=0. This proportionality implies that the impact matrix of the structural shocks at t=0 will be reduced rank.[3] This will automatically imply (4), i.e., that the model possesses a non-fundamental representation.

[3] To be absolutely clear, proportionality of IRFs (and therefore the corresponding matrix of impacts being reduced-rank) only holds on impact. This is a direct logical implication of the fact that, as time goes by, the true nature of the shock is progressively revealed to agents, so that, at all periods after impact, the economy's responses to news and noise shocks will necessarily diverge. A subtle point to stress here is that whereas this holds for the response of the *economy* to news and noise shocks, it does not necessarily hold for other definitions of IRFs (we wish to thank a referee for pointing this out). For example, *on impact* news and noise shocks induce, by their very nature, an identical displacement of agents expectations about the endogenous variables, thus implying that, conditional on information at t=0, *agents*' IRFs to the two shocks are identical. Crucially, however, what matters for our identification strategy is the response of the *economy* to the two shocks—which encodes information about their nature—rather than agents' IRFs conditional on information at t=0.

The reason for the fifth implication is straightforward. As the volatility of noise shocks increases, the signal becomes less and less informative, and agents therefore react to it less and less. In the limit, as the volatility of noise shocks tends to infinity, so that the signal becomes completely uninformative, agents' reaction to it—and therefore to noise shocks—tends to zero. This implies that the absolute magnitude of the IRFs to noise shocks at t=0 is monotonically decreasing in their standard deviation.[4] With the magnitude of the economy's response decreasing as the volatility of noise shocks increases, an obvious question is whether the importance of these shocks at driving macroeconomic fluctuations (in terms of the fractions of FEV of the variables they explain) is increasing or decreasing with their volatility. Although it is not possible to make general statements on this issue, in the model for dividends and stock prices of the next sub-section, the relationship is hump-shaped. Intuitively, when the standard deviation of noise shocks is small, the increase in the volatility of noise shocks dominates the decrease in agents' response to them. As this standard deviation increases, there comes a point where the latter effect becomes the dominant one, and the importance of noise shocks in driving stock prices decreases.

2.1 A simple illustration

2.1.1 A present-value model for dividends and stock prices

We illustrate the news *versus* noise problem with reference to a standard present-value model for dividends and stock prices along the lines of Forni et al. (2017), in which stock prices are equal to the present discounted value of future expected dividends.[5] Labelling stock prices and log dividends as S_t and d_t, respectively, and using notation where $t+j|t$ subscripts on any variable denote expectations at time t of the variable at time $t+j$, we have:

$$\ln S_t = \sum_{j=0}^{\infty} \beta^j d_{t+j|t} \tag{1}$$

where β is the discount rate. Log dividends are assumed be the sum of two unobserved components, a permanent and a transitory one,

$$d_t = d_t^P + d_t^T \tag{2}$$

which evolve according to

$$d_t^P = d_{t-1}^P + \epsilon_t^{NN} + \epsilon_{t-1}^{NE} \tag{3}$$

[4] It is clear that the magnitude of the impact of noise shocks on the economy at t=0 will be zero if their standard deviation is zero (in which case there are no noise shocks to speak of). This implies that the magnitude of the impact of noise shocks on the economy at t=0, as a function of the shocks' standard deviation, exhibits a discontinuity at zero.

[5] Online Appendix A.2 presents an additional illustration based a standard New Keynesian model in which the news versus noise problem pertains to the natural rate of interest.

$$d_t^T = \rho_T d_t^T + v_t \tag{4}$$

where ϵ_t^{NN} and ϵ_t^{NE} are the non-news and news shocks, respectively, with $\epsilon_t^{NN} \sim N(0, \sigma_{NN}^2)$ and $\epsilon_t^{NE} \sim N(0, \sigma_{NE}^2)$; $0 < \rho_T < 1$; and $v_t \sim N(0, \sigma_v^2)$.

Specification (2)-(4) is the same as the one used by Blanchard et al. (2013), with the crucial difference that, in their case, the equation for the permanent component does not feature a 'proper' news shock—defined as a time-t disturbance which impacts upon the relevant variable only at a future date—as it is uniquely driven by a standard 'surprise' disturbance which enters contemporaneously. As we shall discuss below, this is the key reason why, on impact, their IRFs to the noise and surprise shocks are not the same.

Although at time t agents observe d_t, its two individual components, d_t^P and d_t^T, are never observed. Agents however observe a signal, s_t, which reveals some information about the news shock:

$$s_t = \epsilon_t^{NE} + u_t \tag{5}$$

with $u_t \sim N(0, \sigma_u^2)$. Online Appendix A.1 characterizes and solves the agent's signal extraction problem.[6] Based on (1), (2), and (A.5) the solution for log stock prices is given by

$$\ln S_t = \frac{d_{t|t}^P}{1-\beta} + \frac{d_{t|t}^T}{1-\rho_T\beta} + \frac{\beta}{1-\beta}\epsilon_{t|t}^{NE} = \frac{d_{t|t}^P}{1-\beta} + \frac{d_{t|t}^T}{1-\rho_T\beta} + \frac{\beta K_{32}}{1-\beta}\left[\epsilon_t^{NE} + u_t\right], \tag{6}$$

where K_{32} is defined in Online Appendix A.1.

2.1.2 Illustrating the theoretical implications of the 'news *versus* noise' problem

Since, on impact, neither news nor noise shocks affects either $d_{t|t}^P$ or $d_{t|t}^T$ it follows that:

$$\left[\frac{\partial \ln S_t}{\partial \epsilon_t^{NE}}\right]_{t=0} = \left[\frac{\partial \ln S_t}{\partial u_t}\right]_{t=0} = \frac{\beta K_{32}}{1-\beta} \tag{7}$$

In words, since on impact agents are unable to distinguish between news and noise shocks, they will react to either disturbance in the same way, and stock prices will jump by exactly the same amount—i.e., implication (1) above. This feature was also noticed by Forni et al. (2017), who pointed out (see their Section 2.2) that *'[...] the impact responses are identical, since the agents cannot distinguish between the two shocks immediately.'* With this in mind, Forni et al. (2017), develop an identification strategy based on dynamic rotations of the VAR's residuals, which results in a (non-fundamental) structural VARMA representation.[7]

[6] The Online Appendix can be found at Luca Benati's webpage, at: https://sites.google.com/site/lucabenatiswebpage.

[7] Their implied SVARMA obtained this way, however, entails a number of additional restrictions that may not be desirable; we avoid this by estimating the more general VARMA models directly (see Chan et al., 2019, for an in-depth discussion).

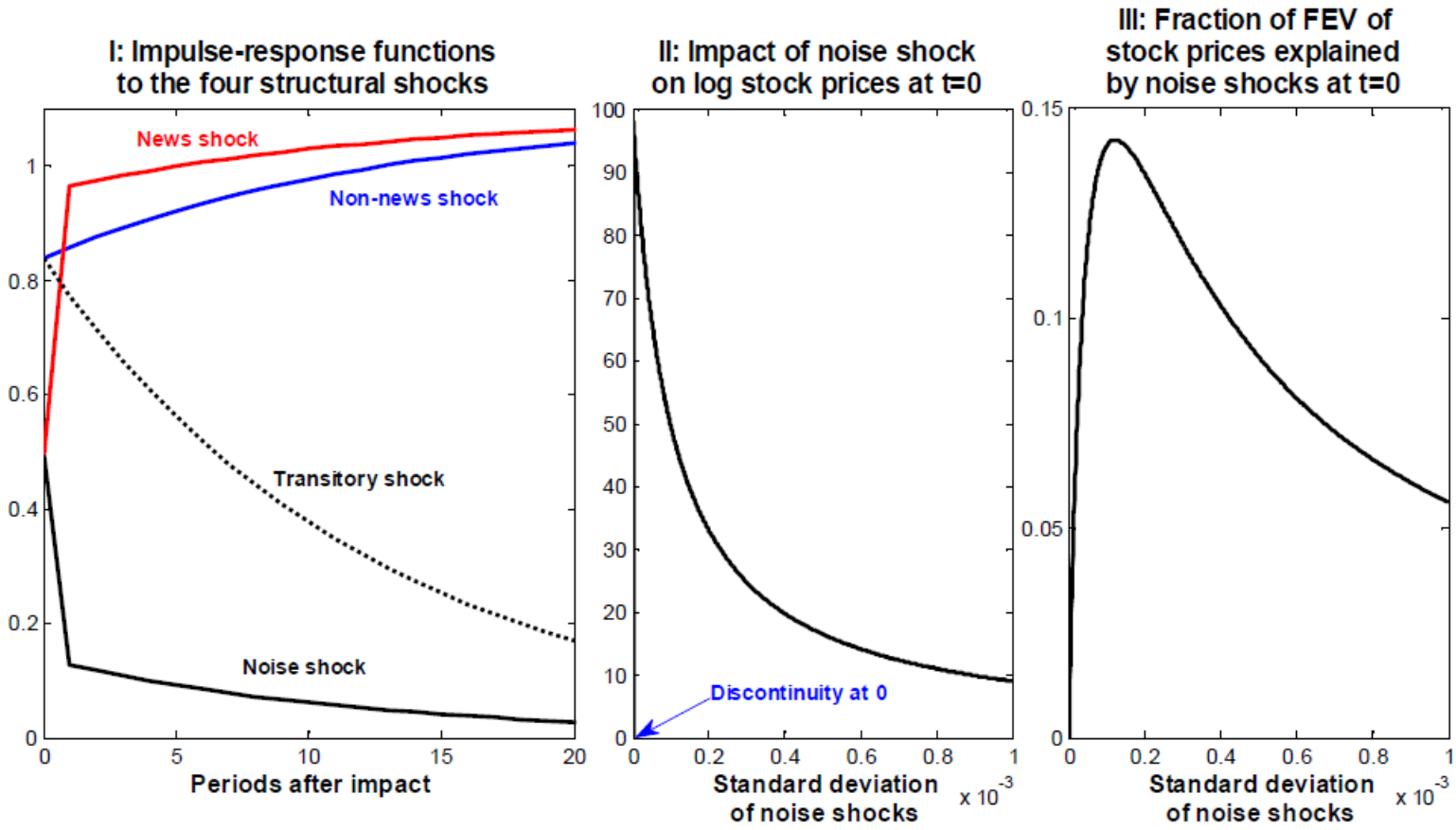

Figure 1 Impulse-response functions of log stock prices to the structural shocks, and fractions of the FEV of stock prices explained by noise shocks, as a function of their standard deviation

The IRFs of Blanchard et al. (2013), on the other hand, do not exhibit this property. The reason is that in their case the process for the permanent component of technology does not feature a news shock. Their analogue to our d_t^P is uniquely driven by a standard surprise disturbance. This implies that, on impact, there cannot be any symmetry between this disturbance and the noise shock. This is because whereas, at t=0, the noise shock only impacts upon the economy via the signal, the surprise shock affects both the signal and (via the permanent component) the relevant variable itself. As a result, the impact upon the economy at t=0 of the two shocks considered by Blanchard et al. (2013) cannot possibly be the same. Noise shocks and news shocks, on the other hand, are perfectly symmetrical at t=0, as they impact upon the economy uniquely via the signal extraction problem, within which they cannot be distinguished. As a result, their impact on the endogenous variables at t=0 must necessarily be identical.

The first panel of Figure 1 plots IRFs to unitary innovations in the four shocks, by setting β=0.99, ρ_T=0.9, σ_{NN}=σ_{NE}=σ_u=σ_v=0.01. As implied by (7), the impact responses to news and noise shocks are identical.[8] As time goes by, agents observe dividends, and therefore learn whether the signal at t=0 was news or noise, and the two IRFs therefore progressively diverge: the IRF to news shocks converges to the new permanent level of stock prices, whereas that to noise shocks converges to zero.

Another point to stress is that, at all horizons beyond t=0, the magnitude of the IRF associated with noise shocks is smaller than the one of the IRF associated with news shocks. The reason is that agents are gradually learning about the nature of the shock. At t=0 they cannot distinguish the two shocks (and so they react in exactly the same way), whereas at t=∞ they can perfectly disentangle them (and their reaction is therefore positive to the news, and zero to the noise). At all other horizons their reaction is intermediate between these two extremes. As we will see in Section 3.2 based on the RBC model of Barsky and Sims (2011) augmented with noise shocks about TFP, this feature is a remarkably robust one.

As for implication (3), with four shocks, and just three observed variables (dividends, stock prices, and the signal), the present model does not allow to properly illustrate the reduced-rank structure of the matrix of the shocks' structural impacts at t=0. For this, we will have to wait for the RBC model of Section 3. As mentioned, non-fundamentalness—implication (4)—is a direct implication of the reduced-rank structure of the matrix of structural impacts at t=0.

With regards to implication (5), the second panel of Figure 1 shows the impact response at t=0 of stock prices to noise shocks as a function of their standard deviation. Apart from the discontinuity at zero, the impact response is monotonically decreasing, reflecting the fact that, as the signal becomes noisier and noisier, agents react to it less and less. Finally, the last panel of Figure 1 illustrates how, in this

[8] Notice that IRFs in Figure 1 are to shocks of size one, so that the impact responses to news and noise shocks are identical for either variable. As mentioned, in general the impacts are proportional, with the factor of proportionality being equal to the ratio between σ_u and σ_{NE}.

model and for these parameter values, the fraction of the FEV of stock prices at t=0 explained by noise shocks, as a function of their standard deviation, exhibits a 'Laffer curve-type' shape. Over the initial portion of the parameter space, the increase in the volatility of noise shocks dominates the decrease in agents' response to them. Beyond a certain point, however, the latter effect becomes the dominant one, and the importance of noise shocks at driving stock prices decreases more and more.

We now turn to Barsky and Sims' (2011) RBC model, which we augment with noise shocks about future TFP.

3 The RBC Model of Barsky and Sims (2011) Augmented With Noise Shocks About TFP

The preceding section examined the properties of news and noise shocks within a simple theoretical model. In our empirical work we will use SVARMA methods, which are atheoretical except for imposing identifying restrictions motivated by theory. In this section we motivate the use of SVARMA methods based on the RBC model of Barsky and Sims (2011), which we augment with noise shocks about TFP. In the next section, working with the theoretical MA representation of the model, we will show that our identifying restrictions can exactly recover news and noise shocks.

The basic elements of the RBC model of Barsky and Sims (2011) can be succinctly described as follows (for details, see Barsky and Sims, 2011, Section 2.2.1). Consumers maximize the discounted sum, with discount factor $0 < \beta < 1$, of the period-t utility flows $U_t \equiv \ln(C_t - bC_{t-1}) - E_t^N(1 + 1/\eta)^{-1} N_t^{1+1/\eta}$, for $t = 0, 1, 2, 3, ..., \infty$. C_t is consumption, N_t is hours worked, E_t^N is a random process whose logarithm, ϵ_t^n, is $N(0, \sigma_n^2)$, $0 < b < 1$ is a parameter capturing the strength of habit formation and η captures the curvature of the labour supply function. Output, Y_t, is the sum of consumption, investment (I_t), and public expenditure (G_t) and it is produced *via* the production function $Y_t = A_t K_t^{\theta} N_t^{1-\theta}$, where K_t is the capital stock, and θ is the Cobb-Douglas parameter. The capital stock evolves according to the law of motion $K_{t+1} = K_t(1 - \delta) + I_t[1 - (\gamma/2)(I_t/I_{t-1} - \tilde{g}_I)^2]$, where δ is the depreciation rate, γ is a parameter capturing the magnitude of capital adjustment costs, and $\tilde{g}_I$ is the gross rate of growth of investment in the steady-state. Finally, public expenditure is postulated to be equal to a stationary random fraction of GDP, that is, $G_t = g_t Y_t$, with $\ln g_t = \ln \bar{g} + \epsilon_t^g$, with ϵ_t^g being $N(0, \sigma_g^2)$.

Productivity is captured by A_t, and we assume that $a_t = \ln(A_t)$ evolves according to

$$a_t = \tilde{a}_t + v_t \tag{8}$$

where $\tilde{a}_t$ is the unobserved permanent component of productivity, and v_t is a white noise disturbance, $v_t \sim N(0, \sigma_v^2)$. The 'news *versus* noise' problem pertains to the productivity process. In particular, we assume that the permanent component of a_t

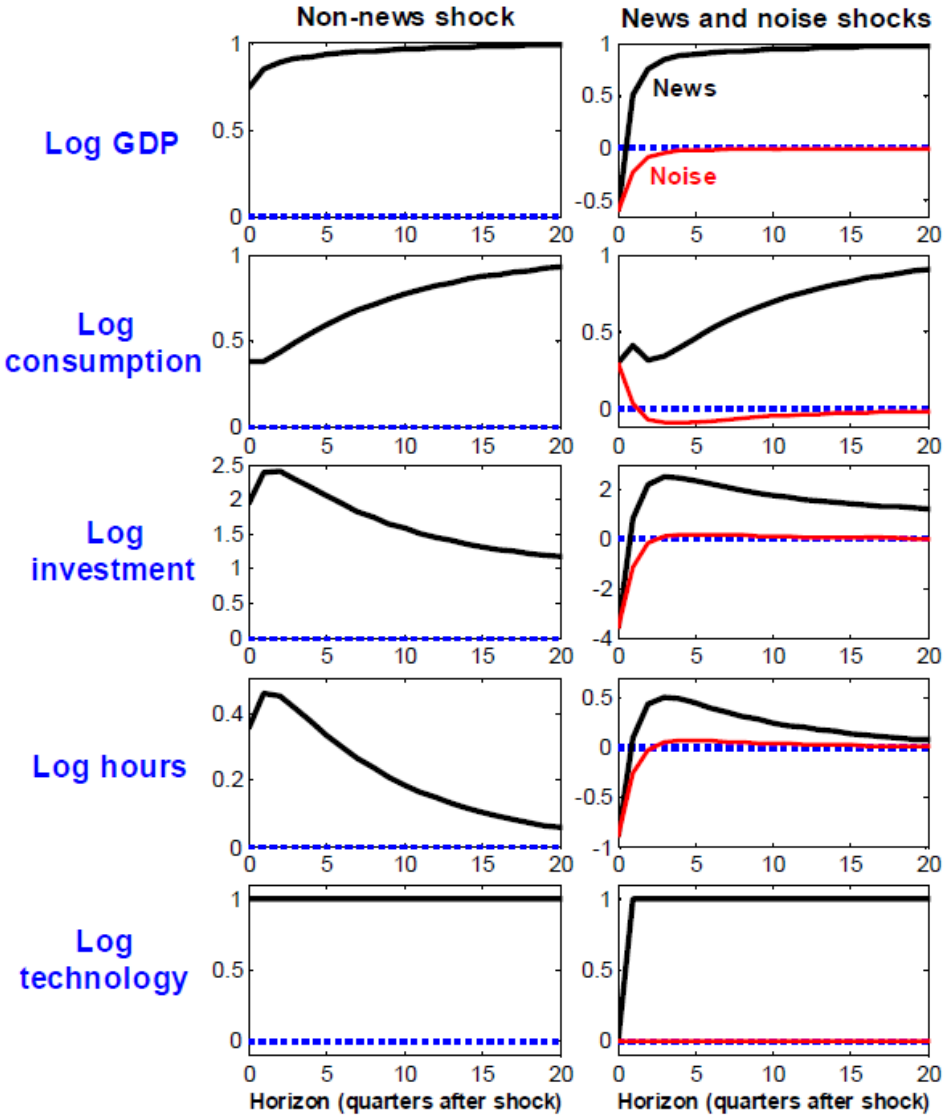

Figure 2 Impulse-response functions to non-news, news and noise shocks for Barsky and Sims' (2011) model augmented with noise shocks about TFP

evolves according to

$$\tilde{a}_t = \tilde{a}_{t-1} + \epsilon_t^{NN} + \epsilon_{t-\tau}^{NE} \tag{9}$$

where, once again, ϵ_t^{NN} and ϵ_t^{NE} are a non-news and a news shock, respectively, and τ is the anticipation horizon for the news shock. Although at time t agents observe a_t, its two individual components, $\tilde{a}_t$ and v_t, are never observed. In each period, however, agents receive a signal, which is equal to the sum of the news shock and of a noise component as in (5). Details of the agent's signal extraction problem, together with the model's solution, are given in Online Appendix C.

Figure 2 shows the IRFs of the logarithms of GDP, consumption, investment, hours and technology to one-standard deviation non-news, news and noise shocks. The model is calibrated using parameter values similar to those in Barsky and Sims (2011) (see Online Appendix C). As expected, the IRFs to news shocks are qualitatively in line with those reported in Figure 1 of Barsky and Sims (2011), with consumption increasing on impact; GDP, investment and hours falling on impact; and GDP, consumption, and investment subsequently converging to their new, higher, steady-state values. Once again, for either variable the responses to news and noise shocks at t=0 are identical. The IRFs of GDP, consumption, and investment to noise shocks fade away to zero quite rapidly, whereas those to news shocks converge to the new steady-state values. As for hours, the response to noise shocks quickly fades away to zero, whereas that to news shocks, although it ultimately also converges to zero, exhibits a strong hump-shaped pattern as in Barsky and Sims (2011). Finally, in response to non-news shocks GDP, consumption, investment and hours all increase on impact.

Figure I.1 in the Online Appendix shows the fractions of FEV explained by non-news, news and noise shocks. Key points to stress are that (*i*) noise shocks explain almost nothing of the FEV of any variable at all horizons; (*ii*) non-news and news shocks explain, at long horizons, about 50% each of the FEV of log TFP, and they explain non-negligible-to-sizeable fractions of the FEV of GDP, consumption, and investment; and (*iii*) none of the three shocks explain an appreciable amount of the FEV of hours.

3.1 Reduced-rank structure of the matrix of structural impacts at t=0

It can be seen immediately that the matrix of structural impacts at t=0 has a reduced rank. In particular, with two signal-extraction problems (news versus noise, and non-news versus transitory) there are two pairs of columns (i.e., of impulse vectors at t=0) which are proportional to each other: the column generated by noise is proportional to the one generated by news, and the same holds true for the two impulse vectors

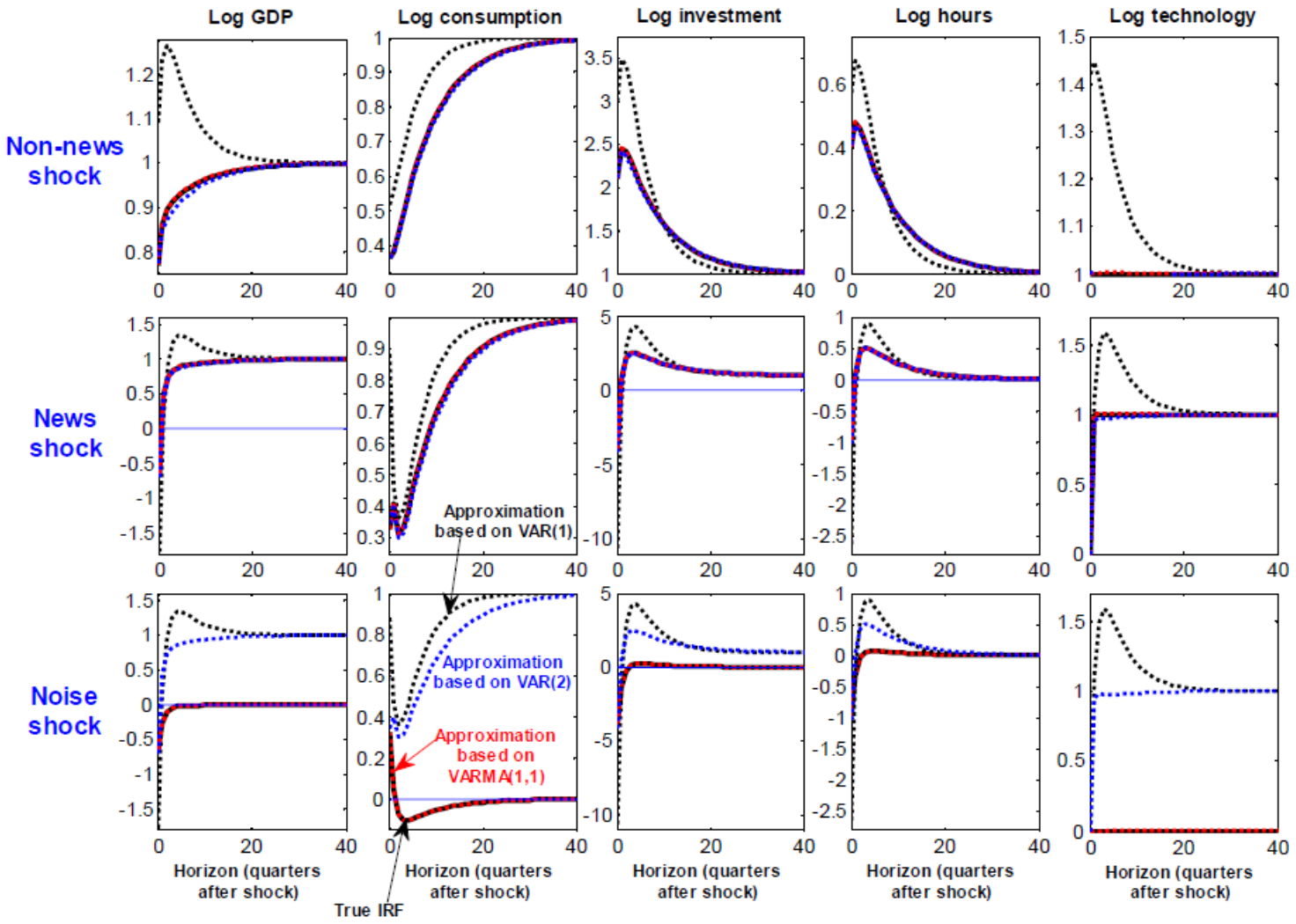

Figure 3 Approximating the IRFs of Barsky and Sims' (2011) RBC model augmented with noise shocks by means of theoretical SVAR(MA) representations

generated by non-news and transitory.[9] This implies that, within the present context, the matrix of the structural shocks' impacts at t=0 has rank 3. We now turn to the issue of whether low-order SVARMAs can well approximate the main features of the model.

3.2 Approximating the theoretical model based on low-order SVARMA(p,q)'s

Figure 3 compares the theoretical IRFs to non-news, news, and noise shocks generated by the model (i.e., the very same IRFs shown in Figure 2) to those produced by either its truncated theoretical SVAR(1) and SVAR(2) representations, or its truncated theoretical SVARMA(1,1) representation.[10] Since, as can be trivially shown, the model possesses a near-exact VARMA(1,1) representation,[11] a truncated VARMA(1,1) recovers the true IRFs almost perfectly.[12] Interestingly, although VARs can never recover, even approximately, the IRFs to noise shocks,[13] a VAR(2) can approximate remarkably well the IRFs to both non-news and news shocks.

Although these results have been obtained based on a fairly simple DSGE model, at the very least they suggests that, in practice, structural VARMA models with a small MA lag order might be sufficient to accurately approximate the main features of the underlying DGP.

3.3 Identifying restrictions

The preceding discussion motivates using the following restrictions in order to identify non-news, news, and noise shocks (in Section 4.2 we will show that these restrictions indeed allow to exactly recover the three shocks in population):

R1 At t=0, the variable of interest—to fix ideas, TFP—is impacted upon by only two disturbances, the non-news shock, and a transitory shock which is disentangled from the non-news shock because, among the two shocks, it explains

[9] Within the context of the present-value model for dividends and stock prices of Section 2.1 this was illustrated in the first panel of Figure 1.

[10] The truncated theoretical SVAR and SVARMA representations of the RBC model have been computed via linear projections methods starting from the RBC model's structural MA representation, which can immediately be recovered from the model's IRFs (see Online Appendix D).

[11] In the sense that all MA matrices after the first one have very small entries.

[12] Qualitatively the same evidence holds for the fractions of FEV: we do not report this evidence for reasons of space, but it is available upon request.

[13] The problem is strictly conceptual: within a VAR framework, there is no way to break the one-to-one mapping at all horizons between the IRFs to news and noise shocks, originating from the fact that they share the very same impulse vector on impact. As a result, a VAR automatically forces the IRFs to news and noise shocks to be identical at all horizons. But since the two sets of IRFs are identical only on impact, and then diverge at longer horizons, it is logically impossible for a VAR to correctly recover the true IRFs. This implies that increasing the VAR's lag order cannot solve the problem.

the minimum fraction of the FEV of TFP at a specific long horizon (which we will take to be 20 years ahead). This restriction is more general than the corresponding restriction used by Barsky and Sims (2011), who identify the non-news shock as the reduced-form innovation in TFP (i.e. the only shock impacting upon TFP at t=0).

R2 The news shock is the one which, among all of the remaining shocks, explains the maximum fraction of the FEV of TFP at a long horizon (again, 20 years ahead).

In order to isolate the noise shock from the remaining disturbances, we impose the following restriction:

R3 The impulse vectors generated on impact by news and noise shocks are proportional to each other.

Although these restrictions are sufficient to eliminate observationally equivalent orthogonal rotations of the three shocks (hence allowing identification), in what follows we will also consider imposing additional restrictions implied by the RBC model, in order to get more precise estimates. We therefore conduct an exercise in the spirit of Canova and Paustian (2011) in order to derive a set of robust restrictions pertaining to news and noise shocks, where 'robust' means 'holding for an overwhelmingly large fraction of plausible random combinations of the model's parameters'.

We consider the following sets of plausible values for most of the model's structural parameters: $b \in [0, 0.99]$; $\delta \in [0.01, 0.08]$, $\eta \in [0.1, \infty)$, $\gamma \in [0.01, 0.1]$, $\bar{g} \in [0.15, 0.25]$, $\tilde{g}_A \in [0, 0.04]$, and we set the standard deviations of news and noise shocks to $\sigma_{NE} = \sigma_u = 1$.[14] The remaining parameters are calibrated as in Barsky and Sims (2011) as described in Online Appendix C. Following Canova and Paustian (2011), we take 100,000 draws for the parameters from Uniform distributions defined over these intervals. For each draw of the parameters, we solve the model and compute IRFs to news and noise shocks and the fractions of the FEV of either variable explained by the two shocks.

Table 1: Proportion of Drawn Parameters which Satisfy the Restriction: News Shock IRFs Greater than Noise Shock IRFs (in absolute value)

Periods after impact	TFP	GDP	Consumption	Investment	Hours
1	1.00	1.00	1.00	0.71	0.98
2	1.00	1.00	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00

[14] To be precise, we set the upper value of the interval for η to be $\frac{1}{\varepsilon}$ where ε is MATLAB's definition of a small number which is equal to 2.2204e-16.

Table 1 reports the proportion of drawn parameters which produce IRFs that satisfy one possible restriction of interest: i.e., that IRFs to news shocks are, in absolute value, greater than those to noise shocks.[15] These proportions turn out to be identical to the proportion of draws for which the percentage of the FEV explained by news shocks is greater than that explained by noise shocks. There is strong support for this restriction for all variables and at all horizons. For horizons of two or more this support becomes overwhelming. Accordingly, in our empirical work some of our specifications will impose the restriction that news shocks have larger IRFs (in absolute value) than noise shocks two periods after impact.[16]

4 Disentangling News and Noise Shocks In Population

In this section we provide a simple illustration of the ability of the proposed identification scheme R1-R3 to exactly recover news and noise shocks in population, based on the RBC model of Barsky and Sims (2011) augmented with noise shocks about future TFP. Specifically, we show that the IRFs to news and noise shocks, as well as the fractions of FEV of the variables they explain, can be exactly recovered by imposing the model-implied restrictions.

4.1 Recovering the IRFs and fractions of FEV in population

Let the theoretical structural MA representation of the RBC model's observables be

$$Y_t = A_0\epsilon_t + A_1\epsilon_{t-1} + A_2\epsilon_{t-2} + A_3\epsilon_{t-3} + \cdots, \tag{10}$$

where $Y_t = [y_t, c_t, i_t, h_t, a_t]'$ and y_t, c_t, i_t, h_t are the logarithms of GDP, consumption, investment, and hours. In addition, $\epsilon_t = [\epsilon_t^{NN}, v_t, \epsilon_t^{NE}, u_t, \epsilon_t^g]'$, where the notation is as before,[17] and $E[\epsilon_t\epsilon_t'] = I_N$, with I_N being the $(N \times N)$ identity matrix, so that any of the i-th columns (with $i = 1, 2, ..., N$) of the MA matrices A_0, A_1, A_2, $A_3, \ldots$ has been divided by the standard deviation of the i-th shock in ϵ_t. Representation (10) can be immediately recovered from the model's IRFs to the structural innovations.

Observationally equivalent *reduced-form* representations of (10) can be obtained by post-multiplying all of the MA matrices A_0, A_1, A_2, $A_3, \ldots$ by an orthogonal rotation matrix R, yielding

$$Y_t = \tilde{A}_0\tilde{\epsilon}_t + \tilde{A}_1\tilde{\epsilon}_{t-1} + \tilde{A}_2\tilde{\epsilon}_{t-2} + \tilde{A}_3\tilde{\epsilon}_{t-3} + \cdots, \tag{11}$$

[15] We do not report results for $t = 0$ because on impact the two sets of IRFs are identical.

[16] This statement assumes that news and noise shocks have the same standard deviation. A slight adjustment is required when they have different standard deviations which is described in section 5.

[17] Since we have five observables, in this section we set to 0 all of the shocks other than the five collected in the vector ϵ_t.

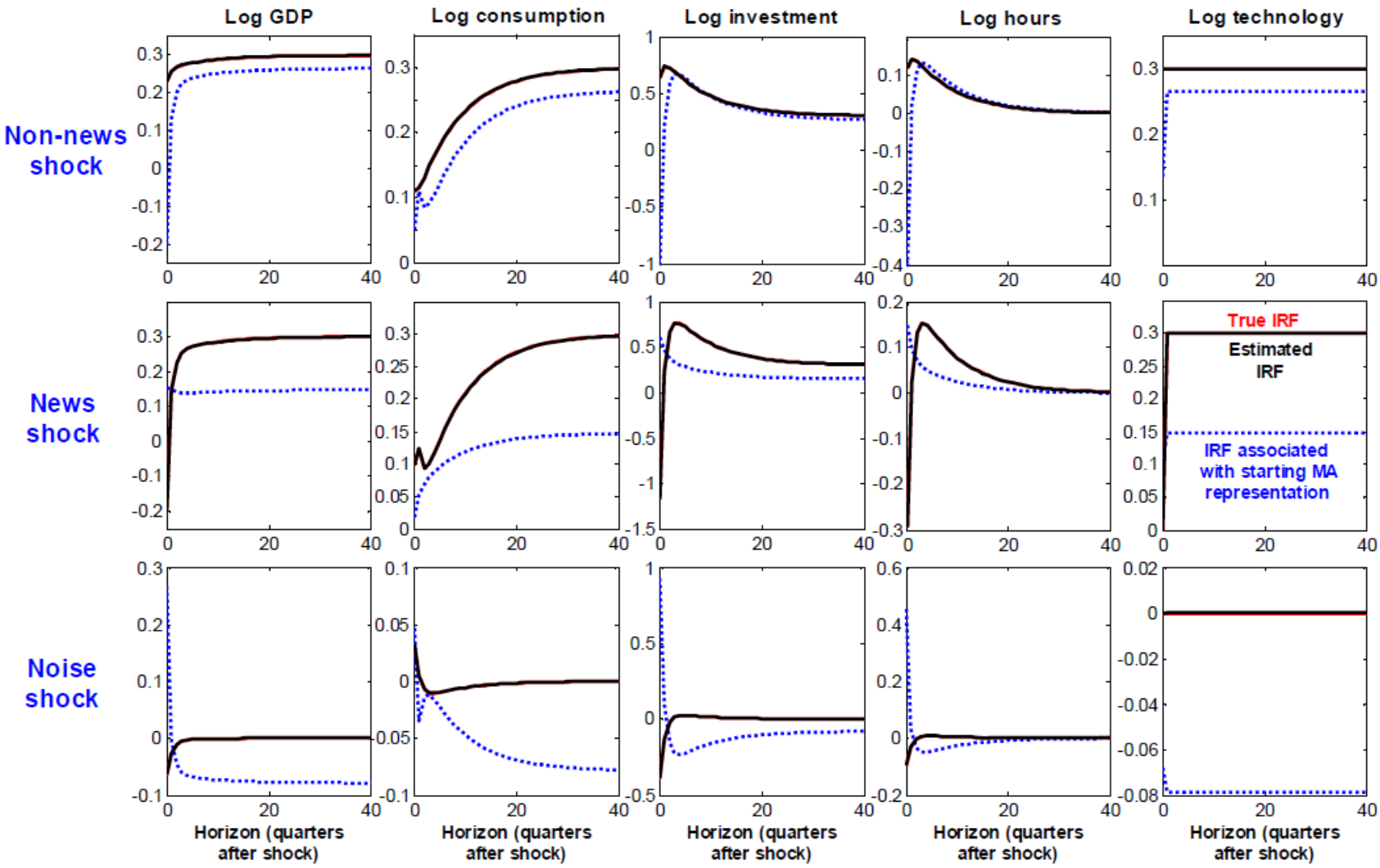

Figure 4 Assessing the identifying restrictions in population: Theoretical and estimated IRFs based on Barsky and Sims' (2011) RBC model augmented with noise shocks

where $\tilde{A}_j = A_j R$ and $\tilde{\epsilon}_{t-j} = R'\epsilon_{t-j}$, $j = 0, 1, 2, 3, \ldots$ We can randomly generate different rotation matrices[18], thus producing different observationally equivalent VMAs. The question we wish to address is whether imposing restrictions R1-R3 from subsection 3.3 on different reduced form VMAs allow us to recover the true structural VMA.

We have performed this exercise 100 times, and for all of these, imposing the restrictions R1-R3 produces IRFs and FEVs which match the true theoretical IRFs and FEVs.[19] We illustrate this in Figure 4, which contains results using one random rotation.[20] The figure shows, in red, the true IRFs as captured by the theoretical MA representation (10); in blue (dotted line) are the IRFs associated with the initial, randomly rotated MA representation (11); in black are the estimated IRFs, i.e. the IRFs we recovered starting from (11) by imposing identifying restrictions R1-R3.

In spite of the fact that the IRFs associated with the initial, randomly rotated MA representation were, in general, far away from the true IRFs, our identifying restrictions were able to recover them with great precision (in fact, the true and estimated IRFs are essentially indistinguishable). In light of the previous discussion this should not be seen as surprising. A crucial difference between news and noise shocks is that whereas the former has a permanent impact on TFP (and GDP, consumption, and investment), the latter does not have a permanent impact on any variable. It is therefore impossible for the two shocks to be observationally equivalent.

4.1.1 Approximating the MA representation with a truncated VARMA(p, q)

An issue arises when the econometrician tries to actually go to the data and obtain an estimate of the reduced form VMA representation of the preceding sub-section to disentangle news and noise shocks. A common practice in macroeconomics is to first estimate a reduced form VAR, then invert it to obtain a reduced form VMA representation. In models with noisy news, this is not possible as shown, e.g., in Blanchard

[18] We generate the $(N \times N)$ random rotation matrix as follows. We start by taking an $(N \times N)$ draw K from the $N(0, 1)$ distribution. Then, we take the QR decomposition of K, and we set the rotation matrix to Q'.

[19] We impose restrictions R1-R3 *via* rotation matrices as follows. Restriction R1's zero restrictions on impact are imposed *via* a Householder reflection. Restriction R2 is imposed *via* Uhlig's (2003, 2004) 'maximum FEV' approach. Proportionality between the impulse vectors generated on impact by news and noise shocks is imposed *via* numerical methods, based on the methodology proposed by Benati (2014, Section 3.2.2). In brief, the methodology is based on the notion of working with the entire set of possible rotation matrices associated with a given structural impact matrix, and optimizing the relevant criterion function over the space of the associated rotation angles (which we do *via* simulated annealing). Section 3.2.3 of Benati (2014) presents Monte Carlo evidence on the excellent performance of this methodology.

[20] Figure 4 presents results for the IRFs. Results for the FEVs are shown in Figure I.2 in the Online Appendix. In addition, we only report results for non-news, news, and noise shocks, but results for the other two shocks are available upon request. They are qualitatively the same to those presented here.

et al. (2013). However, as we shall show in this sub-section, low-order VARMAs can approximate with great precision the MA representation of the observables.

For example, for the RBC model with noise, a truncated[21] VARMA(1,1) already produces theoretical IRFs and fractions of FEV that are indistinguishable from those associated with the model's theoretical MA representation. In fact, as shown via extensive Monte Carlo simulations in Online Appendix E—see in particular Figures E.1-E.4—from an empirical standpoint an effective way of recovering the VMA representation of the observables is through the use of Bayesian VARMA(p,q) models.

To demonstrate these points, we repeat the exercise of the preceding sub-section where we randomly generated different rotation matrices so as to produce different reduced-form specifications. In this subsection, we use a VARMA(1,1) instead of the VMA used previously. To be precise, we use the truncated theoretical VARMA(1,1) representation of the model,

$$Y_t = BY_{t-1} + A_0\epsilon_t + A_1\epsilon_{t-1}, \tag{12}$$

and transform it using randomly generated orthogonal matrices, thus obtaining observationally equivalent VARMA(1,1) representations,

$$Y_t = BY_{t-1} + \tilde{A}_0\tilde{\epsilon}_t + \tilde{A}_1\tilde{\epsilon}_{t-1}. \tag{13}$$

We then impose the identifying restrictions of sub-section 3.3 and obtain IRFs and FEVs. We perform this exercise 100 times, and always find that we are able to recover the model's true IRFs and FEVs with great precision (results for all repetitions are available on request).

This is illustrated in Figure 5 for the IRFs (the comparable figure for the FEVs is in the Online Appendix) for one of the random transformations. The red line in the figure shows the true IRFs from RBC model (i.e. the same lines shown in Figure 4); the blue, dotted lines show the IRFs from the reduced form VARMA(1,1) produced by the random transformation; the black line shows the IRFs that result when we impose the identifying restrictions R1-R3. Again we find that imposing the identifying restrictions allows for the recovery of the model's true IRFs.

This suggests that an effective empirical strategy for estimating noisy-news model is to work with low order VARMAs and then impose the restrictions R1-R3, which identify the news and noise shocks. This is the strategy we will implement in Sections 5 and 6.

[21] As already mentioned, we compute truncated VARMA(p, q) approximations to the model's theoretical MA representation *via* linear projection arguments—see Online Appendix D.

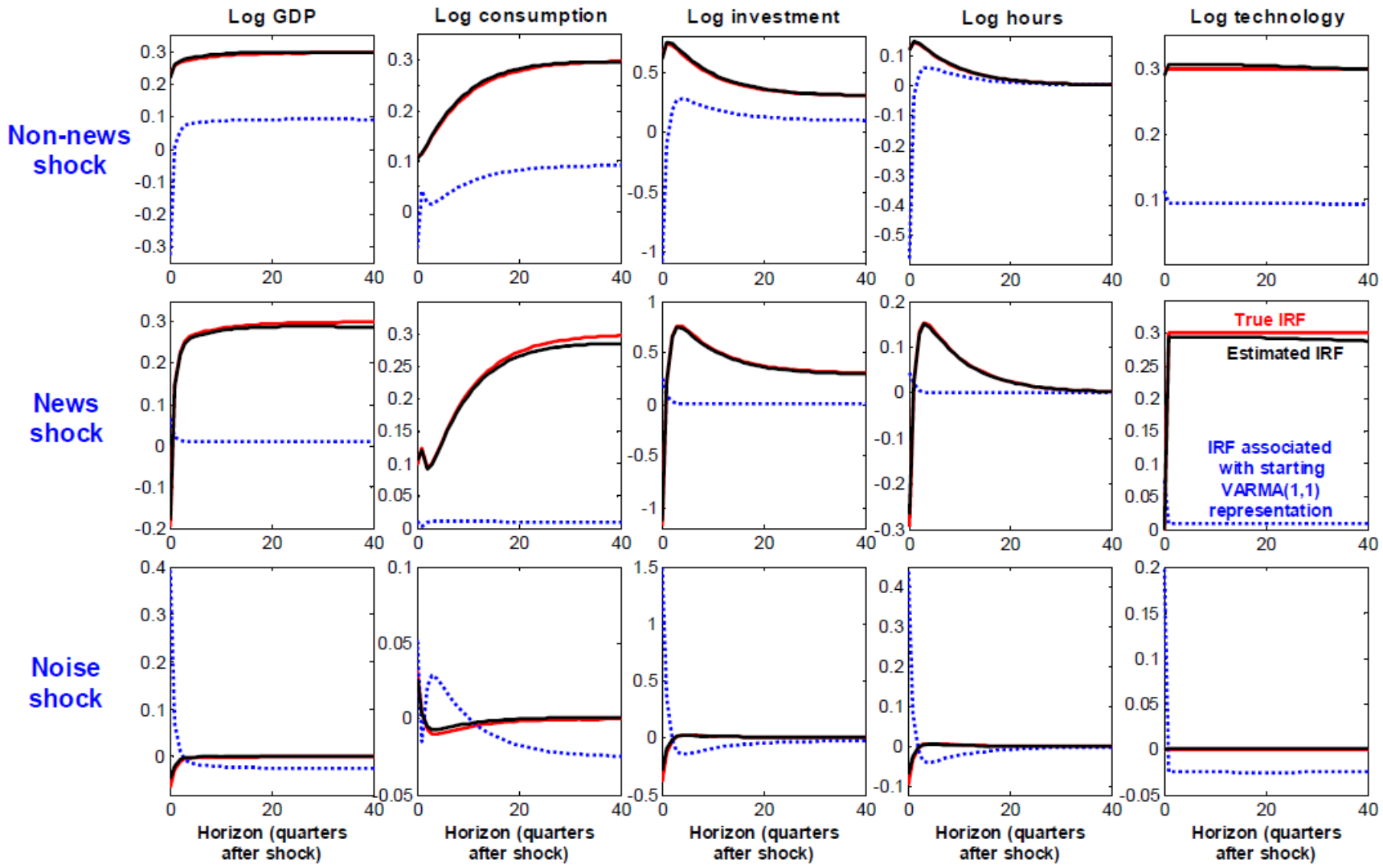

Figure 5 Recovering news and noise shocks in population: Theoretical IRFs, and estimated IRFs based on the theoretical truncated VARMA(1,1) representation of the RBC model augmented with noise shocks about TFP

5 Econometric Methodology

5.1 Discussion of identification

The presence of both news and noise shocks in our theoretical framework implies a non-fundamental VARMA representation. This means that in our specific case it is not possible to impose our key identifying restrictions using an SVAR. An SVARMA is required. To bring out the main ideas, we use an n-variate SVARMA(1,1)

$$Y_t - B_1 Y_{t-1} = u_t = A_0 \epsilon_t + A_1 \epsilon_{t-1}, \qquad \epsilon_t \sim N(0, \mathrm{I}_n). \tag{14}$$

Everything can easily extended to a SVARMA(p,q) with deterministic terms. A simple way to describe the need for a VARMA in our setting is as follows.

Note first that the key aspect of identification relates to the IRFs for t=0, which are in A_0 and that the error covariance matrix, Ω, is $\Omega = A_0 A_0' + A_1 A_1'$. In Sections 2 we derived results such as (7), and reinforced them in the more elaborate setting of Section 3, which showed that news and noise shocks have had an equal response on impact. In terms of the SVARMA, one might think this restriction can be imposed by setting two columns of A_0 to be equal to one another. As already mentioned, this is not correct, since it would imply that the news and noise shocks have the same standard deviations. The correct way of imposing the equal t=0 impacts of news and noise shocks on impact is to restrict two columns of A_0 to be proportional to one another with the factor of proportionality being equal to the ratio between the standard deviations of the two shocks.

If we impose the identification restriction that two columns of A_0 are proportional to one another, then A_0 will be a reduced-rank matrix. This precludes estimation in the SVAR case since, if $A_1 = 0$, Ω will be singular. This is why we need the SVARMA. To explain this another way, if we were to work with an SVAR and thus have $A_1 = 0$, the fact that B_1 will be common to impulse responses at all horizons[22] means that two shocks generating a proportional impulse vector at $t = 0$ will also produce proportional impulse responses at all horizons, with the proportionality factor being exactly the same for all variables and all horizons. This means that within a SVAR framework it is impossible to disentangle news and noise shocks. In Section 3, we demonstrated the theoretical consequences of this within an RBC model. These considerations imply that, unless a researcher is willing to resort to a DSGE-based approach, the SVARMA approach is the only option. Within the SVARMA framework, this problem does not arise since the presence of A_1 breaks the one-to-one mapping between the impulse responses at t=0 and responses over subsequent horizons.

Writing (14) in terms of polynomial matrices in the lag operator, $B(L)Y_t = A(L)\epsilon_t$, we assume following Chan *et al.* (2019) that $\det A(z) \neq 0$, almost everywhere for $z \in \mathbb{C}$. However, a reduced-rank A_0 implies that $\det A(0) = 0$, i.e. $\det A(z)$ has

[22] By this we mean that the IRF is based on the VMA representation, which involves inverting the VAR thus putting B_1 in every lag of the VMA representation.

a root at zero. This means that $A(L)$ is not invertible on the closed unit disk, or equivalently, (14) is non-fundamental. As discussed extensively in Chan *et al.* (2019), the zero root can be 'flipped' using an appropriate Blaschke filter (see Lippi and Reichlin (1994)) to obtain a reduced-form VARMA representation $B(L)Y_t = \Theta(L)\varepsilon_t$, in which Θ_0 is full rank, but Θ_q is reduced rank.

The key insight is that the structural moving average operator $A(L)$ can be decomposed as

$$A(L) = \Theta(L)\Phi_0 \begin{bmatrix} L & \\ & I_{n-1} \end{bmatrix} C_1' \tag{15}$$

where $\Theta(L) = I_n + \Theta_1 L + ... + \Theta_q L^q$ (q being the same polynomial degree as in $A(L)$), Φ_0 is invertible, and

$$C_1 = \begin{bmatrix} 0 & 0 & 1 \\ \frac{c}{\sqrt{1+c^2}} & \frac{c}{\sqrt{1+c^2}} & 0 \\ \frac{c}{\sqrt{1+c^2}} & \frac{c}{\sqrt{1+c^2}} & 0 \end{bmatrix} \tag{16}$$

Hence, the structural VARMA $B(L)Y_t$=$A(L)\epsilon_t$, $\epsilon_t \sim N(0,\mathrm{I}_n)$, is observationally equivalent to the reduced-form VARMA $B(L)Y_t = \Theta(L)\varepsilon_t$, $\varepsilon_t \sim N(0,\Sigma)$, with Σ=$\Phi_0\Phi_0'$. The general strategy is then to estimate $B(L)$, $\Theta(L)$ and Σ in the reduced form, and subsequently use the identifying restrictions R1-R3 to obtain Φ_0, C_1, and ultimately, $A(L)$ by applying (15). The details of this methodology are provided in the following sub-section and Online Appendix B.

In fact, Forni *et al.* (2017) also use this approach, but set $\Theta_j = 0$, for all $j \geq 1$, thus estimating a *reduced-form* VAR. After applying the Blaschke transformation to induce a root at zero, they obtain a SVARMA of the form $B(L)Y_t = A_0\epsilon_t + A_1\epsilon_{t-1}$, where *both* A_0 *and* A_1 have the same two proportional columns (e.g. columns 2 and 3 in our case, since ϵ_2 is the news shock and ϵ_3 is the noise shock). In addition, all other columns of of A_1 are zero. The IRFs and FEVDs they compute are from this structural VARMA representation. However, there is no compelling reason—at least in terms of the theoretical models analyzed in Sections 2 and 3—for having these additional restrictions on A_1. In fact, within the models analyzed in these two sections, such restrictions are incorrect, which suggests that imposing them is, in general, unwarranted. We therefore prefer working with a full VARMA specification throughout.

Another approach to estimating the effects of noise shocks based on a VAR was recently proposed by Chahrour and Jurado (2018). Their methodology derives from certain restrictions on the econometrician's information set that induce an observational equivalence between a 'pure news' representation and a 'pure noise' representation. Since a news representation can be estimated using standard VAR methods, the noise representation can then be obtained using a suitable (dynamic) transformation of the VAR residuals.

The restrictions needed for such observational equivalence to hold, however, are only likely to prevail in small data sets. For example, a central concept to the identifi-

cation strategy of Forni et al. (2017) is that the econometrician observes variables that 'reveal the signal', namely financial variables such as stock prices or survey measures of consumer sentiment. Such observable variables are ruled out by the restrictions used in Chahrour and Jurado (2018), and a data set that includes these variables will generally not produce observationally equivalent news and noise representations.

Moreover, we wish to analyse empirically the relative role of news, noise and surprise shocks using a mixed representation. It can be shown that even when pure news and pure noise representations are observationally equivalent, a mixed representation is unique with respect to either of them, unless restrictive assumptions are imposed on the fundamental process. In Online Appendix I, we provide a simple example to clarify these points; a formal treatment of representation uniqueness is provided in Chan et al. (2019). In our ensuing empirical work, we focus on large data sets and avoid any particular restrictions on the fundamental process. Therefore, VARMA methods emerge as the most suitable choice.

To summarize, the SVARMA approach allows us to identify non-news, news and noise shocks by imposing the restrictions in our identification scheme R1-R3 described in sub-section 3.3. To ensure that the impulse responses derived from (14) are uniquely estimated, we also need to impose restrictions on the location of the non-zero roots of $\det A(z)$, which corresponds to imposing restrictions on all the finite roots of $\det \Theta(z)$.

In our default specification, we impose the restriction that all non-zero roots of $\det A(z)$ are outside the unit circle, formally given by

$$\det A(z) \neq 0 \text{ for all } 0 < |z| \leq 1. \tag{17}$$

This restriction is equivalent to requiring all roots of $\det \Theta(z)$ to be outside the unit circle, thus inducing a *fundamental* reduced-form VARMA representation.[23] Consequently, it is easily implemented in practice. However, because the fundamental and all non-fundamental VARMA representations obtained by flipping MA roots are observationally equivalent (Lippi and Reichlin (1994)) in models with Gaussian errors, by imposing (17), we are selecting one of the many (up to 2^{nq}) possible non-fundamental representations with one (real) root at zero from which to draw inference.

The location of the non-zero roots of $\det A(z)$ carries theoretical implications regarding how quickly agents learn about the shocks (including news and noise). As discussed in Chan et al. (2019), assumption (17) implies that all shocks are revealed to agents after at most two periods through observable economic data. We justify this approach in our setting as follows.

Primarily, SVAR identification strategies generally rely on a similar restriction, namely imposing the fundamental representation with all characteristic MA roots outside the unit circle. In our VARMA setting, therefore, restriction (17) can be

[23] In particular, $\det\Theta(z)$ has no roots at zero and at least one root at infinity. All nonzero roots of $\det\Theta(z)$, on the other hand, are identical to the nonzero roots of $\det A(z)$, which assumption (17) imposes to all be outside the unit circle.

viewed as the closest departure from the standard SVAR literature and, in this sense, produces results that are more comparable to those of related papers (e.g. Barsky and Sims (2011)). This facilitates useful analysis such as the one conducted in subsection 6.1.3 below. In addition, for quarterly data, the assumption that agents learn quickly regarding the nature of news and noise shocks is plausible: assuming (17) implies that all shocks are revealed to agents half a year after impact.

Nevertheless, in spite of these arguments in favour of working with a point identified VARMA, we also check the robustness of our results by deriving set-identified IRFs and fractions of FEV for models with $n \leq 8$ in our first application. In particular, for each draw of the parameters in our Bayesian approach, we compute the IRFs and the fractions of FEV for each representation $i = 1, 2, \ldots, K$, either fundamental or non-fundamental. Then, for each variable and each shock, we consider the lower and upper envelopes of the IRFs and of the fractions of FEV across all of the K representations. This yields draws from the posterior distributions for the lower and upper envelopes of the IRFs and fractions of FEV.[24] In a nutshell, our main conclusions still hold based on the set-identified results, which constitutes further support for employing restriction (17).

Finally, we use this same approach to compute IRFs and fractions of FEV for all of the possible representations in our Monte Carlo study which takes, as DGP, the RBC models with noise shocks of Section 3 (see Online Appendix E). We find overwhelming evidence that the representation satisfying (17) produces IRFs and fractions of FEV which closely match the true IRFs and fractions of FEV produced by the RBC model, whereas the other, non-fundamental representations systematically produce different results. We therefore conclude that, at least for the DSGE model considered in this paper, it is reasonable to impose that the corresponding VARMAs has all non-zero roots outside the unit circle.

Our Monte Carlo results along with the population exercise in sub-section 4.1, clearly demonstrate that for SVARMAs with full-rank $A(L)$, restriction (17) together with the three theoretical restrictions outlined in sub-section 3.3, uniquely identify, up to sign normalizations, non-news, news and noise shocks. As discussed in subsection 3.3, however, relevant theoretical models imply additional restrictions that may be empirically useful. Hence, our main set of results for all of our applications also imposes (in some cases) the over-identifying restriction that, two periods after impact, the IRFs generated by the noise shock are smaller, in absolute value, than the corresponding IRFs generated by the news shock.[25]

[24] The reason for focusing on the lower and upper envelopes of the IRFs and fractions of FEV across all of the K representations is that these are well-defined objects, for which it is possible to meaningfully define a posterior distribution. This allows us to characterize IRFs and fractions of FEV taking into account of the fact that each draw from the posterior distribution of the parameters of the VARMA is potentially associated with more than one representation satisfying the restrictions.

[25] As previously mentioned, a slight adjustment has to be made when we allow news and noise shocks to have different standard deviations and, hence, the restriction imposed is actually $|\text{IRF(news)} > \text{IRF(noise)}/c|$.

5.2 Details of Bayesian econometric methods

The general methodology for estimating stuctural VARMA models of the type discussed in the preceding sub-section is provided in Chan et al. (2019). Here we outline this approach, along with details about imposing R1-R3|which are new in the present paper.

In practice, it proves useful to estimate Bayesian VARMAs using the *expanded* VARMA representation (Chan and Eisenstat, 2017; Chan et al., 2016):

$$\tilde{B}_0 y_t = \tilde{B}_1 y_{t-1} + \cdots + \tilde{B}_{p^*} y_{t-p^*} + \Phi_0 f_t + \Phi_1 f_{t-1} + \cdots + \Phi_{p^*} f_{t-p^*} + \eta_t, \tag{18}$$

where $f_t \sim N(0, \Omega)$, $\eta_t \sim N(0, \Lambda)$, Ω and Λ are diagonal, and Φ_0 is lower triangular with ones on the diagonal. The general approach involves sampling expanded form parameters and then transforming them to the VARMA parameters $\Theta_1, \ldots, \Theta_q$ and Σ. The exact procedure based on *generalized Schur decompositions / generalized eigenvalues* is provided in Section 2.3 of Chan and Eisenstat (2017) and Appendix D of Chan et al. (2016). To economize on space, we do not reproduce it here, but only emphasize that it is a computationally simple procedure, even for large VARMA systems. For a detailed discussion in the context of structural VARMA models, see sub-section 3.3 in Chan *et al.* (2019).

We impose linear restrictions and specify priors directly on the expanded form parameters, assuming $p = p^*$, $q < p$. Specifically, we impose parsimony by first setting (with probability one) $\Phi_j = 0$ for all $j = q+1, \ldots, p^*$, which leads to the restrictions $\Theta_j = 0$ for all $j = q+1, \ldots, p^*$. Next, we specify SSVS priors on the individual free elements of $\tilde{B}_0, \ldots, \tilde{B}_p$ and $\Phi_0, \ldots, \Phi_q$ of the form:

$$\begin{aligned}
(\tilde{B}_{j,ik} \,|\, \gamma^B_{j,ik}) &\sim \gamma^B_{j,ik} N(0,1) + (1 - \gamma^B_{j,ik}) N(0, 0.01), \\
(\Phi_{j,ik} \,|\, \gamma^\Phi_{j,ik}) &\sim \gamma^\Phi_{j,ik} N(0,1) + (1 - \gamma^\Phi_{j,ik}) N(0, 0.01), \\
\Pr(\gamma^B_{j,ik} = 1) &= \Pr(\gamma^\Phi_{j,ik} = 1) = 0.5.
\end{aligned}$$

Through extensive experimentation with the resulting algorithm, we nd these settings to produce satisfactory results in both the Monte Carlo exercises and real data applications. Moreover, moderate changes to these priors (including alternative SSVS settings and rank restrictions) do not materially impact the inference on impulse responses.

To complete the prior specification, we set

$$\begin{aligned}
\Omega_{ii} &\sim \mathcal{IG}(5, 1), \\
\Lambda_{ii} &\sim \mathcal{IG}(0, 0.1),
\end{aligned}$$

where $IG(a, b)$ denotes the *inverse gamma distribution* with shape parameter a and rate parameter b. Note that these settings imply weakly informative priors on Ω_{ii} and improper priors on Λ_{ii}. In the paper, we report results holding fixed all of the above

prior settings, but varying the dimension of the system n as well as the lag-lengths p and q.

To facilitate the use of generic priors such as these, we standardize the scale of all series in y_t before commencing MCMC. Specifically, for each original series $y_{i,t}$, we transform to

$$\tilde{y}_{i,t} = \frac{y_{i,t}}{\sqrt{\frac{1}{T}\sum_{t=2}^{T} \Delta y_{i,t}^2}}.$$

After obtaining MCMC draws, we adjust them such as to remove the e ect of the standardization. Hence, all impulse responses are reported on the original, unscaled variables. The approach is equivalent to working directly with yt, but adjusting the priors by the sample standard deviations, as is often done in Bayesian time-series applications (e.g. VARs with Minnesota priors).

Simulation from the posterior of the expanded form VARMA is implemented with Gibbs sampling by cycling through the following four broad steps:

1. Sample $\left(\gamma_i, \tilde{B}_{(i)}, \Phi_{(i)} \mid f, \Lambda_{ii}, y_i\right)$ for each $i = 1, \ldots, n$, where $\tilde{B}_{(i)}$ denotes the i-th row of $\tilde{B} = (\mathrm{I}_n - \tilde{B}_0, \tilde{B}_1, \ldots, \tilde{B}_p)$, $\Phi_{(i)}$ the i-th row of $\Phi = (\Phi_0, \ldots, \Phi_q)$, and γ_i is the set of all SSVS indicators pertaining to $\tilde{B}_{(i)}, \Phi_{(i)}$.

2. Sample $\left(\Lambda_{ii} \mid \tilde{B}_{(i)}\Phi_{(i)}, \gamma_i, f, y_i\right)$ for each $i = 1, \ldots, n$.

3. Sample $(\Omega_{ii} \mid f_i)$ for each $i = 1, \ldots, n$.

4. Sample $\left(f \mid \tilde{B}, \Phi, \Omega, \Lambda, \gamma, y\right)$.

Details and extensive discussion of each sampling step above are provided in Appendix B of Chan et al. (2016).

In summary, we obtain posterior draws from the impulses responses $K(L)$ identi ed by the structural model as follows:

1. Obtain draws of $\tilde{B}_0, \ldots, \tilde{B}_p$, $\Phi_0, \ldots, \Phi_q$, Ω, Λ using the Gibbs sampling algorithm outlined above.

2. For each draw of the expanded form parameters, transform to draws of $B_1, \ldots, B_p$, $\Theta_1, \ldots, \Theta_q$ and Σ.

3. For each draw of $\Theta_1, \ldots, \Theta_q$ and Σ, transform to draws of $\tilde{A}_0, \ldots, \tilde{A}_q$ using the procedure described in sub-section B.1 of the Online Appendix.

4. For each draw of $\tilde{A}_0, \ldots, \tilde{A}_q$, transform to draws of $A_0, \ldots, A_q$ by applying appropriate orthogonal rotations, as detailed in sub-section B.2 of the Online Appendix.

5. For each draw of $B_1, \ldots, B_p$, $A_0, \ldots, A_q$ compute $K(L) = B(L)^{-1}A(L)$ to obtain draws from the posterior distribution of the impulse responses (and with a further trivial transformation, from the forecast error variance decompositions as well).

6 Empirical Application: News and noise shocks to TFP

This section contains an empirical application which investigates the impact upon the economy of TFP non-news, news and noise shocks.

6.1 Data and specification choices

Our data set involves standard US quarterly macroeconomic variables. The Online Appendix provides exact definitions and data sources. The sample period goes from 1954Q3 (when the Federal Funds rate first becomes available) to 2008Q3 (so that we exclude the period during which the Funds rate has been, in practice, at the zero lower bound).

Our data set includes TFP and several other macroeconomic variables. We estimate SVARMAs with different VAR and MA lag lengths, as well as with different choices for the macroeconomic variables other than TFP. In this section we report the results produced by a SVARMA(4,1) featuring 8 variables. In Online Appendices G and H, we present additional results for other lag lengths, and choices of n between 6 and 15, as well as a model comparison exercise which provides evidence that the $p = 4$, $q = 1$, $n = 8$ choice is supported by the data. A key finding is that results are uniformly robust across the different values of n, p and q. This is the case, in particular, for one of our main findings, that is, the overall minor-to-negligible role played by noise shocks in driving macroeconomic variables.

Our main results use the following eight variables: log TFP; log hours *per capita*; the Federal Funds rate; GDP deflator inflation; the logarithms of real GDP, consumption, and investment *per capita*; and the spread between the 5-year government bond yield and the Federal Funds rate.[26] The full set of 15 variables (all of which are

[26] Following common practice, for TFP we use John Fernald's purified TFP series. See Fernald (2012). The notion that productivity may contain a transitory component that exhibits some persistence, originally introduced by Blanchard *et al.* (2013) in order to rationalize the existence of a signal-extraction problem, requires some discussion. In univariate models, the log of TFP is typically found to be close to a random walk. But a transitory component can be justified in terms of data revision. That is, even if the true TFP is a random walk (driven by both news and non-news shocks), most of the data underlying the construction of TFP are progressively revised and this will introduce a persistent, but transitory, component to the estimate of TFP. Alternatively, Blanchard *et al.* (2013) showed that, in principle, it is possible to construct a process for TFP with a permanent and a transitory component which looks like a random walk in the univariate sense.

standard in the literature) is listed in Online Appendix F.

We estimate all models in (log) levels. The main reason for estimating the models in levels has to do with robustness with respect to cointegration of unknown order. As discussed by Hamilton (1994), estimating the system in levels is the robust thing to do in this case. In recent years, estimating VARs in levels has become standard practice in the macroeconomic literature.[27] All of the results we report are based on imposing identifying restrictions R1-R3, outlined in sub-section 3.3. As for the restrictions on the absolute values of the IRFs to news and noise shocks, we report results either imposing, or not imposing this restriction two quarters after impact.

The Online Appendix presents results for a range of choices of p, q and n. For each specification, we present evidence on the convergence properties of the MCMC algorithm. In particular, for the draws from VARMA's reduced-form parameters, we present the draws' first autocorrelations and the inefficiency factors,[28] which we use in order to assess the convergence of the Markov chain. The first-order autocorrelations are uniformly very low, whereas the inefficiency factors are typically around one, indicating an essentially independent sample of draws.

6.1.1 Evidence

Figures 6 shows, for the baseline 8-variables system, the IRFs to the three shocks, based on *point identification*, whereas Figures 7 shows the corresponding results based on *set identification* (i.e., taking into account of all of the possible representations obtained by 'flipping the roots'). FEVs for this specification and others are available in the Online Appendix. Note in particular that Figures VII.2-VII.3 in the Online Appendix report the corresponding results obtained by imposing the additional restriction on the comparative magnitudes of the IRFs to news and noise shocks. Overall, results based on set identification are qualitatively the same as those based on point identification. Further, imposing or not imposing the restrictions on the magnitudes of the IRFs to news and noise shocks does not make any material difference. Accordingly, in what follows we will focus our discussion on the results based on point identification.

Our main substantive finding pertains to the negligible role played by TFP noise shocks in U.S. macroeconomic fluctuations: in Figure 6 it can be seen that all of the IRFs to noise shocks have credible intervals that include zero at all horizons (results for the fractions of FEV, available in the online appendix, support this conclusion). The fact that noise shocks play such a small role, and have so little impact on any of the variables, is the main substantive point we wish to emphasize. This finding is in

[27] See, e.g., Barsky and Sims (2011) or Kurmann and OtrokSims (2013).

[28] The inefficiency factors are defined as the inverse of the relative numerical efficiency measure of Geweke (1992), $RNE = (2\pi)^{-1}\frac{1}{S(0)}\int_{-\pi}^{\pi} S(\omega)d\omega$, where $S(\omega)$ is the spectral density of the sequence of draws from the Gibbs sampler for the quantity of interest at the frequency ω. We estimate the spectral densities via the lag-window estimator as described in chapter 10 of Hamilton (1994). (We also considered an estimator based on the fast-Fourier transform, and results were very close.)

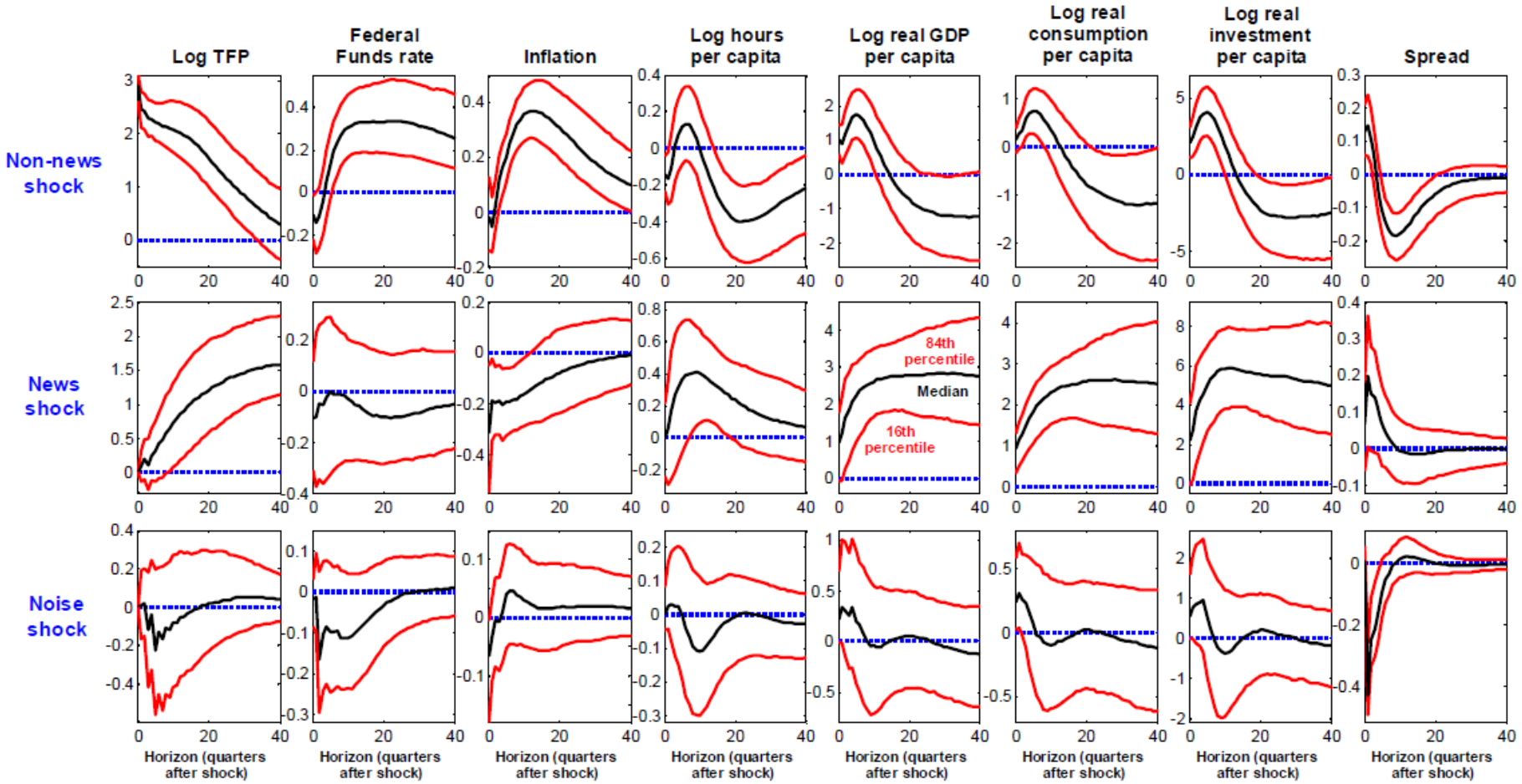

Figure 6 Application with TFP: Impulse-response functions to non-news, news, and noise shocks, based on a VARMA(4,1), and point-identification

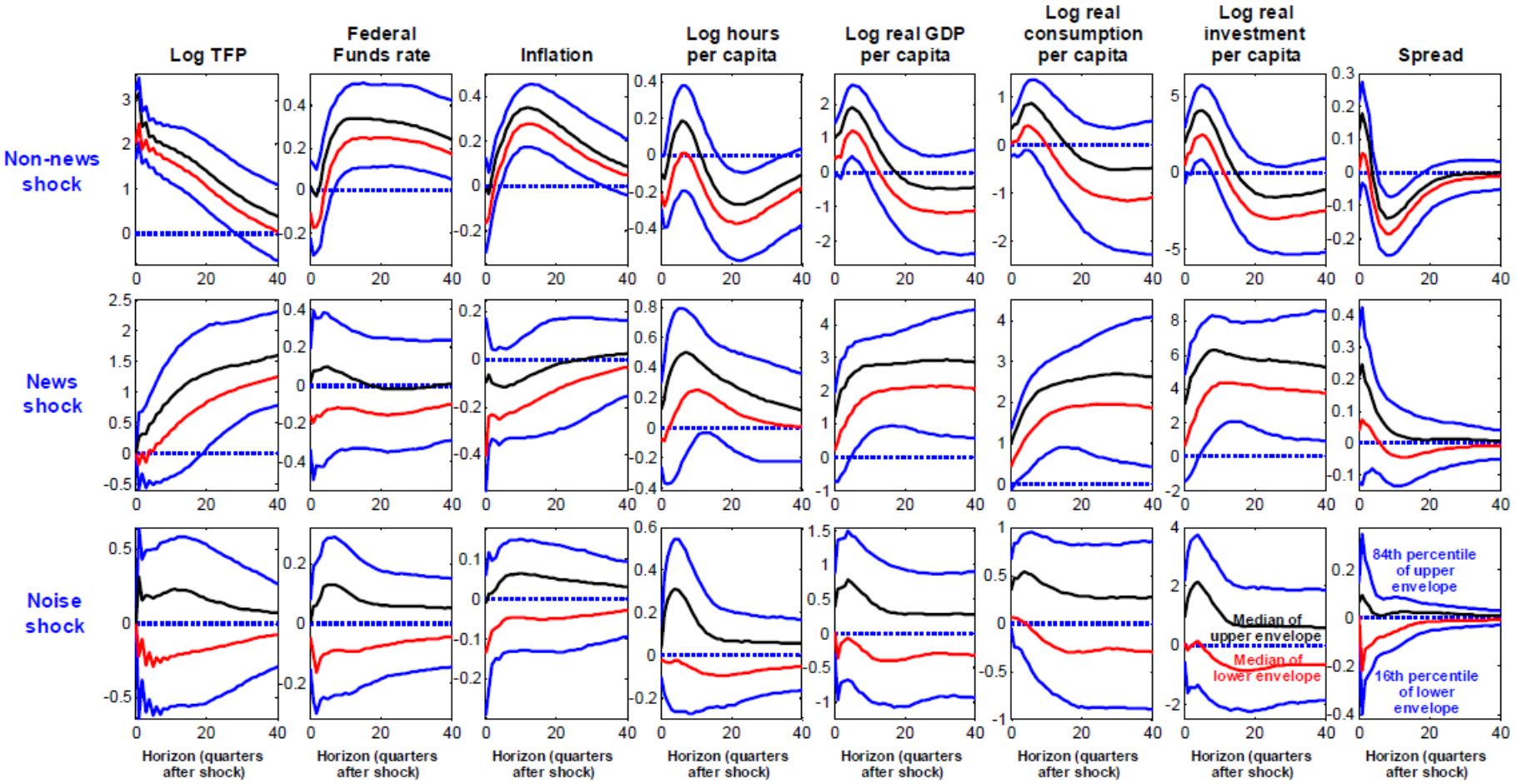

Figure 7 Application with TFP: Impulse-response functions to non-news, news, and noise shocks, based on a VARMA(4,1), and set-identification

contrast to those of Blanchard et al. (2013) and Forni et al. (2017), whereas it is in line with the similar findings of Barsky and Sims (2012) and Chan et al. (2019).

In general, a comparison with the results produced by other authors is not straightforward because of methodological differences (Blanchard et al. (2013), for example, use a DSGE model estimated via Bayesian methods). The only previous paper for which a comparison is meaningful, and it is in fact relatively straightforward, is Forni et al. (2017). As we discussed in Section 5.1, Forni et al.'s methodology is, essentially, a restricted version of ours, in the sense that their SVAR-based approach based on dynamic rotations of the residuals via Blaschke matrices (1) is equivalent to using a SVARMA(p, 1)—as opposed to the more general SVARMA(p, q)'s we use herein; and their MA matrix (i.e., A_1) is quite heavily restricted, with (2) all of the columns other than those pertaining to news and noise shocks being equal to zero, and (3) the two columns pertaining to news and noise being proportional to each other. As we discussed in Section 5.1, from a theoretical perspective neither (1), nor (2) nor (3) are in fact justified. This implies that, as a matter of logic, our results should be regarded as the more reliable.

Figure 6 shows other interesting patterns, which we discuss in the remainder of this sub-section. The IRFs of TFP to news and non-news shocks are in line with the previous literature—see e.g. Barsky and Sims (2011)—with log TFP not jumping, and jumping, respectively, on impact, and then slowly converging to its new long-run value. An important point to stress is that, exactly as in Barsky and Sims (2011), the non-news shock is estimated to be transitory, whereas the news shock clearly has a permanent impact on TFP.

The IRFs for other variables are typically sensible. For non-news shocks, the responses of inflation and the Funds rate are similar to one another. As for news shocks, in line with Barsky and Sims (2011), there is some weak support for the notion that they have a negative impact on inflation at short-horizons.

The IRFs of GDP, consumption, and investment to news shocks mimic those of TFP. Their IRFs to non-news shocks have a different shape from that of TFP, but they are, as for TFP, ultimately transitory. The IRFs of hours to non-news and news shocks are different from those in Barsky and Sims (2011), but, as we will discuss in the next sub-section, they are in line with those produced by a structural VAR identified via the Barsky-Sims methodology based on this dataset. Finally, the impact response of the interest rate spread to news shocks is—as in Kurmann and Otrok (2013)—positive.

6.1.2 Comparison with Barsky and Sims (2011)

How do our results compare to those produced by Barsky and Sims (2011)? Since theirs is an SVAR approach, they can only estimate news and non-news shocks, but not noise shocks. As discussed in Section 3, however, for a sufficiently large lag order, VARs can provide an arbitrarily close approximation to the IRFs for news and non-

news shocks produced by a model also featuring noise shocks. This means that a non-negligible discrepancy between our results and those obtained by Barsky and Sims (2011) would cast doubts on the reliability of our approach, and therefore the meaningfulness of our results. In this sub-section we therefore compare our results for news and non-news shocks to those obtained using an econometric model identical to our own but with no noise shock, and therefore $A_1 = 0$.

Figure I.3 in the Online Appendix reports evidence on this, by showing the IRFs produced by our implementation of Barsky and Sims' (2011) approach for the TFP news and non-news shocks.[29] In order to make these results exactly comparable to those we discussed previously, they have been based on the same eight variables, VAR lag order and estimation sample as was used in the previous sub-section.[30] In line with Barsky and Sims (2011), the two disturbances are identified based on the restrictions that (i) the non-news TFP shock is the only shock that affects TFP on impact, and (ii) the news shock is the one that, among all of the remaining disturbances, explains the maximum fraction of the FEV of TFP at the 20 years ahead horizon. The main finding emerging from Figure I.3 is that the IRFs for the non-news and news shocks produced by the SVAR and by the SVARMA are uniformly very close, which clearly points towards the reliability of the SVARMA-based methodology proposed herein.

7 Conclusions

In this paper we have examined model-based identifying restrictions that allow, in population, to exactly recover news and noise shocks, and we have used them to empirically implemented an identification scheme within a structural VARMA framework. Monte Carlo evidence has shown that our identification scheme has an excellent performance, as it recovers the key features of the postulated data-generation process—the real-business cycle model of Barsky and Sims (2011) augmented with noise shocks about future total factor productivity—with great precision. In an empirical exercise where the news and noise shocks are to TFP, we have provided evidence that noise shocks play a minor role in macroeconomic fluctuations. Additional empirical examples involving stock prices and dividends and defence expenditure[31] produce qualitatively similar results about the minimal role of noise.

[29] We do not show the corresponding results for the fractions of FEV, both for reasons of space, and especially because they are qualitatively the same as those for the IRFs. They are however available upon request.

[30] The VAR is estimated *via* Bayesian methods as in Uhlig (1998, 2005). Specifically, Uhlig's approach is followed exactly in terms of both distributional assumptions—the distributions for the VAR's coefficients and its covariance matrix are postulated to belong to the Normal-Wishart family—and of priors. For estimation details the reader is therefore referred to either the Appendix of Uhlig (1998), or to Online Appendix B of Uhlig (2005). Results are based on 10,000 draws from the posterior distribution of the VAR's reduced-form coefficients and of the covariance matrix of its reduced-form innovations (the draws are computed exactly as in Uhlig (1998, 2005)).

[31] These are available in the working paper version of this paper.

The analysis of identifying restrictions in this paper is potentially useful in a wide variety of contexts, and our scheme for identifying news and noise shocks should also have wide applicability. For instance, it could be used to explore the role played by news and noise shocks with alternative anticipation horizons (i.e., with K news shocks anticipating variation in the relevant variable $1, 2, 3, \ldots, K$ periods ahead) or to investigate the comparative role played by surprise, news and noise shocks in driving fluctuations in other variables, such as real exchange rates. We also provide substantial evidence that SVARMAs constitute a practically useful empirical framework for structural macroeconomic analysis, particularly in settings where SVARs are inappropriate.

References

Robert Barsky and Eric Sims. News shocks and business cycles. Journal of Monetary Economics, 58(3): 273-289, 2011.

Robert Barsky and Eric Sims. Information, animal spirits, and the meaning of innovations in consumer confidence. American Economic Review, 102(4): 1343-1377, 2012.

Luca Benati. Do TFP and the relative price of investment share a common I(1) component? Journal of Economic Dynamics and Control, 45(August 2014): 239-261, 2014.

Olivier Blanchard, Jean-Paul L'Huillier, and Guido Lorenzoni. News, noise, and Fluctuations: An empirical exploration. American Economic Review, 103(7): 3045-70, 2013.

Fabio Canova and Matthias Paustian. Measurement with some theory: Using sign restrictions to evaluate business cycle models. Journal of Monetary Economics, 58: 345-361, 2011.

Ryan Chahrour and Kyle Jurado. News or noise? The missing link. American Economic Review, 108(7): 1702-1736, 2018.

Joshua Chan and Eric Eisenstat. Efficient estimation of Bayesian VARMAs with time-varying coefficients. Journal of Applied Econometrics, 32(7): 1277-1297, 2017.

Joshua Chan, Eric Eisenstat, and Gary Koop. Large Bayesian VARMAs. Journal of Econometrics, 192(2): 374-390, 2016.

Joshua Chan, Eric Eisenstat, and Gary Koop. Quantifying the effects of noise shocks: A structural VARMA approach. Working Paper, 2019. manuscript available at http://joshuachan.org/papers/QuantifyingNoiseShocks.pdf.

John Fernald. A quarterly, utilization-adjusted series on total factor productivity. Federal Reserve Bank of San Francisco Working Paper Series, 2012-19, 2012.

Mario Forni, Luca Gambetti, Marco Lippi, and Luca Sala. Noisy news in business cycles. American Economic Journal: Macroeconomics, 2017.

John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. in J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith (eds.), Bayesian Statistics, Oxford University Press, Oxford, pages 169-193, 1992.

James D. Hamilton. Time Series Analysis. Princeton, NJ, Princeton University Press, 1994.

Lars-Peter Hansen and Thomas J. Sargent. Exact linear rational expectations models: Specification and estimation. Federal Reserve Bank of Minneapolis Staff Report No. 71, 1981.

T. Ito and Danny Quah. Hypothesis testing with restricted spectral density matrices, with an application to uncovered interest parity. International Economic Review, 30(1): 203-215, 1989.

Andre Kurmann and Christopher Otrok. News shocks and the slope of the term structure of interest rates. American Economic Review, 103(6): 2612-32, 2013.

Marco Lippi and Lucrezia Reichlin. VAR analysis, nonfundamental representations, blaschke matrices. Journal of Econometrics, 63(1): 307-325, 1994.

Harald Uhlig. Comment on: The robustness of identified VAR conclusions about money. Carnegie-Rochester Conference Series on Public Policy, 49: 245-263, 1998.

Harald Uhlig. What drives GNP? Unpublished manuscript, Euro Area Business Cycle Network, 2003.

Harald Uhlig. Do technology shocks lead to a fall in total hours worked? Journal of the European Economic Association, 2(2-3): 361-371, 2004.

Harald Uhlig. What are the effects of monetary policy on output? Results from an agnostic identification procedure. Journal of Monetary Economics, 52(2): 381-419, 2005.