A New Approach to Identifying Noise Shocks*

Luca Benati
University of Bern†

Joshua Chan
Australian National University‡

Eric Eisenstat
University of Queensland§

Gary Koop
University of Strathclyde¶

Abstract

A logical implication of agents' inability to distinguish news and noise shocks on impact is that the response of the economy to the two shocks at t=0 will be the same. We provide illustrations of this general property within several macroeconomic models. We then exploit this restriction, together with the fact that whereas news shocks portend future variation in the variable of interest noise shocks do not, to identify news, noise and surprise shocks within a Structural Vector Autoregressive Moving Average (SVARMA) framework. Taking as a data generating process (DGP) Barsky and Sims' (2011) real business cycle (RBC) model augmented with noise shocks about future total factor productivity (TFP), we show in a Monte Carlo study that the proposed SVARMA methodology can reliably recover the DGP's main features in terms of impulse responses and variance decompositions.

In an empirical application, evidence suggests that TFP noise shocks play a minor role in macroeconomic fluctuations, explaining negligible fractions of the forecast error variance of the main macroeconomic variables. In a second applicaton with real dividends and stock prices we show that noise about future dividends explains a very small fraction of the forecast error variance of stock prices.

*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

‡Research School of Economics, Australian National University, Canberra, ACT 0200, Australia. 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

Following the seminal contribution of Beaudry and Portier (2006), the notion that productivity may be driven not only by the familiar 'surprise' (i.e., non-news) shocks, but also by anticipated (i.e., news) shocks is at the forefront of the macroeconomics research agenda.[1] News shocks are defined as those which change expectations about the future value of a specific variable, but have no immediate impact on it.

In recent years, several authors—see in particular Blanchard, L'Huillier, and Lorenzoni (2013) and Forni, Gambetti, Lippi, and Sala (2013)—have started to explore the implications of a straightforward departure from the basic 'news versus non-news' dichotomy by addressing the question: What if some of the news about future technological improvements does not materialize, so that it turns out to have been just noise, rather than news? Blanchard et al. (2013) and Forni et al. (2013) propose two radically different approaches to identifying noise shocks. The former paper estimates, using likelihood methods, a dynamic stochastic general equilibrium (DSGE) model featuring permanent and transitory shocks to technology and a noise shock. The latter paper identifies noise shocks within a Structural Vector Autoregressive (SVAR) framework via dynamic rotations of the VAR's residuals.

In this paper we contribute to this literature by proposing a new, straightforward, and very general approach to disentangling news and noise shocks. Our approach is based on the following general principle. Agents' inability to distinguish news and noise shocks on impact means that the response of the economy to the two shocks at t=0 will be identical. It is only as time goes by that agents learn whether the shock was news or noise. We show how this insight arises in two popular macroeconomic models: a present-value model for dividends and stock prices, and a standard New Keynesian model. We then exploit this restriction, together with the fact that whereas news shocks do anticipate future movements in the relevant variable, noise shocks do not, to identify non-news, news and noise shocks in a SVARMA.

The motivation for use of the SVARMA is that SVAR methods cannot be used, due to our identifying scheme implying a reduced-rank structure for the matrix of impact responses. We show, however, that SVARMA methods can be used to impose such an identifying restriction. In the econometrics literature, a common finding is that VARMAs can be over-parameterized and difficult to estimate, particularly in high dimensional setups such as that used in this paper. Building on previous work with reduced form VARMAs,[2] we derive methods for estimating the SVARMA implied by our identification scheme. We use stochastic search variable selection (SSVS) methods—see in particular George, Sun, and Ni (2008)—to pick out restrictions and ensure parsimony in an otherwise over-parameterized model. In the appendix to this paper is a Monte Carlo study, using Barsky and Sims' (2011) RBC model augmented with noise shocks as a DGP. It shows that our SVARMA methodology can

[1] See Barsky, Basu, and Lee (2014), which provides a summary of the literature.

[2] See Chan, Eisenstat, and Koop (2016).

reliably recover the DGP's main features in terms of impulse responses and variance decompositions whereas an SVAR cannot.

We consider two empirical applications, one where the noisy-news process is embedded in total factor productivity (TFP) and the other involving real dividends and stock prices where the noisy-news process drives dividends. In the former, our main result is that the noise shocks play a minor role in macroeconomic fluctuations, explaining small, or even negligible, fractions of the forecast error variance (FEV) of the main macroeconomic variables at horizons up to 10 years ahead. This contrasts with Blanchard et al. (2013) and Forni et al. (2013) who found noise shocks to play a much larger role. In our second application, we show that noise about future dividends explains very small fractions of the FEV of stock prices. Therefore, although in principle noise about future dividends might reconcile Fama's results about market efficiency with Shiller's findings of excess volatility in stock prices, in fact this does not appear to be the case.

The paper is organized as follows. The next section illustrates the relationship between the impulse response functions (IRFs) on impact to news and noise shocks within a simple present-value model for dividends and stock prices and a standard New Keynesian model driven by shocks to the natural rate of interest. In Section 3 we show that the same property holds within Barsky and Sims' (2011) RBC model augmented with noise shocks about future TFP. In the theoretical SVARMA representation of the augmented Barsky and Sims (2011) model, we show that low-order SVARMAs can approximate very well the model's theoretical IRFs and fractions of FEV explained by individual shocks. This suggests that it should be possible to effectively capture the main features of DGPs featuring noise shocks by estimating low-order SVARMAs. This provides a theoretical rationale for the econometric approach used in the latter half of the paper. Section 4 describes our econometric methodology. Sections 5 and 6 contain our empirical results for applications involving news, noise and surprise TFP shocks and dividend shocks, respectively. Section 7 concludes, and discusses possible directions for future research.

2 Two Simple Theoretical Illustrations

In this section we illustrate the basic features of the response of the economy to news and noise shocks within two theoretical models: a present-value model for dividends and stock prices and a standard New Keynesian model. We do this so as to illustrate the basic concepts in models which can be solved analytically and justify this paper's key identifying restriction: news and noise shocks generate, on impact, identical IRFs for all of the endogenous variables.

2.1 A present-value model for dividends and stock prices

2.1.1 The model and its solution

We start with a standard present-value model for dividends and stock prices, in which log stock prices are equal to the present discounted value of future expected log dividends. Labelling stock prices and log dividends as S_t and d_t, respectively, and using notation where $t|t+j$ 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. The noisy-news feature of the model arises since 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

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

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

where ϵ_t^{NN} and ϵ_t^{NE} are the non-news and news shocks, respectively, with $\epsilon_t^{NN} \sim WN(0, \sigma_{NN}^2)$ and $\epsilon_t^{NE} \sim WN(0, \sigma_{NE}^2)$; $0 < \rho_T < 1$; and $v_t \sim WN(0, \sigma_v^2)$ with WN denoting white noise. Note that the key assumption made in this literature to identify the news shock is to assume that it does not have an immediate impact on the variable of interest, but rather is news about its future variation. In (3), the news shock enters as ϵ_{t-1}^{NE} and there is a one period delay between news and its impact on the permanent component of dividends. Assuming the delay to be longer than one period does not alter the basic theoretical insights of this section. Specification (2)-(4) is the same as the one used by Blanchard, L'Huillier, and Lorenzoni (2013), with the crucial difference that, in their case, the equation for the permanent component of dividends 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.

The role of noise in this model enters through the assumption that, although at time t agents observe d_t, its two individual components, d_t^P and d_t^N, are never observed. Agents do however observe a signal, s_t, which reveals some information about the news shock:

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

with $u_t \sim WN(0, \sigma_u^2)$. Note that, if $u_t = 0$, then the signal is simply equal to the news. By allowing for a non-zero noise shock, u_t, we obtain the noisy-news

specification which is more likely a reasonable characterization for how news arises in the economy.[3]

Appendix A characterizes and solves the agent's signal extraction problem. Based on (1), (2), and (A.5) the solution for log stock prices is given by

$$\ln S_t = \frac{d^P_{t|t}}{1-\beta} + \frac{d^T_{t|t}}{1-\rho_T\beta} + \frac{\beta}{1-\beta}\epsilon^{NE}_{t|t} = \frac{d^P_{t|t}}{1-\beta} + \frac{d^T_{t|t}}{1-\rho_T\beta} + \frac{\beta K_{32}}{1-\beta}\left[\epsilon^{NE}_t + u_t\right], \quad (6)$$

where K_{32} is defined in Appendix A. Since, on impact, neither the news nor the noise shock affects either $d^P_{t|t}$ or $d^T_{t|t}$ it follows that:

$$\left[\frac{\partial \ln S_t}{\partial \epsilon^{NE}_t}\right]_{t=0} = \left[\frac{\partial \ln S_t}{\partial u_t}\right]_{t=0} = \frac{\beta K_{32}}{1-\beta} \quad (7)$$

In words, since, on impact, agents are unable to distinguish between the news and the noise shocks, they will react to either of the two disturbances in the same way, and stock prices will jump by exactly the same amount.

This feature was also noticed, in passing, by Forni, Gambetti, Lippi, and Sala (2013), who pointed out (see their Section 2.2) that

> '[...] the impact responses are identical, since the agents cannot distinguish between the two shocks immediately.'

Forni et al. (2013), however, do not exploit this for identification purposes, as their identification strategy is based on dynamic rotations of the VAR's residuals.

Blanchard et al.'s (2013) IRFs, on the other hand, do not exhibit this property: the reason is that, as previously mentioned, in their case the stochastic process for the permanent component of technology does not feature a news shock. Their analogue to our d^P_t 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 occurs because whereas, at t=0, the noise shock only impacts upon the economy via the signal, the surprise disturbance 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 proper 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 in any way. As a result, their impact on the endogenous variables at t=0 must necessarily be identical.

[3] We use the word news in a broad sense. It can capture authentic announcements of future technological breakthroughs (e.g., Apple announcing that, at some future date, it will release a new and game-changing technology). But it can also reflect the dissemination, within both the technology and investment communities, of the notion that some specific innovation is in the pipeline, and will therefore be realized at some future date, even in the absence of any explicit announcement.

2.1.2 Impulse-response functions to news and noise shocks

The simple model of stock prices and dividends contains four shocks: news (ϵ_t^{NE}), noise (u_t), non-news (ϵ_t^{NN}), and a shock to the transitory process for dividends (v_t). Figure 1 plots IRFs to unitary innovations[4] in these four shocks for the model with 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 the news and noise shocks are identical. 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 important 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 for this is that agents are gradually learning about the nature of the shock. At t=0 agents 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 Barsky and Sims' (2011) RBC model augmented with noise shocks about future TFP, this feature is a remarkably robust one. In particular, we will find two key results in Section 3.2. First, this restriction that IRFs to noise shocks are a dampened versions of those to news shocks holds almost uniformly for all variables and at all horizons after impact. Second, if we look sufficiently far into the future (e.g. our Monte Carlo study suggests at least two periods after impact), it holds for all series and all combinations of the model's parameters.

2.2 A New Keynesian model

The same theoretical properties can also be obtained in a standard forward-looking New Keynesian model:

$$R_t = \phi_\pi \pi_{t+1|t} \tag{8}$$

$$\pi_t = \beta \pi_{t+1|t} + \kappa y_t \tag{9}$$

$$y_t = y_{t+1|t} - \sigma^{-1}[R_t - \pi_{t+1|t} - r_t^N] \tag{10}$$

where R_t, π_t, and y_t are the nominal interest rate, inflation, and the output gap, respectively. r_t^N is the natural rate of interest which is postulated to evolve according to a stationary stochastic process as follows:

$$r_t^N = \tilde{r}_t^N + v_t \tag{11}$$

[4]Note 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. In our empirical work we will plot IRFs to shocks of size equal to one standard deviation. In this case the impact responses to news and noise shocks will not be identical, but they will rather be proportional to each other, with the factor of proportionality being equal to the ratio between σ_u and σ_{NE}.

$$\tilde{r}_t^N = \rho_N \tilde{r}_{t-1}^N + \epsilon_t^{NN} + \epsilon_{t-1}^{NE} \tag{12}$$

where $v_t \sim WN(0, \sigma_v^2)$; $\tilde{r}_t^N$ is the persistent component of the natural rate of interest, with $0 < \rho_N < 1$; and ϵ_t^{NN}, ϵ_t^{NE}, and v_t have the same interpretation, and the same properties, as in sub-section 2.1.

Although at time t agents learn about r_t^N, its two individual components, $\tilde{r}_t^N$ and v_t, are not 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 Appendix A. They imply that

$$\left[\frac{\partial \pi_t}{\partial \epsilon_t^{NE}}\right]_{t=0} = \left[\frac{\partial \pi_t}{\partial u_t}\right]_{t=0} = \kappa\sigma^{-1}\frac{1+\Gamma}{1-\rho_N\Gamma}K_{22}$$

$$\left[\frac{\partial R_t}{\partial \epsilon_t^{NE}}\right]_{t=0} = \left[\frac{\partial R_t}{\partial u_t}\right]_{t=0} = \phi_\pi\kappa\sigma^{-1}\left(1+\rho_N\frac{1+\Gamma}{1-\rho_N\Gamma}\right)K_{22}$$

$$\left[\frac{\partial y_t}{\partial \epsilon_t^{NE}}\right]_{t=0} = \left[\frac{\partial y_t}{\partial u_t}\right]_{t=0} = \sigma^{-1}\left[\frac{(1+\Gamma)(1-\beta\rho_N)}{1-\rho_N\Gamma} - \beta\right]K_{22},$$

where Γ and K_{22} are defined in Appendix A. Just as with the previous model, ϵ_t^{NE} and u_t produce, on impact, the same IRFs for all of the model's endogenous variables (in this case, R_t, π_t, and y_t) whereas, by assumption, they do not impact upon r_t^N (only the news shock impacts upon r_t^N with a one-period delay).

Figure 2 shows IRFs to news and noise shocks for the interest rate, inflation, and the output gap, conditional on a standard calibration of the model's structural parameters.[5] Consistent with the previous discussion, for each variable the impact at t=0 of news and noise shocks is identical. Further, the IRFs to news shocks lie above the corresponding IRFs to noise shocks, reflecting the fact that, just as in the model with dividends and stock prices, agents progressively learn whether a shock was news or noise. In the long run, the IRFs to news shocks progressively converge to their perfect-information counterpart. This can be seen by comparing the black lines and the blue lines in Figure 2, with the former showing the IRFs to news shocks, and the latter representing instead the same IRFs for the case of no noise shocks (i.e. based on the model calibrated as above, but with $\sigma_u^2 = 0$). An implication of this is that separation between the two sets of IRFs will be faster the smaller is the noise, whereas if the noise is substantial (i.e., σ_u^2 is comparatively large), it will take more time for the agents to learn the truth.

2.3 Generality of the result

In our empirical work, we will identify news and noise shocks by imposing the restrictions on the IRFs to the two shocks discussed in the preceding sub-sections. Although

[5] Specifically, we set β=0.99, κ=0.05, σ=1, ϕ_π=1.5, and ρ_N=0.95.

we have demonstrated these properties of news and noise shocks based on two particular theoretical models, it is important to stress their generality. As a matter of logic, these properties should be regarded as a general principle, having nothing to do with the specific nature of the macroeconomic model. For example, it can be trivially shown that the very same result holds within RBC models. In the following section, we will do so in the influential RBC model of Barsky and Sims (2011) augmented with noise shocks about TFP. Nor does it depend on the specific type of news and noise shocks. Nor on whether the process containing the noise and news shocks (e.g. d_t and r_t^N in our examples) is stationary or contains a unit root. The modified Barsky and Sims (2011) model we will use in the next section, for example, features a unit root process for TFP. By the same token, it has also nothing to do with the specific anticipation horizon for the news shock. For example, the same result holds if ϵ_{t-1}^{NE} in (12) is replaced with $\epsilon_{t-\tau}^{NE}$ so that the news is about a shock which will ultimately get realized τ periods ahead. The addition of multiple news shocks with alternative anticipation horizons also does not undermine the general principle.

The key point here is that the very fact that agents cannot distinguish on impact between news and noise—which is the essence of the entire 'news versus noise' problem—automatically implies that, when facing the signal, they will react to the two shocks in exactly the same way. Indeed, if agents' reaction to the two shocks were not the same this would imply that, in some way, they are in fact able to tell the two shocks apart, which contradicts the assumption that the two shocks are not observed, and agents can only observe their sum.

Finally, another point to stress is that this logic suggests that each news shock has its own noise shock. Every time we have announcements about future structural disturbances, they may be authentic news or just plain noise. This should be regarded as the general case with the '100 per cent news' case analyzed, e.g., by Beaudry and Portier (2006), Barsky and Sims (2011), and Kurmann and Otrok (2013) being the extreme case, and the partly news, partly noise case analyzed herein being the general one.

This also clarifies why we choose to frame the problem as we do, instead of as in Blanchard et al. (2013). Remember that their postulated stochastic process for the permanent component of technology (which is analogous to our d_t^P in the stock price-dividends model) does not feature a proper news shock, but only a standard surprise disturbance with a contemporaneous impact. This implies that, when Blanchard et al.'s (2013) model is taken to the data, their surprise shock ought to play the role of both the non-news and the news shocks.

This has two consequences. First, Blanchard et al.'s (2013) approach ignores the fact that news and noise shock produce identical IRFs on impact. Since the surprise disturbance they identify is a linear combination of the authentic surprise and news shocks, it cannot generate, on impact, the same impulse vector as the noise shock.[6]

[6] This holds true unless technology is uniquely driven by news shocks, so that there are no surprise shocks in the data.

This is an important piece of information about the nature of the DGP which is being ignored in estimation. Second, Blanchard et al.'s (2013) approach, by construction, cannot address the crucial issue of the relative importance of news and noise shocks. In order to do that, indeed, it is necessary to separately estimate non-news and news structural disturbances. What their approach provides is rather an assessment of the relative importance of noise shocks, on the one hand, and of 'news plus surprise' shocks, on the other hand.

Based on the previous discussion it is apparent that the magnitude of the impact of noise shocks on the economy will exhibit a hump-shaped relationship with the standard deviation of these shocks. In particular, the economy's reaction to noise shocks will be equal to zero at two extremes: if the standard deviation of these shocks is equal to zero or very large. In the former case, the relevant variable is perfectly observed and there is no role for noise shocks. In the latter case the signal is so noisy that there is no point in reacting to it. In intermediate cases the reaction of the economy to noise shocks will be at first increasing (in absolute value) as the standard deviation of the shocks progressively increases from zero. Beyond a certain value, however, noise shocks become so large that the signal becomes exceedingly noisy, so that it become rational not to react to it, as it carries no meaningful information. All of this implies a sort of 'Laffer curve' for the magnitude of the impact of noise shocks on the economy as a function of their standard deviation.

An implication of this is that there will always exist an upper bound on the importance of the role played by noise shocks in economic fluctuations. In contrast to other shocks, it is not the case that their importance—in terms of the fractions of the FEV of macro variables they explain—is a uniformly increasing function of their standard deviation. Beyond a certain point, it is instead the case that the larger noise shocks are, the smaller the role they play in driving macroeconomic variables.

3 Barsky and Sims' (2011) RBC Model Augmented With Noise Shocks About Future TFP

The preceding section examined the properties of news and noise shocks in simple theoretical models. In our empirical work, we will use SVARMA methods which are atheoretical except for imposing structural identification restrictions motivated by the theory. This section presents evidence that the SVARMA approach is reasonable. In particular, it addresses the questions: "if the underlying theoretical model is an extension of the influential RBC model of Barsky and Sims (2011), can a SVARMA well approximate it? Can a SVAR?" and finds the answers to be "yes" and "no", respectively.

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 $WN(0, \sigma_n^2)$, $0 < b < 1$ is a parameter capturing the strength of habit formation and η captures the curvature of the labor 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 $WN(0, \sigma_g^2)$.

In this model, productivity is captured by A_t and we assume $a_t = \ln(A_t)$ evolves according to

$$a_t = \tilde{a}_t + v_t \tag{13}$$

where $\tilde{a}_t$ is the unobserved permanent component of productivity, and v_t is a white noise disturbance, $v_t \sim WN(0, \sigma_v^2)$. We add noisy-news to this model through the productivity process. In particular, we assume that the permanent component of a_t evolves according to

$$\tilde{a}_t = \tilde{a}_{t-1} + \epsilon_t^{NN} + \epsilon_{t-\tau}^{NE} \tag{14}$$

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 Appendix C.

Figure 3 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 (see Appendix C). As expected, the IRFs to news shocks are qualitatively in line with those reported in Barsky and Sims' (2011) Figure 1, 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 4 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 per cent 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.

In Appendix E we present evidence from a Monte Carlo study that our econometric approach can reliably recover the main features of Barsky and Sims' RBC model augmented with noise shocks—both in terms of IRFs and FEVs—based on samples of typical size. As a preliminary step, in this sub-section we explore, in population, the issue of whether low-order SVARs and SVARMAs can well approximate the RBC model's features, in terms of both IRFs and fractions of FEV.

Figures 5 and 6 compare the theoretical IRFs to non-news, news, and noise shocks generated by the RBC model, and the theoretical fractions of FEV explained by these shocks to those produced by several of its truncated theoretical SVARMA representations.[7] Basically, they show that SVARs cannot capture key features, but low order SVARMAs can. In particular, the following facts emerge from the two figures.

First, a simple SVAR provides reasonably good approximations for non-news shocks in terms of both IRFs[8] and fractions of FEV, and progressively adding MA terms does not materially improve the accuracy.

Second, a SVAR over-estimates quite significantly the magnitudes of the impacts at t=0 for news and noise shocks. The IRFs to news shocks converge to their true values quite fast, so that at sufficiently long horizons the approximation provided by a simple VAR is reasonably accurate. But for noise shocks, approximations are uniformly poor. Turning to the fractions of FEV, those for news shocks are systematically under-estimated—sometimes quite significantly so—whereas those for noise shocks are uniformly over-estimated, although by a lesser extent. The intuition for this is straightforward and will be elaborated on in the next section. But the basic idea is that, within an SVAR 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, the role played by the two shocks in driving macroeconomic fluctuations will end up being necessarily distorted, with that of news shocks being systematically downplayed, and that of noise shocks being uniformly exaggerated.

Third, adding MA terms changes things completely, because it breaks the just-mentioned one-to-one mapping between the two shocks, thus allowing for their separa-

[7] The truncated theoretical SVARMA representations of the RBC model are computed via linear projections methods. Appendix D describes how the truncated SVARMA representations are computed starting from the RBC model's structural moving-average representation, which can immediately be recovered from the model's IRFs.

[8] The IRFs produced by the truncated theoretical SVARMA representations have been normalized as follows. Those to non-news and news shocks have been normalized in such way that the response of TFP at a 'very long' horizon is equal to one (we set the horizon to 1,000 periods after impact). As for those to noise shocks, we normalize them in such a way that the impact vector at t=0 is the same for news and noise shocks.

tion. As a result, the approximations provided by truncated theoretical SVARMA(1, q) representations are superior to those provided by an SVAR(1), in terms of both IRFs and FEV decompositions.

Fourth, an SVARMA(1,1) already provides a reasonably good approximation for either shock to both the true IRFs, and the true fractions of FEV, whereas the approximation provided by an SVARMA(1,2) is uniformly excellent. Although this result has been obtained in a fairly simple DSGE model, at the very least it suggests that, in practice, structural VARMA models with a small MA lag order will be sufficient to accurately approximate the main features of the underlying DGP.

Motivated by the previous discussion, in our empirical work we will impose the following restrictions in order to identify non-news, news, and noise shocks. Following Barsky and Sims (2011),

(1) the non-news shock is identified as the reduced-form innovation to TFP, whereas

(2) 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 (which we will take to be 20 years ahead).

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

(3) news and noise have no immediate impact on TFP and, for the other variables, the IRFs generated by news and noise shocks on impact are proportional to each other.

Restrictions (1)-(3) define what we will label as 'Identification scheme I'.

'Identification scheme II' is the same as Scheme I except for adding another shock to the TFP equation. Hence, (1) is replaced by:

(1') at t=0, TFP is impacted upon by two disturbances, the non-news shock, and a transitory TFP shock which is disentangled from the non-news shock because it explains the minimum fraction of the FEV of TFP at a specific long horizon (again, 20 years ahead).

It is worth stressing from the outset that the two identification schemes produce empirically nearly identical results (the full sets of results based on the two schemes are reported in Online Appendices I and II, respectively). Among the two schemes, Scheme I is the closest to Barsky and Sims' (2011) original approach. From a strictly logical point of view, however, Scheme I suffers from the following problem. As the previous discussion of the RBC model augmented with noise shocks made clear, for these shocks to play any role it ought to be the case that the permanent component of TFP is not perfectly observed, which is obtained by having a white noise shock impacting upon TFP at t=0. By postulating that the non-news shock is the only one impacting upon TFP at t=0, on the other hand, Scheme I is implicitly assuming that the permanent component of TFP is, in fact, perfectly observed, which from a logical point of view rules out the possibility that noise shocks may play a role. Because of this, in what follows we will report and discuss results based on Scheme II.

Although these restrictions are sufficient to identify the three shocks, in what fol-

lows we will also consider imposing additional restrictions, in order to get more precise estimates. The question arises as to whether these restrictions are reasonable in the context of a theoretical model such as that considered in this section. Accordingly, we 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: [0, 0.99]; δ: [0.01, 0.08]; η: [0.1, ∞)[9]; γ: [0.01, 0.1]; $\bar{g}$: [0.15; 0.25]; $\tilde{g}_A$: [0; 0.04] and set the standard deviations of news and noise shocks to σ_{NE}=σ_u=1. The remaining parameters are calibrated as in Barsky and Sims (2011) as described in 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

Table 1 reports the proportion of drawn parameters which produce IRFs which satisfy one possible restriction of interest: that IRFs to news shocks are, in absolute value, greater than those to noise shocks.[10] 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 impose the restriction that news shocks have larger IRFs (in absolute value) than noise shocks.[11]

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

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

[11] 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 the following section.

Another restriction that might be plausible is that news and noise shocks' IRFs have the same sign. Table 2 reports the proportion of drawn parameter values which satisfy this restriction. The numbers in this table are much lower, thus indicating that large regions of the RBC model's parameter space are producing IRFs which are of opposite sign. Accordingly, we do not impose this restriction in our empirical work.

Table 2: Proportion of Drawn Parameters which Satisfy the Restriction: Signs of IRFs to News and Noise Shocks are the Same

Periods after impact	TFP	GDP	Consumption	Investment	Hours
1	0.00	0.04	0.16	0.16	0.04
2	0.00	0.01	0.06	0.41	0.01
3	0.00	0.00	0.01	0.64	0.12
4	0.00	0.00	0.01	0.86	0.46
6	0.00	0.09	0.01	0.92	0.65
8	0.00	0.14	0.00	0.92	0.67

4 Econometric Methodology

4.1 Justification for SVARMA

In this paper we use SVARMAs instead of the SVARs which dominate the empirical macroeconomic literature. In general, several authors (e.g. Cooley and Dwyer (1998)) argue that theoretical macroeconomic models such as DSGE models lead to VMA representations which may not be well approximated by VARs, especially parsimonious VARs with short lag lengths. Such authors suggest VARMA models may be more desirable. And some econometricians (e.g. Poskitt and Yao (2012)) point out that the errors which arise when approximating a VARMA with a finite order VAR can potentially be substantial.[12] In addition to these general reasons for preferring VARMAs to VARs, in our specific case it is not possible to impose our key identifying restriction using an SVAR. An SVARMA is required.

To bring out the main ideas, we use an n-variate SVARMA$(1,1)$

$$\mathbf{y}_t - \mathbf{B}_1\mathbf{y}_{t-1} = \mathbf{u}_t = \mathbf{A}_1\boldsymbol{\epsilon}_{t-1} + \mathbf{A}_0\boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathrm{I}_n). \tag{15}$$

Everything can easily extended to a SVARMA(p,q) with deterministic terms.

Note first that the key aspect of identification relates to the IRFs for $t = 0$ which are in $\mathbf{A}_0$ and that the error covariance matrix, Ω, is $\Omega = A_0'A_0 + A_1'A_1$. In Section 2, we derived results such as (7) which showed that news and noise shocks had an equal response on impact. In terms of the SVARMA, one might think this restriction can be imposed by setting two columns of $\mathbf{A}_0$ to be equal to one another. This is not

[12] Chari, Kehoe, and McGrattan (1998) offer a similar criticism of the SVAR approach.

quite correct, since this 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 to restrict two columns of $\mathbf{A}_0$ to be proportional to one another. It can be shown that the factor of proportionality is equal to the ratio between the standard deviations of the news and noise shocks.

If we impose the identification restriction that two columns of $\mathbf{A}_0$ are proportional to one another, then $\mathbf{A}_0$ 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[13] 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 in 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.

To summarize, the SVARMA approach allows us to identify non-news, news and noise shocks by imposing the restrictions in our Identification Schemes I and II described in the preceding section. Restriction (3) of these Schemes is the key new one for identifying news and noise shocks. In terms of our SVARMA we impose this using the following restrictions (assuming the variable impacted by news and noise shocks, such as TFP, is ordered first in the vector of variables and non-news, news and noise shocks are ordered first, second, and third in the vector of shocks, and $A_{0,ij}$ denotes the $(ij)^{th}$ element, and $A_{0,i}$ the $(i)^{th}$ column of A_0):

- News and noise shocks do not affect TFP at t=0: $A_{0,12} = A_{0,13} = 0$.
- News and noise shocks have a proportional impact on all variables at t=0: $A_{0,2} = cA_{0,3}$ for scalar c.

Our main set of results for both our applications imposes these restrictions along with the dampening restriction on the IRFs discussed in Section 3. The latter imposes the 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. This reflects the dampening effect of agents progressively learning that a shock was noise (as opposed to news) on the way they respond to the shock.[14]

[13] By this we mean that the IRF is based on the VMA respresentation which involves inverting the VAR which puts B_1 in every lag of the VMA representation.

[14] As mentioned in footnote 11, 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 $|\text{IRF(news)} > \text{IRF(noise)}/c|$.

4.2 Estimating the SVARMA

Despite having many attractive statistical properties, VARMAs are rarely used in practice since they can be over-parameterized and identification and computation can be difficult. That is, even VARs are often over-parameterized. Ignoring deterministic terms, a VAR(p) will have pn^2 VAR coefficients, which can be large, even with the moderately sized VARs often used in the structural macroeconomics literature (e.g. Barsky, Basu and Lee, 2014, use a VAR with $n = 9$). Adding a VMA component increase this to $(p + q) n$ making over-parameterization concerns even worse. With structural SVARs, there is a need for structural identification restrictions. With SVARMAs there is the same need, but even the reduced form VARMA suffers from a lack of identification which arises from the possibility of common factors in the VAR and VMA parts of the model. These issues are discussed in detail in Chan et al. (2016). A crucial aspect of this approach is the use of prior shrinkage to mitigate over-parameterization concerns. Chan et al. (2016) also develops Bayesian methods for estimating VARMAs which work even with large values of n. We adapt these methods to the SVARMA described in the preceding sub-section, resulting in an efficient Markov Chain Monte Carlo (MCMC) algorithm. Complete details of how we carry out Bayesian estimation are given in Appendix B. Here we provide a basic outline of our strategy.

As discussed in Chan and Eisenstat (2015) and Chan et al. (2016), estimation of VARMA models is more easily done when they are put in so-called expanded form. The expanded form writes the VARMA as a state space model and standard Bayesian Markov MCMC algorithms for state space models can be used, greatly simplifying computation.[15] In Appendix B, we show how the SVARMA written in (15) with the identification restrictions imposed on A_0 can be written in equivalent form as an expanded form VARMA. We use a prior designed to ensure shrinkage and parsimony in the potentially over-parameterized SVARMA. Several different priors are popular in the Bayesian VAR literature, including the Minnesota prior and various hierarchical shrinkage priors (e.g. various LASSO priors, spike-and-slab priors). In this paper, we use the SSVS prior introduced to the Bayesian VAR literature by George, Sun, and Ni (2008) and used in many VAR papers.[16] The basic idea can be explained in terms of a generic VAR or MA coefficient, say θ. SSVS specifies a hierarchical prior (i.e. a prior expressed in terms of parameters which in turn have a prior of their own) which is a mixture of two Normal distributions:

$$\theta|\gamma \sim (1-\gamma)\, N\left(0, \kappa_0^2\right) + \gamma N\left(0, \kappa_1^2\right), \tag{16}$$

where $\gamma \in \{0, 1\}$ is an unknown parameter. If $\gamma = 0$ the prior for θ is given by the

[15] An alternative is the random walk Metropolis-Hastings (M-H) algorithm. However, M-H algorithms tend to be less efficient than Gibbs sampling algorithms even in small models, much less high-dimensional models such as the ones considered in this paper.

[16] E.g., among many others, Koop (2013), and Korobilis (2013).

first Normal distribution, and if it is $\gamma = 1$ its prior is given by the second. The prior is hierarchical since γ is treated as an unknown parameter which is estimated in a data-based fashion. The first prior variance, κ_0^2, is chosen to be "small" (so that the coefficient is constrained to be virtually equal to zero) and the second prior variance, κ_1^2, to be "large" (implying a relatively noninformative prior for the corresponding coefficient). Thus, SSVS allows for the data to decide which coefficients should be set to zero so as to ensure parsimony in the SVARMA. The only subjective prior information that is required is the choice of κ_0^2 and κ_1^2, but standard methods exist for their choice. Details of how this is done and the MCMC algorithm which results from use of the SVARMA model with SSVS prior are given in Appendix B.

5 Application 1: News and Noise Shocks to TFP

In this section we provide evidence on the impact upon the economy of TFP non-news, news and noise shocks.

5.1 Data and Specification Choices

Our data set includes TFP plus several other macroeconomic variables. We have estimated SVARMAs with different lag lengths and different choices for the other macroeconomic variables. In the main text we report the results produced by a SVARMA(4,1) using an 8-variable system. In Appendices G and H we present additional empirical results for other lag lengths and choices of n, 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 FED Funds rate.[17] In our Empirical Appendix, we consider up to 15 variables. All these variables are standard in the literature and described in Appendix F. The sample period goes from 1954Q3 (when the Federal Funds rate first

[17]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 which 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.

becomes available) to 2008Q3 (so that we exclude the period during which the FED Funds rate has been, in practice, at the zero lower bound).[18]

We estimate all models in (log) levels. The main reason for estimating the model in levels has to do with robustness with respect to cointegration of an unknown order. As discussed by Hamilton (1994), estimating the system in levels is the conservative and robust thing to do in this case. In recent years, estimating VARs in levels has become standard practice in the macroeconomic literature (see, e.g., Barsky and Sims (2011) or Kurmann and Otrok (2013)).

The results in this section impose all of the Scheme II identifying restrictions specified in sub-section 4.1. In addition, we include the restriction on the absolute magnitudes of the IRFs to noise and news shocks at a specific horizon (two quarters after impact). Results are also presented without the latter restriction.

The Online Appendices associated with this paper 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 the ergodic distribution of the VARMA's reduced-form parameters, we present the draws' first autocorrelations and the inefficiency factors,[19] which we use in order to assess the convergence of the Markov chain. The first autocorrelations are uniformly very low, whereas the inefficiency factors are all around one, which is much lower than the 20-25 value which is typically taken as signalling problems in convergence to the ergodic distribution.

5.2 Empirical Results

Figures 7 and 8 show the medians, together with credible intervals (i.e. the 16-84 percentiles of the posterior distributions), of the fractions of the k-step-ahead FEV due to non-news, news, and noise shocks, and of the impulse-response functions to the shocks, respectively. The main finding emerging from these figures is the negligible role played by noise shocks for U.S. macroeconomic fluctuations. In Figure 7 it can be seen that noise shocks explain very small fractions of the FEV of all series at all horizons. In Figure 8 it can be seen that all of the IRFs to noise shocks have credible intervals that include zero at all horizons. The fact that noise shocks explain such a small role and have so little impact on any of the variables is the main point we want

[18] The FED Funds rate reached 0.16 per cent in December 2008, and it has been around that level ever since.

[19] 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.)

to emphasize in our results. However, Figures 7 and 8 show other interesting patterns relating to the other shocks which we discuss in the remainder of this sub-section.

News shocks explain small fractions of the FEV of TFP at short horizons, but their importance progressively increases. At the 10 year horizon, news shocks are playing a very big role: they explain around 40% of the FEV for TFP. News shocks also play a very important role for GDP, consumption and investment, especially at long horizons. For the remaining variables, they play a smaller but non-negligible role (e.g. the point estimates indicate they tend to explain around 20% of the FEV).

Non-news shocks account for 100% of the FEV of TFP on impact, but after this their importance progressively diminishes. However, even at the 10 year horizon they still account for roughly 50% of the FEV. As for other series, non-news shocks tend to play a small role in the FEV. Only for inflation and, to a lesser extent, the Fed Funds rate, do they have a non-negligible role in explaining the FEV (e.g. at longer horizons point estimates indicate they explain roughly 20% of the FEV).

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 in line with Barsky and Sims (2011), the non-news shock is estimated to be transitory, whereas the news shock clearly has a permanent impact on TFP. As noted above, the noise shock does not seem to have any impact on TFP at any horizon.

The IRFs for the other variables are sensible. The responses of inflation and the Fed Funds rate are similar to one another. They are negative for news and noise shocks, but the credible intervals rarely exclude zero. There is some weak support for the idea, in line with Barsky and Sims (2011), that news shocks have a negative impact on inflation at short-horizons. However, the IRFs for the non-news shocks are strongly positive for both variables. Consistent with our FEV findings, the non-news shock is largely associated with these two variables.

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.

5.2.1 Comparison with Barsky and Sims (2011)

How do our results compare with those produced by Barsky and Sims' (2011) approach? Since theirs is an SVAR approach, they can only estimate TFP news and non-news shocks, but not noise shocks. But our approach only differs through the

presence of the A_1 matrix and, hence, we would expect results for news and non-news shocks to be similar between the two approaches. Accordingly, as a gauge of the reliability of our approach, in this sub-section we 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 $A_1 = 0$.

Figure 9 report 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.[20] 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 section.[21] In line with Barsky and Sims (2011), and with the way we previously identified non-news and news TFP shocks, the two disturbances are identified based on the restrictions that (i) the non-news TFP shocks is the only shock which affects TFP on impact, and (ii) the news shock is the one which, among all of the remaining disturbances, explains the maximum fraction of the FEV of TFP at the 80 quarters ahead horizon.

The main finding emerging from Figure 9 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.

We now turn to an application to a model with dividends and stock prices.

6 Application 2: Evidence on Noise Shocks About Dividends

One interesting feature of noise shocks is that, in principle, noise about future dividends might reconcile Fama's results about market efficiency with Shiller's findings of excess volatility in stock prices. The reason for this is straightforward, and it can be immediately grasped from the solution for log stock prices given in (6). The crucial feature is that, since S_t depends on noise shocks, a sufficiently large variance of these shocks could cause rationally determined stock prices to exhibit excess volatility compared to their fundamental value. An important caveat to this is that, for the reasons we discussed in Section 2.4, there is an upper limit to the fractions of

[20] 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.

[21] 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 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)).

FEV of any variable which can be explained by noise shocks. This upper limit is however model-specific, so that, in principle, noise about future dividends might in fact reconcile the two strands of literature.

Whatever might be possible in principle, however, in fact this does not appear to be the case in practice. Figures 10-11 show results from a 7-variable VARMA(4,1) for inflation, the FED Funds rate, the spread between the 5-year government bond yield and the FED Funds rate, and the logarithms of real dividends, real stock prices, hours per capita and labor productivity. All specification choices are the same as in the previous application. The identifying scheme assumes that the noisy-news process affects dividends.

For the sake of brevity, we will leave the reader to view the results for each individual variable, but highlight only our main finding. This is given in panel (3,2) of Figure 11, showing how noise about future dividends plays an uniformly minor role in driving stock prices. Therefore, although in principle noise about future dividends might reconcile Fama's results about market efficiency with Shiller's findings of excess volatility in stock prices, in fact this does not appear to be the case.

7 Conclusions

News and noise shocks are, by definition, impossible to distinguish on impact. Thus, the response of the economy to the two shocks at t=0 will be the same. In this paper, we have provided several illustrations of this property within different theoretical macroeconomic models. We then use this insight to develop a new scheme for identifying news and noise shocks in empirical macroeconomic models. We show how SVARs are not able to incorporate this scheme in theory and in practice can approximate it poorly and, thus, use a SVARMA framework. Taking as DGP Barsky and Sims' (2011) RBC model augmented with noise shocks about future TFP, we show via Monte Carlo that the proposed SVARMA-based methodology can reliably recover the DGP's main features in terms of both IRF and FEV decompositions. In two empirical applications, we present evidence suggesting that noise shocks play a minor role in macroeconomic fluctuations, explaining negligible fractions of the forecast error variance of the main macroeconomic variables.

The methodology introduced in this paper is potentially useful in a wide variety of contexts. In terms of the econometric methods, we established that SVARMA models of the large dimension increasingly used in macroeconomics can be easily and robustly estimated. SVARMAs allow a researcher to do things which are impossible to do within the SVAR framework. 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 fiscal news and noise shocks with alternative anticipation horizons (i.e., with K news shocks anticipating variation in the relevant fiscal variable 1, 2, 3, ..., K periods ahead) or to investigate the comparative role played by surprise, news and noise shocks in driving fluctuations in other variables such as the real exchange rate.

References

Barsky, R., S. Basu, and K. Lee (2014): "Whither News Shocks?," *in Jonathan Parker and Michael Woodford, editors, NBER Macroeconomics Annuals*, 29, 225–264.

Barsky, R., and E. Sims (2011): "News Shocks and Business Cycles," *Journal of Monetary Economics*, 58(3), 273–289.

Beaudry, P., and F. Portier (2006): "Stock Prices, News, and Economic Fluctuations," *American Economic Review*, 96(4), 1293–1307.

Blanchard, O., J.-P. L'Huillier, and G. Lorenzoni (2013): "News, Noise, and Fluctuations: An Empirical Exploration," *American Economic Review*, 103(7), 3045–70.

Canova, F., and M. Paustian (2011): "Measurement with Some Theory: Using Sign Restrictions to Evaluate Business Cycle Models," *Journal of Monetary Economics*, 58, 345–361.

Celeux, G., F. Forbes, C. Robert, and D. Titterington (2006): "Deviance Information Criteria for Missing Data Models," *Bayesian Analysis*, 1(4), 651–674.

Chan, J., and E. Eisenstat (2015): "Efficient Estimation of Bayesian VARMAs With Time-Varying Coefficients," manuscript available at http://people.anu.edu.au/joshua.chan/papers/VARMA_tvp.pdf.

Chan, J., E. Eisenstat, and G. Koop (2016): "Large Bayesian VARMAs," *Journal of Econometrics*, 192(2), 374–390.

Chan, J., and A. Grant (2016): "Fast Computation of the Deviance Information Criterion for Latent Variable Models," *Computational Statistics and Data Analysis*, 100, 847–859.

Chari, V., P. Kehoe, and E. McGrattan (1998): "Are Structural VARs With Long-Run Restrictions Useful in Developing Business Cycle Theory?," *Journal of Monetary Economics*, 55(8), 1337–1352.

Cooley, T., and M. Dwyer (1998): "Business Cycle Analysis Without Much Theory: A Look at Structural VARs," *Journal of Econometrics*, 83, 57–88.

Dadda, C., and A. Scorcu (2003): "On the Time Stability of the Output-Capital Ratio," *Economic Modelling*, 20(6), 1175–1189.

Fernald, J. (2012): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," Federal Reserve Bank of San Francisco Working Paper Series, 2012-19.

Forni, M., L. Gambetti, M. Lippi, and L. Sala (2013): "Noisy News in Business Cycles," mimeo.

George, E., D. Sun, and S. Ni (2008): "Bayesian Stochastic Search for VAR Model Restrictions," *Journal of Econometrics*, 142, 553–580.

Geweke, J. (1992): "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*, pp. 169–193.

Ghosh, J., and D. B. Dunson (2009): "Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis," *Journal of Computational and Graphical Statistics*, 18(2), 306–320.

Gustafson, P. (2005): "On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables," *Statistical Science*, 20(2), 111–140.

Hamilton, J. (1994): *Time Series Analysis*. Princeton, NJ, Princeton University Press.

Imai, K., and D. A. van Dyk (2005): "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation," *Journal of Econometrics*, 124(2), 311Ű–334.

Koop, G. (2013): "Forecasting with Medium and Large Bayesian VARs," *Journal of Applied Econometrics*, 28, 177–203.

Koop, G., R. León-González, and R. W. Strachan (2010): "Efficient Posterior Simulation for Cointegrated Models with Priors on the Cointegration Space," *Econometric Reviews*, 29(2), 224–242.

——— (2012): "Bayesian Model Averaging in the Instrumental Variable Regression Model," *Journal of Econometrics*, 171(2), 237 – 250.

Korobilis, D. (2013): "VAR Forecasting Using Bayesian Variable Selection," *Journal of Applied Econometrics*, 28, 204–230.

Kuo, L., and B. Mallick (1997): "Bayesian Semiparametric Inference for the Accelerated Failure-Time Model," *Canadian Journal of Statistics*, 25(4), 457–472.

Kurmann, A., and C. Otrok (2013): "News Shocks and the Slope of the Term Structure of Interest Rates," *American Economic Review*, 103(6), 2612–32.

Lippi, M., and L. Reichlin (1994): "VAR Analysis, Nonfundamental Representations, Blaschke Matrices," *Journal of Econometrics*, 63(1), 307–325.

Liu, J. S., and Y. N. Wu (1999): "Parameter Expansion for Data Augmentation," *Journal of the American Statistical Association*, 94(448), 1264–1274.

Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer-Verlag.

Meng, X.-L., and D. van Dyk (1999): "Seeking Efficient Data Augmentation Schemes via Conditional and Marginal Augmentation," *Biometrika*, 86(2), 301–320.

Plagborg-Møller, M. (2016): "Bayesian Inference on Structural Impulse Response Functions," available at: http://scholar.harvard.edu/files/plagborg/files/irf_bayes.pdf.

Poskitt, D., and W. Yao (2012): "VAR Modeling and Business Cycle Analysis: A Taxonomy of Errors," Department of Econometrics and Business Statistics, Monash University, Working Paper 11-12.

Sargent, T. J. (1987): *Macroeconomic Theory, II Edition*. Orlando, Academic Press.

Spiegelhalter, D., N. Best, B. Carlin, and A. vanderLinde (2002): "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society Series B*, 64(4), 583–639.

Uhlig, H. (1998): "Comment On: The Robustness of Identified VAR Conclusions About Money," *Carnegie-Rochester Conference Series on Public Policy*, 49, 245–263.

——— (1999): "A Toolkit for Analysing Nonlinear Dynamic Stochastic Models Easily," *in Ramon Marimon and Andrew Scott, eds, Computational Methods for the Study of Dynamic Economies, Oxford University Press*, pp. 30–61.

——— (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52(2), 381–419.

A Solutions to the Theoretical Models

A.1 A present-value model for dividends and stock prices

The agent's signal-extraction problem can be characterized as follows. First, define $Y_t \equiv [d_t,\, s_t]'$, $\xi_t \equiv [d_t^P,\, d_t^N,\, \epsilon_t^{NE}]'$, $w_t \equiv [\epsilon_t^{NN}, v_t, \epsilon_t^{NE}]'$, $\eta_t \equiv [0,\, u_t]'$, and

$$
\begin{aligned}
F &\equiv \begin{bmatrix} 1 & 0 & 1 \\ 0 & \rho_T & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad H \equiv \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}' \\
Q &\equiv \begin{bmatrix} \sigma_{NN}^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_{NE}^2 \end{bmatrix} \qquad R \equiv \begin{bmatrix} 0 & 0 \\ 0 & \sigma_u^2 \end{bmatrix}
\end{aligned}
$$

equations (2)-(5) can be cast in state-space form as:

$$Y_t = H'\xi_t + \eta_t \tag{A.1}$$

$$\xi_t = F\xi_{t-1} + w_t \tag{A.2}$$

The estimate of the state vector conditional on information at time t, $\xi_{t|t}$, together with its estimated covariance matrix, $P_{t|t} \equiv E[(\xi_t - \xi_{t|t})(\xi_t - \xi_{t|t})'|t]$ can be obtained via the following Kalman filtering recursions (see Hamilton (1994)):

$$\xi_{t|t} = F\xi_{t-1|t-1} + K_t[Y_t - H'F\xi_{t-1|t-1}] \tag{A.3}$$

$$P_{t|t} = FP_{t-1|t-1}F' + Q - K_tH'(FP_{t-1|t-1}F' + Q) \tag{A.4}$$

where $K_t \equiv (FP_{t-1|t-1}F' + Q)H[H'(FP_{t-1|t-1}F' + Q)H + R]^{-1}$ is the Kalman gain. The steady-state value of the precision matrix $P_{t|t}$ is obtained by iterating on (A.4) starting from $P_{0|0} = Q$, thus also obtaining the steady-state value of the Kalman gain, which, being time-invariant, in what follows will simply be referred to as K. Based on the steady-state Kalman gain, and defining $\tilde{K} \equiv (I_3 - KH')F$, where I_3 is the 3×3 identity matrix, the solution to the the signal-extraction problem is therefore given by

$$
\underbrace{\begin{bmatrix} d_{t|t}^P \\ d_{t|t}^T \\ \epsilon_{t|t}^{NE} \end{bmatrix}}_{\equiv \xi_{t|t}} = \underbrace{\begin{bmatrix} \tilde{K}_{11} & \tilde{K}_{12} & \tilde{K}_{13} \\ \tilde{K}_{21} & \tilde{K}_{22} & \tilde{K}_{23} \\ 0 & 0 & 0 \end{bmatrix}}_{\equiv \tilde{K}} \underbrace{\begin{bmatrix} d_{t-1|t-1}^P \\ d_{t-1|t-1}^T \\ \epsilon_{t-1|t-1}^{NE} \end{bmatrix}}_{\equiv \xi_{t-1|t-1}} + \underbrace{\begin{bmatrix} K_{11} & 0 \\ K_{21} & 0 \\ 0 & K_{32} \end{bmatrix}}_{\equiv K} \underbrace{\begin{bmatrix} d_t^P + d_t^T \\ \epsilon_t^{NE} + u_t \end{bmatrix}}_{\equiv Y_t}. \tag{A.5}
$$

A.2 A new Keynesian model

The agents' signal-extraction problem can be characterized as follows. By defining $Y_t \equiv [r_t^N,\, s_t]'$, $\xi_t \equiv [\tilde{r}_t^N,\, \epsilon_t^{NE}]'$, $w_t \equiv [\epsilon_t^{NN},\, \epsilon_t^{NE}]'$, $\eta_t \equiv [v_t,\, u_t]'$, and

$$F \equiv \begin{bmatrix} \rho_N & 1 \\ 0 & 0 \end{bmatrix} \quad H \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad Q \equiv \begin{bmatrix} \sigma_{NN}^2 & 0 \\ 0 & \sigma_{NE}^2 \end{bmatrix} \quad R \equiv \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix}$$

equations (11), (12), and (5) can be cast in the state-space form (A.1)-(A.2). As before, the solution to the signal-extraction problem can be obtained by applying the Kalman filter recursions (A.3)-(A.4) to the state-space form (A.1)-(A.2), thus obtaining the solution

$$\underbrace{\begin{bmatrix} \tilde{r}_{t|t}^N \\ \epsilon_{t|t}^{NE} \end{bmatrix}}_{\equiv \xi_{t|t}} = \underbrace{\begin{bmatrix} \tilde{K}_{11} & \tilde{K}_{12} \\ 0 & 0 \end{bmatrix}}_{\equiv \tilde{K}} \underbrace{\begin{bmatrix} \tilde{r}_{t-1|t-1}^N \\ \epsilon_{t-1|t-1}^{NE} \end{bmatrix}}_{\equiv \xi_{t-1|t-1}} + \underbrace{\begin{bmatrix} K_{11} & 0 \\ 0 & K_{22} \end{bmatrix}}_{\equiv K} \underbrace{\begin{bmatrix} \tilde{r}_t^N + v_t \\ \epsilon_t^{NE} + u_t \end{bmatrix}}_{\equiv Y_t}. \tag{A.6}$$

where K is still the steady-state Kalman gain at time t, and $\tilde{K} \equiv (I_2 - K)F$, where I_2 is the 2×2 identity matrix.

To obtain the model's solution, we first substitute (8) into (10) to obtain

$$y_t = y_{t+1|t} - \sigma^{-1}[(\phi_\pi - 1)\pi_{t+1|t} - r_t^N] = \tag{A.7}$$

$$= -\sigma^{-1}[(\phi_\pi - 1)\pi_{t+1|t} - r_t^N] + y_{t+2|t} - \sigma^{-1}[(\phi_\pi - 1)\pi_{t+2|t} - r_{t+1|t}^N]$$

From (11)-(12) we have that $r_{t+1|t}^N = \rho_N \tilde{r}_{t|t}^N + \epsilon_{t|t}^{NE}$, so that the previous equation becomes

$$y_t = -\sigma^{-1}[(\phi_\pi - 1)\pi_{t+1|t} - r_t^N] + y_{t+2|t} - \sigma^{-1}[(\phi_\pi - 1)\pi_{t+2|t} - (\rho_N \tilde{r}_{t|t}^N + \epsilon_{t|t}^{NE})] \tag{A.8}$$

From (9) we get $y_t = \kappa^{-1}[\pi_t - \beta \pi_{t+1|t}]$, and substituting this into the previous expression, we get the following expectational difference equation for inflation:

$$\pi_t - \pi_{t+1|t}[\beta - \kappa\sigma^{-1}(\phi_\pi - 1)] - \pi_{t+2|t}[1 - \kappa\sigma^{-1}(\phi_\pi - 1)] + \beta\pi_{t+3|t} =$$

$$= \kappa\sigma^{-1}[r_t^N + \rho_N \tilde{r}_{t|t}^N + \epsilon_{t|t}^{NE}] \tag{A.9}$$

Assuming that the condition for determinacy is satisfied (which, as it can easily be checked, boils down to ϕ_π being greater than 1), the solution can be found via the method of undetermined coefficients. Postulating that inflation is a linear function of the three states—r_t^N, $\tilde{r}_{t|t}^N$, and $\epsilon_{t|t}^{NE}$—that is,

$$\pi_t = \alpha_1 r_t^N + \alpha_1 \tilde{r}_{t|t}^N + \alpha_1 \epsilon_{t|t}^{NE} \tag{A.10}$$

the solution turns out to be equal to

$$\pi_t = \kappa\sigma^{-1} r_t^N + \kappa\sigma^{-1}\rho_N \frac{1+\Gamma}{1-\rho_N\Gamma} \tilde{r}_{t|t}^N + \kappa\sigma^{-1} \frac{1+\Gamma}{1-\rho_N\Gamma} \epsilon_{t|t}^{NE} \tag{A.11}$$

with the analogous solutions for R_t and y_t being

$$R_t = \phi_\pi \rho_N \kappa \sigma^{-1} \left(1 + \rho_N \frac{1+\Gamma}{1-\rho_N \Gamma}\right) \tilde{r}^N_{t|t} + \phi_\pi \kappa \sigma^{-1} \left(1 + \rho_N \frac{1+\Gamma}{1-\rho_N \Gamma}\right) \epsilon^{NE}_{t|t} \quad \text{(A.12)}$$

$$y_t = \sigma^{-1} r^N_t + \rho_N \sigma^{-1} \left[\frac{(1+\Gamma)(1\text{-}\beta\rho_N)}{1-\rho_N\Gamma}\text{-}\beta\right] \tilde{r}^N_{t|t} + \sigma^{-1} \left[\frac{(1+\Gamma)(1\text{-}\beta\rho_N)}{1-\rho_N\Gamma}\text{-}\beta\right] \epsilon^{NE}_{t|t} \quad \text{(A.13)}$$

where

$$\Gamma \equiv \beta - \kappa\sigma^{-1}(\phi_\pi - 1) + \rho_N[1 - \kappa\sigma^{-1}(\phi_\pi - 1)] - \beta\rho_N^2.$$

B Econometric Methods

B.1 General Framework

The econometric methodology employed in this paper extends the approach of Barsky and Sims (2011) to VARMA processes. A general approach to estimating structural models in the VAR literature can be characterized as estimating a reduced-form VAR and then recovering structural shocks from VAR residuals by applying *constant* orthogonal rotations to each time t realization of the VAR residuals. In the context of VARMAs, this idea can be generalized to an approach that recovers structural VARMA shocks from reduced-form VARMA residuals by applying *dynamic* orthogonal rotations to combinations of past, present, and future realizations of the VARMA residuals.

In particular, consider the n-dimensional VARMA(p, q) process given by[22]

$$\mathbf{B}(L)\mathbf{y}_t = \tilde{\mathbf{\Theta}}(L)\tilde{\boldsymbol{\epsilon}}_t, \qquad \tilde{\boldsymbol{\epsilon}}_t \sim \mathcal{N}(0, \mathbf{\Sigma}), \quad \text{(B.1)}$$

where $\mathbf{B}(L) = \mathrm{I}_n - \mathbf{B}_1 L - \cdots - \mathbf{B}_p L^p$ and $\tilde{\mathbf{\Theta}}(L) = \mathrm{I}_n + \tilde{\mathbf{\Theta}}_1 L + \cdots + \tilde{\mathbf{\Theta}}_q L^q$ are matrix polynomials in the lag operator L that satisfy

$$\det \mathbf{B}(z) \neq 0 \text{ for all } |z| < 1, \qquad \det \tilde{\mathbf{\Theta}}(z) \neq 0 \text{ for all } |z| \leq 1,$$

along with the restriction $1 \leq \tilde{\mathbf{\Theta}}_q \leq n - 1$. Observe that the VARMA process in (B.1) is *fundamental*, and as shown next, the structural VARMA(p, q) of interest can be derived as a *basic non-fundamental* representation of this process (see Lippi and Reichlin (1994) for a precise definition and further discussion).

[22] For clarity of exposition, we suppress all deterministic terms as they do not affect any of the computations in this section. In practice, we include an intercept in all estimated models.

To see how the structural representation is obtained from (B.1), consider the matrix polynomial:

$$\mathbf{C}(L) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{\tilde{c}}{\sqrt{1+\tilde{c}^2}}L & -\frac{1}{\sqrt{1+\tilde{c}^2}}L & 0 \\ 0 & \frac{1}{\sqrt{1+\tilde{c}^2}} & \frac{\tilde{c}}{\sqrt{1+\tilde{c}^2}} & 0 \\ 0 & 0 & 0 & \mathrm{I}_{n-3} \end{pmatrix},$$

$$\mathbf{C}(L^{-1})' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{\tilde{c}}{\sqrt{1+\tilde{c}^2}}L^{-1} & \frac{1}{\sqrt{1+\tilde{c}^2}} & 0 \\ 0 & -\frac{1}{\sqrt{1+\tilde{c}^2}}L^{-1} & \frac{\tilde{c}}{\sqrt{1+\tilde{c}^2}} & 0 \\ 0 & 0 & 0 & \mathrm{I}_{n-3} \end{pmatrix},$$

where L^{-1} is the forward operator, forms the *inverse* of $\mathbf{C}(L)$ in the sense that $\mathbf{C}(L)\mathbf{C}(L^{-1})' = \mathrm{I}_n$. Hence, $\mathbf{C}(L)$ is a *Blaschke matrix* with the well known property that for any n-dimensional orthogonal white noise $\mathbf{u}_t$ (i.e. with $\mathrm{Var}(\mathbf{u}_t) = \mathrm{I}_n$), the operation $\tilde{\mathbf{u}}_t = \mathbf{C}(L^{-1})'\mathbf{u}_t$ also yields an orthogonal white noise process $\tilde{\mathbf{u}}_t$. It is convenient to view Blaschke matrices as *dynamic* generalizations of orthogonal matrices.

Consequently, let $\mathbf{\Theta}_0\mathbf{\Theta}_0' = \mathbf{\Sigma}$ and define

$$\begin{aligned} \mathbf{A}(L) &= \tilde{\mathbf{\Theta}}(L)\mathbf{\Theta}_0\mathbf{C}(L)\mathbf{\Gamma}, \\ \boldsymbol{\epsilon}_t &= \mathbf{\Gamma}'\mathbf{C}(L^{-1})'\mathbf{\Theta}_0^{-1}\tilde{\boldsymbol{\epsilon}}_t, \end{aligned}$$

for some orthogonal matrix matrix $\mathbf{\Gamma}$ satisfying $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathrm{I}_n$. It can be seen that $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathrm{I}_n)$ and the VARMA(p, q)

$$\mathbf{B}(L)\mathbf{y}_t = \mathbf{A}(L)\boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathrm{I}_n), \tag{B.2}$$

is observationally equivalent to (B.1).

To ensure that we recover the structural shocks of interest in $\boldsymbol{\epsilon}_t$, it is necessary to employ a suitable choice of $\mathbf{\Theta}_0$ and $\mathbf{\Gamma}$. To this end, suppose $\mathbf{\Theta}_0$ is chosen such that the second column of $\mathbf{\Theta}_q = \tilde{\mathbf{\Theta}}_q\mathbf{\Theta}_0$ is zero, where such a $\mathbf{\Theta}_0$ exists if and only if $\tilde{\mathbf{\Theta}}_q \leq n-1$. Then multiplying

$$\begin{aligned} \tilde{\mathbf{\Theta}}(L)\mathbf{\Theta}_0\mathbf{C}(L) &= \mathbf{\Theta}(L)\mathbf{C}(L), \\ &= (\mathbf{\Theta}_0 + \mathbf{\Theta}_1 L + \cdots + \mathbf{\Theta}_q L^q)(\mathbf{C}_0 + \mathbf{C}_1 L), \\ &= \mathbf{\Theta}_0\mathbf{C}_0 + (\mathbf{\Theta}_0\mathbf{C}_1 + \mathbf{\Theta}_1\mathbf{C}_0)L + \cdots + (\mathbf{\Theta}_{q-1}\mathbf{C}_1 + \mathbf{\Theta}_q\mathbf{C}_0)L^q + \mathbf{\Theta}_q\mathbf{C}_1 L^{q+1}. \end{aligned}$$

Note that since $\mathbf{C}_0 = n-1$ and it has the third column proportional to the second column (by a factor $\tilde{c}$), while $\mathbf{C}_1$ only has non-zero entries in the second row, we obtain $\mathbf{\Theta}_q\mathbf{C}_1 = 0$ and $(\mathbf{\Theta}_0\mathbf{C}_0) = n-1$ with the third column of $\mathbf{\Theta}_0\mathbf{C}_0$ proportional to the second column (by a factor $\tilde{c}$). Moreover, $(\mathbf{\Theta}_{q-1}\mathbf{C}_1 + \mathbf{\Theta}_q\mathbf{C}_0) = \mathbf{\Theta}_q + 1$, except in the special case where the second column of $\mathbf{\Theta}_{q-1}$ is either zero or lies in the column space of $\mathbf{\Theta}_q$ (then, $(\mathbf{\Theta}_{q-1}\mathbf{C}_1 + \mathbf{\Theta}_q\mathbf{C}_0) = \mathbf{\Theta}_q$).

Consequently, the resulting VMA representation $\tilde{\mathbf{A}}(L) = \mathbf{\Theta}(L)\mathbf{C}(L)$ is of the same order q as $\tilde{\mathbf{\Theta}}(L)$, but is *not fundamental* since it contains exactly one root inside the unit circle.[23] However, it is related to the structural VMA representation of interest by (constant) orthogonal rotations $\mathbf{\Gamma}$, which are chosen to satisfy the identifying restrictions in Section 3 in similar fashion to the VAR case.

Based on this, our approach to estimating the structural VARMA involves the following steps:

1. Estimate the reduced-form, fundamental VARMA(p, q) in (B.1) subject to the restriction $1 \leq \tilde{\mathbf{\Theta}}_q \leq n-1$. We note that a wide range of methods exist to estimate fundamental VARMA systems, and in principle, any such method can be employed here. We use Bayesian methods (described in detail below) because they are particularly well suited for working with large systems of the type we focus on in our applications (i.e. with $n \geq 8$). Moreover, Bayesian methods provide a convenient framework for imposing non-linear over-identifying restrictions, such as the "dampening" restrictions that we analyze in our empirical work.

   It is worth emphasizing, however, that while in theory the VARMA specified in (B.1) admits a VAR(∞) representation, it is not possible to estimate a *truncated* VAR to implement this step because any truncation of the VAR would make it incompatible with the rank restriction on $\tilde{\mathbf{\Theta}}_q$. In other words, a finite-order VAR cannot be inverted to recover a VARMA with $1 \leq \tilde{\mathbf{\Theta}}_q \leq n-1$, which is necessary to obtain the structural VARMA representation, as described above.

2. Decompose $\mathbf{\Theta}_0\mathbf{\Theta}_0' = \mathbf{\Sigma}$ and set $\mathbf{\Theta}_j = \tilde{\mathbf{\Theta}}_j\mathbf{\Theta}_0$ for $j = 1, \ldots, q$ such that the 2nd column of $\mathbf{\Theta}_q = 0$. One practical way to do this is to first set $\tilde{\mathbf{\Theta}}_0$ to be the Cholesky factor of $\mathbf{\Sigma}$ and then construct the orthogonal matrix $\mathbf{\Gamma}_0$ with:

   - the second column $\mathbf{\Gamma}_{0,2}$ set to an $n \times 1$ vector in the null space of $\tilde{\mathbf{\Theta}}_q\tilde{\mathbf{\Theta}}_0$, normalized such that $\|\mathbf{\Gamma}_{0,2}\| = 1$, and
   - and the remaining columns $\mathbf{\Gamma}_{0,1}, \mathbf{\Gamma}_{0,3}, \ldots, \mathbf{\Gamma}_{0,n}$ set to the $n-1$ vectors orthogonal to $\mathbf{\Gamma}_{0,2}$ and normalized such that $\|\mathbf{\Gamma}_{0,i}\| = 1$.

   Consequently, setting $\mathbf{\Theta}_0 = \tilde{\mathbf{\Theta}}_0\mathbf{\Gamma}_0$ preserves the property $\mathbf{\Theta}_0\mathbf{\Theta}_0' = \mathbf{\Sigma}$ while ensuring that the second column of $\mathbf{\Theta}_q = \tilde{\mathbf{\Theta}}_q\mathbf{\Theta}_0 = \tilde{\mathbf{\Theta}}_q\tilde{\mathbf{\Theta}}_0\mathbf{\Gamma}_0$ is zero.

[23] Clearly, other structural non-fundamental representations can be constructed by "flipping" additional roots of the determinant of the fundamental $\tilde{\mathbf{\Theta}}(L)$ (Lippi and Reichlin (1994)). Motivated by the goal to work with a model that entails a minimum necessary deviation from the SVAR, we only consider structural $\mathbf{A}(L)$ that have all roots of the determinant outside the unit circle, aside from the one necessary to impose $\mathbf{A}_0 = n-1$.

3. Arbitrarily set $\tilde{c} = 1$ and compute some $\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_q$ satisfying $\tilde{\mathbf{A}}_0 = n - 1$ through linear transformations of $\mathbf{\Theta}_0, \ldots, \mathbf{\Theta}_q$, namely

$$\begin{aligned}\tilde{\mathbf{A}}_0 &= \mathbf{\Theta}_0 \mathbf{C}_0, \\ \tilde{\mathbf{A}}_j &= \mathbf{\Theta}_{j-1} \mathbf{C}_1 + \mathbf{\Theta}_j \mathbf{C}_0, \qquad j = 1, \ldots, q.\end{aligned}$$

4. Obtain the structural $\mathbf{A}_0, \ldots, \mathbf{A}_q$ that satisfy the identifying restrictions by applying a series of (constant) orthogonal rotations to $\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_q$ as in typical VAR settings. For example, in our 'Identification scheme I' (outlined in Section 3 of the text and reproduced here for convenience), we assume $n \geq 4$, $\epsilon_{1,t}$ is non-news, $\epsilon_{2,t}$ is news, $\epsilon_{3,t}$ is noise, and we follow Barsky and Sims (2011) in imposing the restrictions:

   (a) the non-news shock is identified as the reduced-form innovation to TFP, such that it is the only shock affecting TFP on impact, i.e. the first *row* in $\mathbf{A}_0$ has zero entries everywhere except the first element ($A_{0,1,1} \neq 0$ and $A_{0,1,i} = 0$ for all $i = 2, \ldots, n$);

   (b) 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 (which we will take to be 20 years ahead).

   However, we also add the restriction that allows us to disentangle news from noise shocks:

   (c) news and noise have no immediate impact on TFP and, for the other variables, the IRFs generated by news and noise shocks on impact are proportional to each other, i.e. the third column in $\mathbf{A}_0$ is proportional to the second column ($\mathbf{A}_{0,3} = c\mathbf{A}_{0,2}$).

   Implementing these restrictions involves applying three types of orthogonal rotations, $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, and $\mathbf{\Gamma}_3$, such that their product $\mathbf{\Gamma} = \mathbf{\Gamma}_1\mathbf{\Gamma}_2\mathbf{\Gamma}_3$ yields the comprehensive set of orthogonal rotations that transform $\tilde{\mathbf{A}}(L)$ into the structural representation of interest $\mathbf{A}(L)$, where the two VMA polynomials are related by $\mathbf{A}_j = \tilde{\mathbf{A}}_j\mathbf{\Gamma}$ for $j = 0, \ldots, q$.

   Specifically, $\mathbf{\Gamma}_1$ is determined by setting the first column $\mathbf{\Gamma}_{0,1} = \tilde{\mathbf{A}}'_{0,(1)}/\|\tilde{\mathbf{A}}_{0,(1)}\|^2$, where $\tilde{\mathbf{A}}_{0,(1)}$ denotes the first row of $\tilde{\mathbf{A}}_0$, and the remaining columns $\mathbf{\Gamma}_{0,i}$ for $i = 2, \ldots, n$ equal to the $n-1$ vectors that are orthogonal to $\mathbf{A}'_{0,(1)}$ (normalized such that $\|\mathbf{\Gamma}_{0,i}\| = 1$).

   Next, let $\tilde{\mathbf{K}}(L) = \mathbf{B}(L)^{-1}\tilde{\mathbf{A}}(L)\mathbf{\Gamma}_1$ be the impulse responses obtained after applying the first set of orthogonal rotations $\mathbf{\Gamma}_1$, and define $\tilde{\mathbf{K}}_{j,1,2:n}$ for $j \geq 0$ as the $1 \times n - 1$ row vector constructed from the first row and columns 2 to

n of $\tilde{\mathbf{K}}_j$. Compute the eigenvalue decomposition of $\sum_{j=1}^{20} \tilde{\mathbf{K}}'_{j,1,2:n}\tilde{\mathbf{K}}_{j,1,2:n}$, with eigenvalues sorted in *descending* order, and store the eigenvectors in $\mathbf{\Delta}_2$. The orthogonal matrix that identified news shocks according to restriction 4b above is then given by

$$\mathbf{\Gamma}_2 = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Delta}_2 \end{pmatrix}.$$

We normalize the sign of the news shock by requiring that the maximum impulse response (over the horizon $0:20$) of TPF to news is positive.

Let $\check{\mathbf{A}}(L) = \tilde{\mathbf{A}}(L)\mathbf{\Gamma}_1\mathbf{\Gamma}_2$ be the VMA representation obtained after applying the first two sets of orthogonal rotations. At this stage, non-news and news shocks are identified according to restrictions 4a and 4b, but the noise shock is not identified in the sense that following multiplication by $\mathbf{\Gamma}_1\mathbf{\Gamma}_2$, the third column of $\check{\mathbf{A}}_0$ will generally *not* be proportional to the second. To enforce the proportionality restriction, we construct a third orthogonal matrix

$$\mathbf{\Gamma}_3 = \begin{pmatrix} \mathrm{I}_2 & 0 \\ 0 & \mathbf{\Delta}_3 \end{pmatrix},$$

where the first column $\mathbf{\Delta}_{3,1}$ of the $n-2 \times n-2$ orthogonal matrix $\mathbf{\Delta}_3$ must satisfy $(\check{\mathbf{A}}_{0,3}, \ldots, \check{\mathbf{A}}_{0,n})\mathbf{\Delta}_{3,1} = c\check{\mathbf{A}}_{0,2}$.

By construction[24], the $n \times n-1$ matrix $(\check{\mathbf{A}}_{0,2}, \ldots, \check{\mathbf{A}}_{0,n})$ has rank $n-2$ and, therefore, there exists a $n-1 \times 1$ vector $\mathbf{z} = (z_1, \mathbf{z}_2')'$, $\|\mathbf{z}\| = 1$ such that

$$(\check{\mathbf{A}}_{0,2}, \ldots, \check{\mathbf{A}}_{0,n})\mathbf{z} = 0$$

(i.e. $\mathbf{z}$ is the orthonormal basis for the null space of $(\check{\mathbf{A}}_{0,2}, \ldots, \check{\mathbf{A}}_{0,n})$). Accordingly, set

$$c = \frac{|z_1|}{1 - z_1^2}$$
$$\mathbf{\Delta}_{3,1} = -(z_1)\frac{\mathbf{z}_2}{1 - z_1^2},$$

and the remaining columns $\mathbf{\Delta}_{3,2}, \ldots, \mathbf{\Delta}_{3,n-2}$ of $\mathbf{\Delta}_3$ to be the $n-3$ vectors orthogonal to $\mathbf{\Delta}_{3,1}$ (normalized such that $\|\mathbf{\Delta}_{3,i}\| = 1$). Subsequently, multiplying $\mathbf{A}_j = \check{\mathbf{A}}_j\mathbf{\Gamma}_3$ for all $j = 0, \ldots, q$ yields the desired representation $\mathbf{A}(L)$ where $\mathbf{A}_{0,3} = c\mathbf{A}_{0,2}$ while preserving the restrictions 4a and 4b.

In 'Identification Scheme II', we modify assumption 4a to be:

(a) at $t = 0$, TFP is impacted upon by two disturbances, the non-news shock, and a transitory TFP shock which is disentangled from the non-news shock because it explains the minimum fraction of the FEV of TFP at a specific long horizon (again, 20 years ahead).

[24] Recall that $\tilde{\mathbf{A}}_0$ has proportional second and third columns, transformation by $\mathbf{\Gamma}_1$ preserves the linear independence of the first column and transformation by $\mathbf{\Gamma}_2$ only alters columns 2 to n.

Therefore, Scheme II only differs from Scheme I in that it requires $n \geq 5$ and allows one other shock (besides non-news) to impact TFP at $t = 0$. Assume for notational convenience that this transitory TFP shock is ordered last (i.e. $\epsilon_{n,t}$), and let the constant orthogonal rotations that need to be applied to $\tilde{\mathbf{A}}(L)$ under Scheme II be given by $\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}_1 \boldsymbol{\Upsilon}_2$.

We first construct $\boldsymbol{\Upsilon}_1$ to identify the transitory TFP shock. Hence, let $\tilde{\mathbf{K}}(L) = \mathbf{B}(L)^{-1}\tilde{\mathbf{A}}(L)$ be the impulse responses obtained based on $\tilde{\mathbf{A}}(L)$, and define $\tilde{\mathbf{K}}_{j,(1)}$ for $j \geq 0$ as the first row of $\tilde{\mathbf{K}}_j$. Compute the eigenvalue decomposition of $\sum_{j=1}^{20} \tilde{\mathbf{K}}'_{j,(1)}\tilde{\mathbf{K}}_{j,(1)}$, with eigenvalues sorted in *descending* order, and store the eigenvectors in $\boldsymbol{\Upsilon}_1$.

Finally, let

$$\boldsymbol{\Upsilon}_2 = \begin{pmatrix} \boldsymbol{\Gamma} & \\ & 1 \end{pmatrix},$$

where $\boldsymbol{\Gamma}$ is now $n-1 \times n-1$. To construct $\boldsymbol{\Gamma}$, we proceed in nearly identical fashion to the procedure described above for Scheme I, except taking into account only the first $n-1$ shocks $\epsilon_{1,t}, \ldots, \epsilon_{n-1,t}$ in all computations.

B.2 Bayesian Algorithms

The ultimate goal of a Bayesian approach to estimating noise shocks is to obtain draws from the posterior distribution of the Wold representation

$$\mathbf{y}_t = \mathbf{K}(L)\boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathrm{I}_n), \tag{B.3}$$

where $\mathbf{K}(L) = \mathbf{B}(L)^{-1}\mathbf{A}(L)$. Indeed, the Bayesian framework offers a great deal of flexibility in designing sampling algorithms for this purpose. For example, Plagborg-Møller (2016) develops an MCMC algorithm that samples directly from a truncated approximation to (B.3). Such an approach is suitable when working with stationary data and has the advantage of allowing restrictions on impulse responses to be imposed directly in the sampling.

Since in our applications we wish to use data on the log-levels of our variables and allow for possible co-integration, a truncated approximation to the Wold representation is not appropriate, and it is necessary to work with a finite order VARMA representation, such as the VARMA(p, q) specified in (B.1). A well known feature of VARMAs, however, is that the parameters of $\mathbf{B}(L)$ and $\tilde{\boldsymbol{\Theta}}(L)$ will generally not be identified without further restrictions. The same is true for the structural representation (B.2). The reason for that is that, even though $\mathbf{K}(L) = \mathbf{B}(L)^{-1}\mathbf{A}(L)$ is uniquely determined by identifying restrictions such as 4a-4 above, such restrictions do not guarantee uniqueness of $\mathbf{B}(L)$ and $\mathbf{A}(L)$ since there may exist some $\mathbf{D}(L)$ such that $\mathbf{B}(L)^\dagger = \mathbf{D}(L)\mathbf{B}(L)$ is of order p, $\mathbf{A}(L)^\dagger = \mathbf{D}(L)\mathbf{A}(L)$ is of order q, and both lead to the same Wold representation $\mathbf{K}(L) = (\mathbf{B}(L)^\dagger)^{-1}\mathbf{A}(L)^\dagger$.

Identification issues in VARMAs are further complicated by the fact that representations such as (B.1) and (B.2) are observationally equivalent. Consequently, when a unique VARMA representation is required for estimation purposes, it is typically specified as a *fundamental* process in the canonical *echelon* form. This unique specification is derived by starting with:

$$\tilde{\mathbf{B}}_0\mathbf{y}_t = \tilde{\mathbf{B}}_1\mathbf{y}_{t-1} + \cdots + \tilde{\mathbf{B}}_{p^*}\mathbf{y}_{t-p^*} + \tilde{\mathbf{B}}_0\mathbf{u}_t + \mathbf{M}_1\mathbf{u}_{t-1} + \cdots + \mathbf{M}_{p^*}\mathbf{u}_{t-p^*}, \qquad \mathbf{u}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \tag{B.4}$$

where $\tilde{\mathbf{B}}_0$ is lower triangular with ones on the diagonal, which we refer to as the *semi-structural* VARMA form. Then, two types of restrictions are imposed on this representation to ensure uniqueness:

1. exclusion restrictions on $\tilde{\mathbf{B}}_0, \ldots, \tilde{\mathbf{B}}_{p^*}$, $\mathbf{M}_1, \ldots, \mathbf{M}_{p^*}$ according to the *row degrees* $p_1, \ldots, p_n$ that define the lag structure of each equation in the system (with $p^* = \max(p_1, \ldots, p_n)$);

2. non-linear restrictions on $\mathbf{M}_1, \ldots, \mathbf{M}_{p^*}$ to ensure all roots of $\mathbf{M}(L)$ lie outside the unit circle.

When these restrictions hold, the semi-structural VARMA is said to be in *echelon* form.

Note that the coefficients in (B.1) are related to the semi-structural VARMA by:

$$\mathbf{B}_j = \tilde{\mathbf{B}}_0^{-1}\tilde{\mathbf{B}}_j \qquad\qquad \tilde{\boldsymbol{\Theta}}_j = \tilde{\mathbf{B}}_0^{-1}\mathbf{M}_j.$$

However, estimating a VARMA in the *echelon* canonical form is challenging. First, imposing type 2 restrictions on the roots of $\mathbf{M}(L)$ becomes exceedingly difficult as the size of the system increases. Moreover, imposing type 1 exclusion restrictions requires knowledge of the row degrees $p_1, \ldots, p_n$, which themselves need to be estimated in practice.[25]

Fortunately, point identification is not necessary in the Bayesian framework. That is, the posterior distribution will be well-defined even when the likelihood does not uniquely identify the parameters in the model, as long as proper prior distributions are specified for the parameters. In the VARMA case, this means that as long as proper priors are specified for $\mathbf{B}_1, \ldots, \mathbf{B}_p$, $\tilde{\boldsymbol{\Theta}}_1, \ldots, \tilde{\boldsymbol{\Theta}}_q$, and $\boldsymbol{\Sigma}$, we can readily obtain draws from

$$p(\mathbf{B}_1, \ldots, \mathbf{B}_p, \tilde{\boldsymbol{\Theta}}_1, \ldots, \tilde{\boldsymbol{\Theta}}_q, \boldsymbol{\Sigma} \,|\, \mathbf{y}),$$

even though this posterior may not be characterized by a *unique* mode, or may simply resemble the joint prior distribution (in the extreme case where the likelihood provides no information on the model parameters).

[25] Note that we only provide a brief summary of the identification issues and classical methods designed to deal with them in estimating VARMA systems. An in-depth discussion is beyond the scope of this paper, and we refer the interested reader to Lütkepohl (2005) for a textbook treatment, including further details and explicit formulae.

The key insight in a Bayesian approach to analyzing VARMA models is that parameters $\mathbf{B}_1, \ldots, \mathbf{B}_p$, $\tilde{\mathbf{\Theta}}_1, \ldots, \tilde{\mathbf{\Theta}}_q$, and $\mathbf{\Sigma}$ themselves are not of primary interest, but rather quantities such as forecasts and impulses responses, which are uniquely identified even when the AR and MA coefficients are not. Therefore, it is possible to obtain draws from the posterior of unidentified parameters, then transform them to draws from the posterior of quantities which are, in fact, identified.

In general, Bayesians routinely build sampling algorithms on unidentified parameter spaces to obtain computational efficiency (examples include Gustafson (2005); Imai and van Dyk (2005); Ghosh and Dunson (2009); Koop, León-González, and Strachan (2010); Koop, León-González, and Strachan (2012), among many others). Indeed, early work such as Meng and van Dyk (1999) and Liu and Wu (1999) suggest that artificially expanding the parameter space may reduce auto-correlation in Markov Chain Monte Carlo (MCMC) sampling algorithms, in terms of the identified quantities of interest, thus further improving computation. Nevertheless, identification is an important concept in the Bayesian framework to the extent that it provides parsimony in over-parameterized systems. From a practical viewpoint, both parsimony and identification are features of the model that are implemented entirely through the appropriate specification of prior distributions.

Building on these ideas, Chan and Eisenstat (2015) and Chan, Eisenstat, and Koop (2016) develop MCMC algorithms on the *expanded* VARMA representation:

$$\tilde{\mathbf{B}}_0 \mathbf{y}_t = \tilde{\mathbf{B}}_1 \mathbf{y}_{t-1} + \cdots + \tilde{\mathbf{B}}_{p^*} \mathbf{y}_{t-p^*} + \mathbf{\Phi}_0 \mathbf{f}_t + \mathbf{\Phi}_1 \mathbf{f}_{t-1} + \cdots + \mathbf{\Phi}_{p^*} \mathbf{f}_{t-p^*} + \boldsymbol{\eta}_t, \qquad \text{(B.5)}$$

where $\mathbf{f}_t \sim \mathcal{N}(0, \mathbf{\Omega})$, $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \mathbf{\Lambda})$, $\mathbf{\Omega}$ and $\mathbf{\Lambda}$ are diagonal, and $\mathbf{\Phi}_0$ is lower triangular with ones on the diagonal. Expanded form parameters are related to the VARMA parameters in (B.4) by the mapping:

$$\sum_{l=j}^{p^*} \tilde{\mathbf{\Theta}}_l \mathbf{\Sigma} \tilde{\mathbf{\Theta}}_{l-j}' = \sum_{l=j}^{p^*} \mathbf{\Phi}_l \mathbf{\Omega} \mathbf{\Phi}_{l-j}' + \mathbb{1}(j=0)\mathbf{\Lambda}, \quad \text{for all } j = 0, \ldots, p^*, \qquad \text{(B.6)}$$

whereas $\tilde{\mathbf{B}}_j$ in the expanded form is identical to the corresponding $\tilde{\mathbf{B}}_j$ in the semi-structural form for $j = 0, \ldots, p^*$. Consequently, draws from (B.4) can be obtained by sampling directly from the expanded form (B.5) and then computing $\mathbf{M}_1, \ldots, \mathbf{M}_{p^*}$, $\mathbf{\Sigma}$ from each draw of $\mathbf{\Phi}_0, \ldots, \mathbf{\Phi}_{p^*}$, $\mathbf{\Omega}$, and $\mathbf{\Lambda}$ using the mapping in (B.6). The exact procedure based on *generalized eigenvalues* is provided in Section 2 of Chan and Eisenstat (2015) and Appendix D of Chan, Eisenstat, and Koop (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.

The advantage of the expanded form is that it can be regarded as a *linear state space* model, and therefore, admits straightforward and efficient MCMC sampling algorithms. Moreover, there is no need to impose non-linear restrictions directly in the MCMC since restrictions on the roots of $\mathbf{M}(L)$ can be easily implemented in the

post-processing of draws (i.e. when constructing $\mathbf{M}_1, \ldots, \mathbf{M}_{p^*}, \boldsymbol{\Sigma}$ from $\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_{p^*}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Lambda}$).

At the same time, it provides an extremely flexible approach to estimating VARMAs. For example, Chan, Eisenstat, and Koop (2016) demonstrate how to construct a prior on the expanded form parameters—using *stochastic search variable selection* (SSVS) methods (see Kuo and Mallick (1997); George, Sun, and Ni (2008))—such that the implied draws from the semi-structural from (B.4) satisfy the echelon form restrictions at every iteration. Hence, the expanded form can be used to estimate *unique* VARMA systems, although this may still lead to computationally intensive algorithms in larger VARMAs. On the other hand, it is also possible to obtain more computational efficiency by employing priors that *approximate* the echelon form in the sense that they lead to exact identifying restrictions holding with some probability (less than one) in the posterior. The Bayesian approach based on the expanded form, therefore, affords a great deal of flexibility in designing algorithms that target an optimal balance between computational efficiency and parsimony.

In this paper, we employ such an *approximate identification* approach. In particular, starting from the expanded form (B.5) and assuming $p = p^*$, $q < p$, we impose parsimony by first setting (with probability one) $\boldsymbol{\Phi}_j = 0$ for all $j = q+1, \ldots, p^*$ and

$$\boldsymbol{\Phi}_{j,3} = \cdots = \boldsymbol{\Phi}_{j,n} = 0 \qquad \text{for all } j = 1, \ldots, q,$$

where $\boldsymbol{\Phi}_{j,k}$ denotes the kth column of $\boldsymbol{\Phi}_j$. This leads to the restrictions $\tilde{\boldsymbol{\Theta}}_j = 0$ for all $j = q+1, \ldots, p^*$ and $\tilde{\boldsymbol{\Theta}}_q = 2$. Next, we specify SSVS priors on the individual free elements of $\tilde{\mathbf{B}}_0, \ldots, \tilde{\mathbf{B}}_p$ and $\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_q$ of the form:

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

Through extensive experimentation with the resulting algorithm, we find 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 $\mathcal{IG}(a,b)$ 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 $\mathbf{y}_t$ 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 effect of the standardization. Hence, all impulse responses are reported on the original, unscaled variables. The approach is equivalent to working directly with $\mathbf{y}_t$, 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 $\left(\boldsymbol{\gamma}_i, \mathbf{B}_{(i)}, \boldsymbol{\Phi}_{(i)} \,|\, \mathbf{f}, \Lambda_{ii}, \mathbf{y}_i\right)$ for each $i = 1, \ldots, n$, where $\mathbf{B}_{(i)}$ denotes the i-th row of $\mathbf{B} = (\mathbf{I}_n - \mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_p)$, $\boldsymbol{\Phi}_{(i)}$ the i-th row of $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_q)$, and $\boldsymbol{\gamma}_i$ is the set of all SSVS indicators pertaining to $\mathbf{B}_{(i)}, \boldsymbol{\Phi}_{(i)}$.

2. Sample $\left(\Lambda_{ii} \,|\, \mathbf{B}_{(i)}, \boldsymbol{\Phi}_{(i)}, \boldsymbol{\gamma}_i, \mathbf{f}, \mathbf{y}_i\right)$ for each $i = 1, \ldots, n$.

3. Sample $(\Omega_{ii} \,|\, \mathbf{f}_i)$ for each $i = 1, \ldots, n$.

4. Sample $(\mathbf{f} \,|\, \mathbf{B}, \boldsymbol{\Phi}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\gamma}, \mathbf{y})$.

Details and extensive discussion of each sampling step above are provided in Appendix B of Chan, Eisenstat, and Koop (2016).

In summary, we obtain posterior draws from the impulses responses $\mathbf{K}(L)$ identified by the structural model as follows:

1. Obtain draws of $\tilde{\mathbf{B}}_0, \ldots, \tilde{\mathbf{B}}_p$, $\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_q$, $\boldsymbol{\Omega}$, $\boldsymbol{\Lambda}$ using the Gibbs sampling algorithm outlined above.

2. For each draw of the expanded form parameters, transform to draws of $\mathbf{B}_1, \ldots, \mathbf{B}_p$, $\tilde{\boldsymbol{\Theta}}_1, \ldots, \tilde{\boldsymbol{\Theta}}_q$ and $\boldsymbol{\Sigma}$.

3. For each draw of $\tilde{\boldsymbol{\Theta}}_1, \ldots, \tilde{\boldsymbol{\Theta}}_q$ and $\boldsymbol{\Sigma}$, transform to draws of $\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_q$ by applying the Blaschke matrix transformation.

4. For each draw of $\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_q$, transform to draws of $\mathbf{A}_0, \ldots, \mathbf{A}_q$ by applying appropriate orthogonal rotations.

5. Fore each draw of $\mathbf{B}_1, \ldots, \mathbf{B}_p$, $\mathbf{A}_0, \ldots, \mathbf{A}_q$ compute $\mathbf{K}(L) = \mathbf{B}(L)^{-1}\mathbf{A}(L)$ 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).

C Barsky and Sims' (2011) RBC model augmented with noise shocks about future TFP

C.1 The first-order conditions

The first-order conditions with respect to C_t, I_t, N_t, and K_{t+1} are given by

$$\Sigma_t^N N_t^{\theta+1/\eta} = \mu_t A_t (1-\theta) K_t^{\theta} \tag{C.1}$$

$$\mu_t = (C_t - bC_{t-1})^{-1} - b\beta (C_{t+1} - bC_t)^{-1} \tag{C.2}$$

$$\lambda_t = \beta E_t \left[(1-\delta)\lambda_{t+1|t} + \theta \mu_{t+1} A_{t+1} N_t^{1-\theta} K_t^{\theta-1} \right] \tag{C.3}$$

$$\mu_t = \lambda_t \left\{ \left[1 - \frac{\gamma}{2} \left(\frac{I_t}{I_{t-1}} - \tilde{g}_I \right)^2 \right] - \gamma \frac{I_t}{I_{t-1}} \left(\frac{I_t}{I_{t-1}} - \tilde{g}_I \right) \right\} + \\ + \beta E_t \left[\gamma \lambda_{t+1} \frac{I_t^2}{I_{t-1}^2} \left(\frac{I_t}{I_{t-1}} - \tilde{g}_I \right) \right] \tag{C.4}$$

where λ_t and μ_t are two Lagrange multipliers.

C.2 The process for TFP

The process for $a_t = \ln(A_t)$ is given by

$$a_t = \tilde{a}_t + v_t \tag{C.5}$$

where $\tilde{a}_t$ is the unobserved permanent component of productivity and v_t is a white noise disturbance, $v_t \sim WN(0, \sigma_v^2)$. The permanent component of a_t evolves according to

$$\tilde{a}_t = \tilde{a}_{t-1} + \epsilon_t^{NN} + \epsilon_{t-1}^{NE} \tag{C.6}$$

where, once again, ϵ_t^{NN} and ϵ_t^{NE} are a non-news and a news shock, respectively. We consider a 1-period 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)—that is: $s_t = \epsilon_t^{NE} + u_t$—with u_t being once again $WN(0, \sigma_u^2)$.

C.3 The agents' signal-extraction problem about TFP

By defining $\xi_t = [\Delta\tilde{a}_t, \epsilon_t^{NE}, v_t, v_{t-1}]'$ and $S_t = [\Delta a_t, s_t]'$, the model (5), (C.5), and (C.6) can be put into state-space form, with state equation

$$\begin{bmatrix} \Delta\tilde{a}_t \\ \epsilon_t^{NE} \\ v_t \\ v_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{A} \begin{bmatrix} \Delta\tilde{a}_{t-1} \\ \epsilon_{t-1}^{NE} \\ v_{t-1} \\ v_{t-2} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{B} \begin{bmatrix} \epsilon_t^{NN} \\ \epsilon_t^{NE} \\ v_t \\ u_t \end{bmatrix} \tag{C.7}$$

and observation equation

$$\begin{bmatrix} \Delta a_t \\ s_t \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{C} \begin{bmatrix} \Delta\tilde{a}_t \\ \epsilon_t^{NE} \\ v_t \\ v_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{D} \begin{bmatrix} \epsilon_t^{NN} \\ \epsilon_t^{NE} \\ v_t \\ u_t \end{bmatrix} \tag{C.8}$$

The solution to the agents' signal-extraction problem is still given by expressions (A.3)-(A.4).

C.4 Stationarizing the model's variables

We stationarize all variables except hours as in Barsky and Sims (2011).[26] Specifically, defining $\Gamma_t \equiv A_t^{\frac{1}{1-\theta}}$, we stationarize output, consumption, investment, and the capital stock as $X_t^* \equiv X_t/\Gamma_t$, with $X = Y$, C, I, and $K_t^* \equiv K_t/\Gamma_{t-1}$;[27] and we stationarize the two Lagrange multipliers, λ_t and μ_t, as $\lambda_t^* \equiv \lambda_t \cdot \Gamma_t$ and $\mu_t^* \equiv \mu_t \cdot \Gamma_t$. Then, $\hat{\lambda}_t^*$ is the log-deviation from the steady-state of λ_t^*, $\hat{y}_t^*$ is the log-deviation from the steady-state of Y_t^*, and so on.

C.5 The log-linearized equations for the stationarized variables

Log-linearizing the model's transformed equations for the stationarized variables we obtain the following expressions:

$$\hat{\mu}_t^* + \hat{c}_t^*[S_C\Delta_G + (1-S_C)(1-\Delta_G)] + (1-S_C)\Delta_G\hat{c}_{t+1|t}^* +$$

$$+S_C(1-\Delta_G)\hat{c}_{t-1}^* + S_C\frac{1-\Delta_G}{1-\theta}\Delta a_t - \frac{(1-S_C)\Delta_G}{1-\theta}\Delta a_{t+1|t} - \epsilon_t^c = 0 \tag{C.9}$$

$$-\hat{\mu}_t^* + \left(\theta + \frac{1}{\eta}\right)\hat{n}_t + \frac{\theta}{1-\theta}\Delta a_t - \theta\hat{k}_t^* - \epsilon_t^n = 0 \tag{C.10}$$

$$-\hat{\lambda}_t^* + \beta S_K\hat{\mu}_{t+1|t}^* + \beta S_K^*\hat{y}_{t+1|t}^* - \beta S_K\hat{k}_{t+1|t}^* + \beta(1-S_K)\hat{\lambda}_{t+1|t}^* - \beta\frac{(1-S_K)}{1-\theta}\epsilon_t^{NE} = 0 \tag{C.11}$$

$$\hat{y}_t^* - \frac{\alpha_C\hat{c}_t^* + \alpha_I\hat{\imath}_t^* + (1-\alpha_C-\alpha_I)\epsilon_t^g}{\alpha_C+\alpha_I} = 0 \tag{C.12}$$

$$\hat{k}_{t+1}^* + \frac{\alpha_K + \gamma(1-\alpha_K)}{1-\theta}\Delta a_t + (1-\alpha_K)(\gamma-1)\hat{\imath}_t^* - \alpha_K\hat{k}_t^* - \gamma(1-\alpha_K)\hat{\imath}_{t-1}^* = 0 \tag{C.13}$$

[26] We wish to thank Eric Sims for providing extensive details about the solution to their original model.

[27] The capital stock is divided by Γ_{t-1}, rather than by Γ_t, in order to make sure that the stationarized capital stock, K_t^*, is still predetermined at time t.

$$\hat{y}_t^* + \frac{\theta}{1-\theta}\Delta a_t - (1-\theta)\hat{n}_t - \theta\hat{k}_t^* = 0 \tag{C.14}$$

$$\hat{\mu}_t^* - \hat{\lambda}_t^* + \gamma\hat{\imath}_t^* + \frac{\gamma}{1-\theta}\Delta a_t - \gamma\hat{\imath}_{t-1}^* - \epsilon_t^i = 0 \tag{C.15}$$

with $\hat{\lambda}_t^*$ and $\hat{\mu}_t^*$ being the log-deviations from the steady-state of the two stationarized Lagrange multipliers; $\hat{n}_t$ being the log-deviation from the steady-state of hours worked (which are already stationary); $\hat{x}_t^*$, with $x = y$, c, i, k, being the log-deviation from the steady-state of the stationarized output, consumption, investment, and the capital stock, respectively; and ϵ_t^c and ϵ_t^i being white noise shocks with variances σ_c^2 and σ_i^2, respectively. We add the latter to Barsky and Sims' original model in order to eliminate stochastic singularity. Finally, the following objects are convolutions of the model's structural parameters, and are defined as follows: S_C=(1-$b\tilde{g}_A^{-1/(1-\theta)}$)$^{-1}$/[(1-$b\tilde{g}_A^{-1/(1-\theta)}$)$^{-1}$-$b\beta(\tilde{g}_A^{1/(1-\theta)}$-$b)^{-1}$]; Δ_G=$\tilde{g}_A^{1/(1-\theta)}$/($\tilde{g}_A^{1/(1-\theta)}$-$b$); S_K=$\theta\bar{\mu}_{ss}^*\rho_{YK}$/[$\theta\bar{\mu}_{ss}^*\rho_{YK}$+(1-$\delta$)$\bar{\lambda}_{ss}^*\tilde{g}_A^{-1/(1-\theta)}$]; α_C=$\bar{C}/\bar{Y}$; α_I=$\bar{I}/\bar{Y}$; α_K=(1-δ)/[(1-δ)+$\delta\tilde{g}_A^{1/(1-\theta)}$]; $\bar{\mu}_{ss}^*$=[(1-$b\tilde{g}_A^{-1/(1-\theta)}$)$^{-1}$-$b\beta(\tilde{g}_A^{1/(1-\theta)}$-$b)^{-1}$]/$\bar{C}$; $\bar{\lambda}_{ss}^*$=$\beta\theta\bar{\mu}_{ss}^*\rho_{YK}$/[1-$\beta$(1-$\delta$)$\tilde{g}_A^{-1/(1-\theta)}$], where $\bar{C}$, $\bar{I}$, $\bar{Y}$, and $\bar{K}$ are the values taken by consumption, investment, GDP, and the capital stock in the steady-state, and ρ_{YK}=$\bar{Y}/\bar{K}$ is the value taken by ratio between GDP and the capital stock in the steady-state, and $\bar{\lambda}_{ss}^*$ and $\bar{\mu}_{ss}^*$ are the values taken by the stationarized Lagrange multipliers in the steady-state.

In the benchmark calibration, we set most of the model's parameters as in Barsky and Sims (2011). Specifically, we set β=0.99, δ=0.05, θ=1/3, γ=0.05, $\bar{g}$=0.2, $\tilde{g}_A$=$1.02^{1/4}$, $\bar{c}$=2/3, α_C=2/3, α_I=0.2. We then set b=0 (so that in the benchmark calibration the model features no habit formation in consumption), and $(1/\eta)$=0 (so that the utility function is linear in hours worked). As for the standard deviations of the structural shocks, we set them to σ_{NN}=0.3, σ_{NE}=0.3, σ_u=0.25, σ_v=0.25, σ_c=0.25, σ_n=0.25, σ_g=0.25, σ_i=0.25. As for ρ_{YK}, we calibrate it based on the estimate of the steady-state capital-output ratio for the United States, which Dadda and Scorcu (2003) based on long-run data, estimate at 1.7, so that we have ρ_{YK}=1/1.7=0.5882. Finally, we set $\tilde{g}_I$=$\tilde{g}_A^{1/(1-\theta)}$: the rationale for doing this is simply that, in the steady-state, I_t/I_{t-1}=$\tilde{g}_I$, and since the steady-state gross rate of growth of investment is equal to $\tilde{g}_A^{1/(1-\theta)}$, it ought to be the case that $\tilde{g}_I$=$\tilde{g}_A^{1/(1-\theta)}$.

C.6 Model solution

Following Blanchard et al.'s (2013) Appendix, we compute the solution via the method of undetermined coefficients as follows. We start by putting the RBC model in the form

$$FY_{t+1|t} + GY_t + HY_{t-1} + MS_t + NS_{t+1|t} + Z\epsilon_t = 0 \tag{C.16}$$

where F, G, H, M, N, and Z are matrices of coefficients; Y_t is a vector containing the stationarized endogenous variables for the log-linearized model, that is, Y_t =

$[\hat{\lambda}_t^*, \hat{\mu}_t^*, \hat{y}_t^*, \hat{c}_t^*, \hat{\imath}_t^*, \hat{n}_t, \hat{k}_t^*, \hat{k}_{t+1}^*]'$; and ϵ_t contains all shocks except those pertaining to the signal-extraction problem $(\epsilon_t^{NN}, \epsilon_{t-\tau}^{NE}, u_t, v_t)$, that is, $\epsilon_t = [\epsilon_t^n, \epsilon_t^g, \epsilon_t^c, \epsilon_t^i]'$.

Given the model in the form (C.16), we conjecture that the solution for Y_t takes the form

$$Y_t = PY_{t-1} + QS_t + R\xi_{t|t} + V\epsilon_t \tag{C.17}$$

where P solves the quadratic equation[28]

$$FP^2 + GP + H = 0, \tag{C.18}$$

Q and V are given by

$$Q = -(G + FP)^{-1}M \tag{C.19}$$

$$V = -(G + FP)^{-1}Z \tag{C.20}$$

and, given P and Q, R is obtained by solving iteratively the expression

$$(G + FP)R + [NC + F(QC + R)]A = 0. \tag{C.21}$$

Finally, $\xi_{t|t}$ is the agents' estimate of the vector ξ_t based on information at time t, which is generated by the Kalman filter within the context of the signal-extraction problem. Equations (C.18), (C.19), and (C.21) are the same as in Blanchard et al.'s (2013) Appendix, whereas the additional expression we have, equation (C.20), originates from the fact that we here have additional shocks, over and above those pertaining the signal-extraction problem.

D Computing Truncated Theoretical SVARMA Representations of the RBC Model via Linear Projections

In Section 4.1.2, we compare the theoretical IRFs to non-news, news, and noise shocks produced by Barsky and Sims' (2011) RBC model augmented with noise shocks, and the IRFs produced by several of the model's truncated theoretical SVARMA representations. In this appendix we discuss how we compute such truncated theoretical SVARMA representations based on linear projections arguments (for a discussion of linear projections, see e.g. Sargent (1987)).

Let

$$Y_t = \tilde{A}_0\epsilon_t + \tilde{A}_1\epsilon_{t-1} + \tilde{A}_2\epsilon_{t-2} + \tilde{A}_3\epsilon_{t-3} + \tilde{A}_4\epsilon_{t-4} + \tilde{A}_5\epsilon_{t-5} + ... \tag{D.1}$$

be the infinite structural MA representation of the RBC model, which can be recovered from the model's IRFs to unit-variance structural shocks, and let $\Omega = E[\epsilon_t\epsilon_t']$ be the covariance matrix of the structural innovations. For chosen VAR and MA lag

[28] See Uhlig (1999) for the solution to the quadratic equation.

orders p and q, we compute the model's truncated (in general) theoretical SVARMA representation

$$Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + ... + B_p Y_{t-p} + A_0 \epsilon_t + A_1 \epsilon_{t-1} + ... + A_q \epsilon_{t-q} \tag{D.2}$$

as follows.

The theoretical value of the VAR matrix B_1 in (D.2) is the linear projection of Y_t onto Y_{t-1}. By lagging (D.1) by one period we have

$$Y_{t-1} = \tilde{A}_0 \epsilon_{t-1} + \tilde{A}_1 \epsilon_{t-2} + \tilde{A}_2 \epsilon_{t-3} + \tilde{A}_3 \epsilon_{t-4} + \tilde{A}_4 \epsilon_{t-5} + \tilde{A}_5 \epsilon_{t-6} + ... \tag{D.3}$$

and B_1 is therefore given by

$$B_1 = [\text{Var}(Y_{t-1})]^{-1} \text{Cov}(Y_t, Y_{t-1}) \tag{D.4}$$

where

$$\text{Var}(Y_{t-1}) = \text{Var}(Y_t) = \sum_{j=0}^{\infty} \tilde{A}_j \Omega \tilde{A}_j' \quad \text{and} \quad \text{Cov}(Y_t, Y_{t-1}) = \sum_{j=0}^{\infty} \tilde{A}_{j+1} \Omega \tilde{A}_j' \tag{D.5}$$

Similarly, the theoretical value of the generic VAR matrix B_k, $k = 1, 2, ..., p$, is given by

$$B_k = [\text{Var}(Y_{t-k})]^{-1} \text{Cov}(Y_t, Y_{t-k}) \tag{D.6}$$

where $\text{Var}(Y_{t-k}) = \text{Var}(Y_t)$, and

$$\text{Cov}(Y_t, Y_{t-k}) = \sum_{j=0}^{\infty} \tilde{A}_{j+k} \Omega \tilde{A}_j' \tag{D.7}$$

Having computed the theoretical values of the VAR matrices B_1, B_2, ..., B_p, the next step is then computing the theoretical values of the MA matrices A_0, A_1, ..., A_q, which we do as follows. From (D.1) we have

$$Y_t = \sum_{j=0}^{\infty} \tilde{A}_j \epsilon_{t-j} =$$

$$= \sum_{j=0}^{\infty} \tilde{A}_j \epsilon_{t-j} + (B_1 Y_{t-1} - B_1 Y_{t-1}) + (B_2 Y_{t-2} - B_2 Y_{t-2}) + ... + (B_p Y_{t-p} - B_p Y_{t-p}) =$$

$$= B_1 Y_{t-1} + B_2 Y_{t-2} + ... + B_p Y_{t-p} + \sum_{j=0}^{\infty} \tilde{A}_j \epsilon_{t-j} -$$

$$-B_1[\tilde{A}_0 \epsilon_{t-1} + \tilde{A}_1 \epsilon_{t-2} + \tilde{A}_2 \epsilon_{t-3} + \tilde{A}_3 \epsilon_{t-4} + \tilde{A}_4 \epsilon_{t-5} + \tilde{A}_5 \epsilon_{t-6} + ...]$$

$$-B_2[\tilde{A}_0 \epsilon_{t-2} + \tilde{A}_1 \epsilon_{t-3} + \tilde{A}_2 \epsilon_{t-4} + \tilde{A}_3 \epsilon_{t-5} + \tilde{A}_4 \epsilon_{t-6} + \tilde{A}_5 \epsilon_{t-7} + ...]$$

$$
\begin{aligned}
&-...-\\
&-B_p[\tilde{A}_0\epsilon_{t-p}+\tilde{A}_1\epsilon_{t-(p+1)}+\tilde{A}_2\epsilon_{t-(p+2)}+\tilde{A}_3\epsilon_{t-(p+3)}+\tilde{A}_4\epsilon_{t-(p+4)}+\tilde{A}_5\epsilon_{t-(p+5)}+...]=\\
&=B_1Y_{t-1}+B_2Y_{t-2}+...+B_pY_{t-p}+\tilde{A}_0\epsilon_t+\\
&+[\tilde{A}_1-B_1\tilde{A}_0]\epsilon_{t-1}+\\
&+[\tilde{A}_2-B_1\tilde{A}_1-B_2\tilde{A}_0]\epsilon_{t-2}+\\
&+...+\\
&+[\tilde{A}_p-B_1\tilde{A}_{p-1}-B_2\tilde{A}_{p-2}-...-B_p\tilde{A}_0]\epsilon_{t-p}+\sum_{j=p+1}^{\infty}\tilde{A}_j\epsilon_{t-j} \qquad \text{(D.8)}
\end{aligned}
$$

so that by setting $A_0=\tilde{A}_0$, $A_1=[\tilde{A}_1-B_1\tilde{A}_0]$, $A_2=[\tilde{A}_2-B_1\tilde{A}_1-B_2\tilde{A}_0]$, ..., $A_p=[\tilde{A}_p-B_1\tilde{A}_{p-1}-B_2\tilde{A}_{p-2}-...-B_p\tilde{A}_0]$, we have the truncated theoretical SVARMA representation (D.2), which differs from (D.1) by the term

$$\sum_{j=p+1}^{\infty}\tilde{A}_j\epsilon_{t-j} \qquad \text{(D.9)}$$

To the extent that, for all $j>p$, the elements of the matrices $\tilde{A}_j$'s are sufficiently 'small', the truncated representation (D.2) provides a good approximation to (D.1), which is what Figure 5 is intended to illustrate. Although this property holds in population, still this naturally suggests that, even in finite samples, it should be possible to meaningfully capture the stochastic properties of the DGP via a SVARMA(p, q) with p and q small.

E Monte Carlo Evidence on the Performance of the Proposed Econometric Methodology

In this appendix we present Monte Carlo evidence on the performance of the proposed econometric methodology, taking Barsky and Sims' (2011) RBC model, augmented with noise shocks about future TFP as the DGP. We focus on two main questions: (i) Can the proposed identification scheme and estimation approach correctly recover the response of the economy to non-news, news, and noise shocks? (ii) Can they correctly capture how important the three shocks are at driving the dynamics of individual variables?

We start by briefly describing the Monte Carlo experiment's setup.

E.1 Details of the Monte Carlo experiment

Based on the DGP described in Section 3 and Appendix C, we generate 1,000 artificial samples of length equal to the actual sample length in the application of Section 5, that is, $T = 216$. Based on each sample we then estimate a VARMA(2,1) based on the econometric methodology described in Section 4, imposing *exactly* the same restrictions we impose when we work with the actual data with a single difference, which we now discuss.

As discussed in Section 3 and Appendix C, for noise shocks to play a role in the model, it ought to be the case that the permanent component of TFP is not observed, which requires a transitory TFP shock (i.e., v_t in equation (C.5) in Appendix C). A key point to stress is that if v_t were not there, the entire signal-extraction problem would disappear, and noise shocks would play no role whatsoever. This, however, implies that in the DGP we are using for the Monte Carlo exercise, the non-news shock is not the only shock impacting TFP contemporaneously, so that imposing the restriction that it is would end up distorting the estimates. In the Monte Carlo exercise we therefore allow for two shocks to impact upon TFP at t=0, and we disentangle them by rotating them in such a way that the transitory one explains the minimum fraction of TFP at the relevant long horizon.

Finally, we impose the restriction that the largest eigenvalue of the VARMA is (i) greater than or equal to 0.99, and (ii) strictly smaller than 1. Restriction (ii) is not crucial, and it is imposed simply in order to rule out explosive IRFs and paths. Restriction (i), on the other hand, is crucial only for the Monte Carlo exercise, whereas it is completely irrelevant for the applications based on actual data of Sections 5 and 6. In the Monte Carlo exercise, not imposing this restriction produces IRFs which tend to mean-revert to zero too fast, which prevents them from effectively capturing the permanent nature of TFP shocks on TFP, GDP, consumption, and investment.

E.2 Evidence

Figure E.1 reports, for either non-news, news, or noise shocks, the means, across all of the 1,000 Monte Carlo simulations, of the 50th, 16th, and 84th percentiles of the posterior distributions of the IRFs, whereas Figure E.2 reports the means of the 50th, 16th, and 84th percentiles of the posterior distributions of the fractions of FEV explained by either of the three shocks.

The evidence in Figure E.1 is qualitatively similar to that reported in Barsky and Sims' Figure 1 for news shocks, with the model's true IRFs to either of the three shocks lying almost uniformly inside the 16-84 bands. Results are especially good for the main shock of interest in this paper—the noise shock—with the means of the 50th percentiles of the posterior distributions of the IRFs being typically close to the true IRFs. As for non-news shocks, results are likewise excellent for investment and hours, whereas for TFP, GDP, and consumption the true IRFs are still fully inside the 16-84 bands, but the means of the 50th percentiles of the posterior distributions

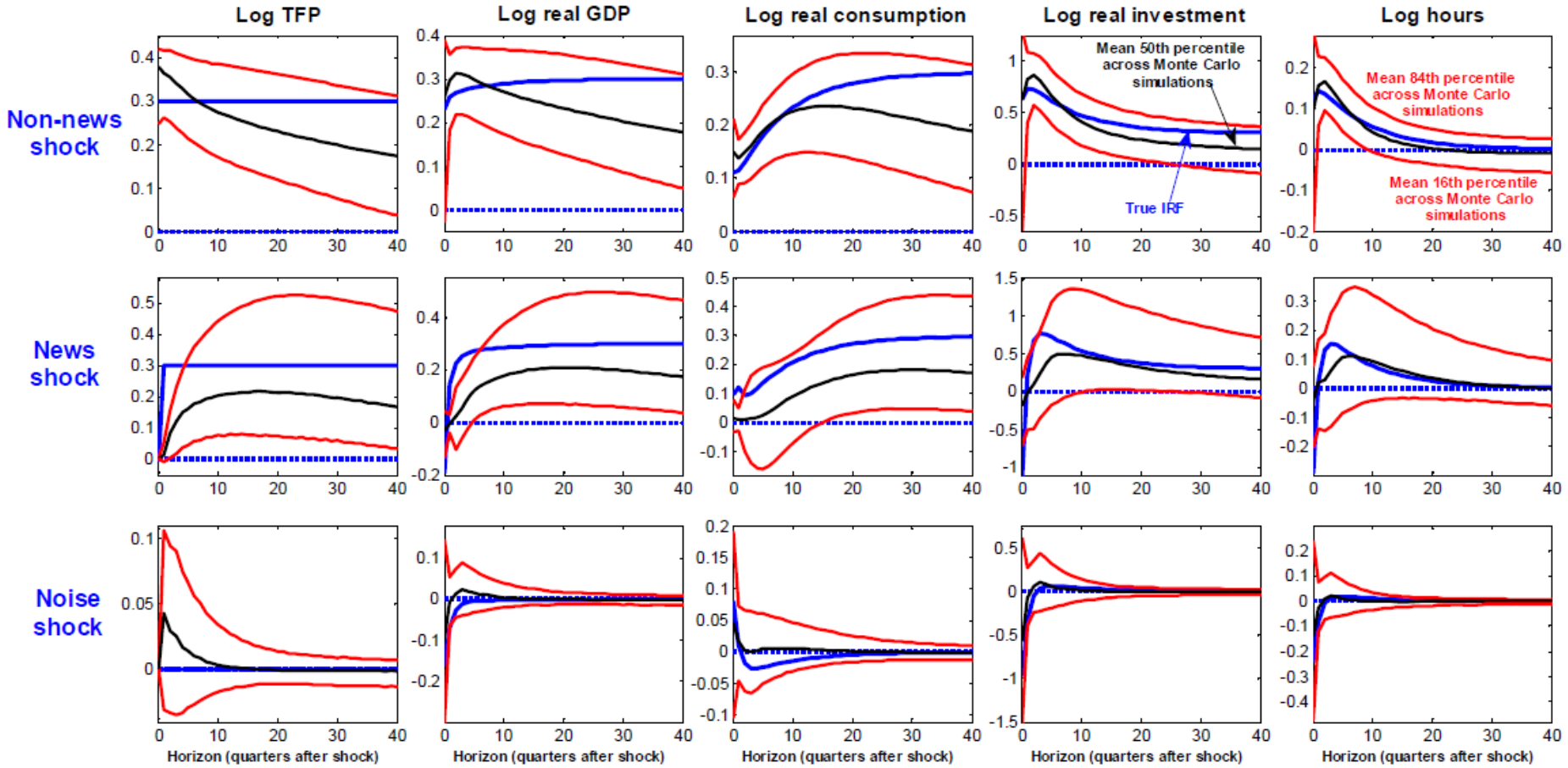

Figure E.1 Results from the Monte Carlo exercise: Means, across all of the Monte Carlo simulations, of the 50th, 16th, and 84th percentiles of the posterior distributions of the impulse-response functions

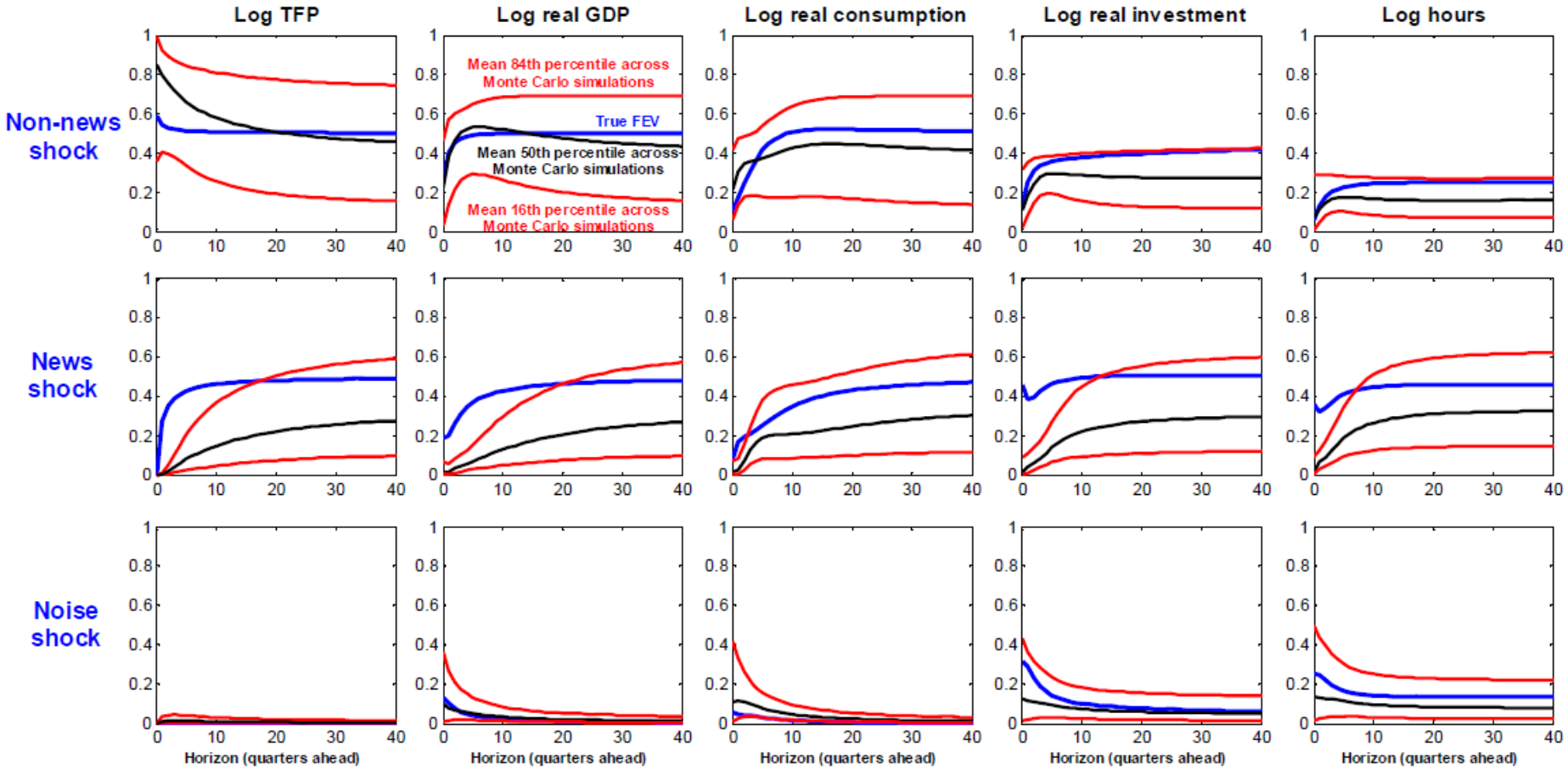

Figure E.2 Results from the Monte Carlo exercise: Means, across all of the Monte Carlo simulations, of the 50^{th}, 16^{th}, and 84^{th} percentiles of the posterior distributions of the fractions of forecast error variance

are not close to them. As for news shocks, the 16-84 bands still capture the true IRFs almost uniformly, but this is not the case for the impacts at t=0 for GDP, consumption, investment, and hours, whose magnitude is uniformly under-estimated. Further, estimated IRFs tend to rise more slowly than the true IRFs towards the new long-run equilibrium, with the result that for both TFP and GDP the true IRFs remain outside the 16-84 tunnels for about a year-year and a half after the impact.

Turning to the evidence reported in Figure E.2 (it is to be noticed that Barsky and Sims (2011) did not report Monte Carlo evidence for the fractions of FEV), results are once again excellent for the noise shock, with the true fractions of FEV lying uniformly inside the 16-84 band, and the means of the 50th percentiles of the posterior distributions being very close to the true fractions. Results are likewise uniformly good for non-news shocks, whereas for news shocks the performance is qualitatively the same as that for the IRFs, with the 16-84 bands systematically capturing the true fractions of FEV at medium-to-long horizons, but failing to do so at the short horizons.

Overall, our own assessment is that the performance of the proposed estimation and identification methodology against this specific DGP although by no means perfect, is nonetheless good. This is especially the case for the noise shock, which is the main object of interest in the present paper.

F The Data

We use John Fernald's purified TFP series available from the San Francisco Fed's website. A seasonally adjusted series for real GDP (GDPC96) is from the U.S. Department of Commerce: Bureau of Economic Analysis. Inflation has been computed as the log-difference of the GDP deflator (GDPCTPI) taken from the St. Louis Fed's website. Hours worked by all persons in the nonfarm business sector (HOANBS) is from the U.S. Department of Labor, Bureau of Labor Statistics. The seasonally adjusted series for real chain-weighted investment, consumption of non-durables and services, and their deflators (which we use in order to compute the chain-weighted relative price of investment) have been computed based on the data found in Tables 1.1.6, 1.1.6B, 1.1.6C, and 1.1.6D of the National Income and Product Accounts. Whereas real consumption and its deflator pertain to non-durables and services, real investment and its deflator have been computed by chain-weighting the relevant series pertaining to durable goods; private investment in structures, equipment, and residential investment; Federal national defense and non-defense gross investment; and State and local gross investment. All these variables are available at the quarterly frequency.

The remaining variables are available at a monthly frequency and have been converted to the quarterly frequency by taking averages within the quarter. The Federal funds rate (FEDFUNDS) and the 5-year government bond yield (GS5) are taken from the St. Louis Fed's website. They are quoted at a non-annualized rate in

order to make their scale exactly comparable to that of inflation.[29] Seasonally unadjusted nominal dividends and stock prices (the S&P 500 index) are both from Robert Shiller's website. They have then been deflated by the GDP deflator. Civilian noninstitutional population (CNP16OV) is from the U.S. Department of Labor, Bureau of Labor Statistics.

G Model Comparison Exercise

G.1 Deviance Information Criterion

The Deviance Information Criterion (DIC) was introduced in Spiegelhalter, Best, Carlin, and vanderLinde (2002). For latent variable models there are a few distinct variants depending on the exact notion of the likelihood (Celeux, Forbes, Robert, and Titterington (2006)). Given a likelihood function $f(\mathbf{y} \mid \theta)$, the DIC is defined as:

$$\text{DIC} = \overline{D(\theta)} + p_D,$$

where

$$\overline{D(\theta)} = -2E_\theta[\ln f(\mathbf{y} \mid \theta) \mid \mathbf{y}]$$

is the posterior mean deviance and p_D is the effective number of parameters. That is, the DIC is the sum of the posterior mean deviance, which can be used as a Bayesian measure of model fit or adequacy, and the effective number of parameters that measures model complexity. The effective number of parameters is in turn defined as

$$p_D = \overline{D(\theta)} - D(\tilde{\theta}),$$

where $D(\theta)$=-2ln $f(\mathbf{y} \mid \theta)$, and $\tilde{\theta}$ is an estimate of θ, which is typically taken as the posterior mean.

Following Chan, Eisenstat, and Koop (2016), we use the likelihood implied by the system

$$\mathbf{y}_t = \sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \sum_{j=1}^{q} \mathbf{\Theta}_j \boldsymbol{\epsilon}_{t-j} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{\Sigma}). \tag{G.1}$$

where all the parameters are identified and can be recovered from the main sampling algorithm.

To derive this density, we stack (G.1) over t and obtain:

$$\mathbf{y} = \mathbf{a} + \mathbf{\Theta}\boldsymbol{\epsilon} \tag{G.2}$$

where $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_\mathbf{1}', ..., \boldsymbol{\epsilon}_\mathbf{T}']' \sim \mathcal{N}(0, I_T \otimes \mathbf{\Sigma})$, $a = [(\sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{1-j})', ..., (\sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{T-j})']'$, and $\mathbf{\Theta}$ is a $Tn \times Tn$ lower triangular matrix with the identity matrix $\mathbf{I}_n$ on the main

[29] To be clear, if we define an interest rate series as R_t—with its scale such that, e.g., a ten per cent rate is represented as 10.0—the rescaled series is computed as r_t=$(1+R_t/100)^{1/4}$-1.

diagonal block, $\mathbf{\Theta}_1$ on the first lower diagonal block, $\mathbf{\Theta}_2$ on the second lower diagonal block, and so forth. Hence, we have

$$(\mathbf{y} \mid \mathbf{A}_1, ...\mathbf{A}_p, \mathbf{\Theta}_1, ...\mathbf{\Theta}_q, \mathbf{\Sigma}) \sim \mathcal{N}(a, \mathbf{\Theta}(\mathbf{I}_T \otimes \mathbf{\Sigma})\mathbf{\Theta}').$$

Since the covariance matrix $\mathbf{\Theta}(\mathbf{I}_T \otimes \mathbf{\Sigma})\mathbf{\Theta}'$ is a band matrix, this Normal density can be evaluated quickly using the band matrix algorithms discussed in Chan and Grant (2016).

G.2 Estimated DIC values for alternative models

We work with a set of $n-$variate VARMA(p,1) models and consider various choices of n and p. All the DICs are computed using the marginal distribution of the six variables in the $n = 6$ case as the likelihood. A model with a smaller DIC value is preferred. They are reported in Table G1. Models with $p = 4$ clearly dominate for all choices of n and all results presented in this paper use this lag length. With regard to n, the choice n=8dominates choices of a similar dimension. Since medium-size VARs of approximately this dimension are typically used in this literature, the results presented in the body of the paper use n=8. However, the lowest value of DIC is obtained for the larger VARMA with $n = 15$. In the next appendix, we present IRFs and FEVs for this case and find them to be quite similar, but slightly less precisely estimated than those with n=8.

Table G1: Estimated DIC values[a] and associated numerical standard errors (in parentheses)

	$n = 6$	$n = 8$	$n = 9$	$n = 10$	$n = 15$
$p = 2$	3064.7	3016.5	3044.4	3003.5	2908.6
	(0.10)	(0.05)	(0.16)	(0.27)	(0.16)
$p = 4$	3022.4	3000.7	3008.4	2981.7	2883.0
	(0.13)	(0.13)	(0.37)	(0.27)	(0.20)

[a] The DICs are computed using the marginal distribution of the six variables in the $n = 6$ case as the likelihood.

G.3 Evidence based on the model selected by the DIC criterion

Figures G.1-G.6 report evidence for the model with $n = 15$, which, based on the results reported in Table G1, is the one preferred by the DIC criterion. Since, as discussed in the paper, the results produced by models in which we do, or we do not impose restrictions on the absolute values of the IRFs to news and noise shocks are very close, the evidence reported in Figures G.1-G.6 comes from a model in which we have *not* imposed such restrictions (the only reason for doing so is that, with

n =15, imposing such restrictions is very computationally intensive). The evidence reported in Figure G.3 confirms the main finding in Section 5: Noise shocks explain uniformly negligible fractions of the FEV of all variables at all horizons. Further, the fractions of FEV are typically estimated quite precisely: This is especially the case for noise and non-news shocks, whereas it is less so for news shocks. As for the IRFs, the broad pattern for non-news and news shocks is the same as in Figure 8, with the main difference being the smaller extent of precision. As for noise shocks, on the other hand, the IRFs in Figure G.6 are so imprecisely estimated that it is essentially impossible to say anything about the response of the economy to these disturbances.

H Additional Empirical Results

As mentioned in the text, online Appendices I and II contain the entire sets of results based on identification schemes I and II, respectively. Specifically, for the application with TFP, we present (i) results for $n = 6$ based on VARMA(p, q)'s with $p = 2$, 4 and $q = 1$, 2, 3 without imposing restrictions on the absolute values of the IRFs to news and noise shocks; and (ii) results for $n = 8$ based on VARMA(4, 1)'s either imposing or not imposing restrictions on the absolute values of the IRFs to news and noise shocks. The main point to stress here is that our key finding—noise shocks play a uniformly negligible role—is remarkably robust across all specifications. Further, both the IRFs and the fractions of FEVs are, likewise, very similar across specifications.

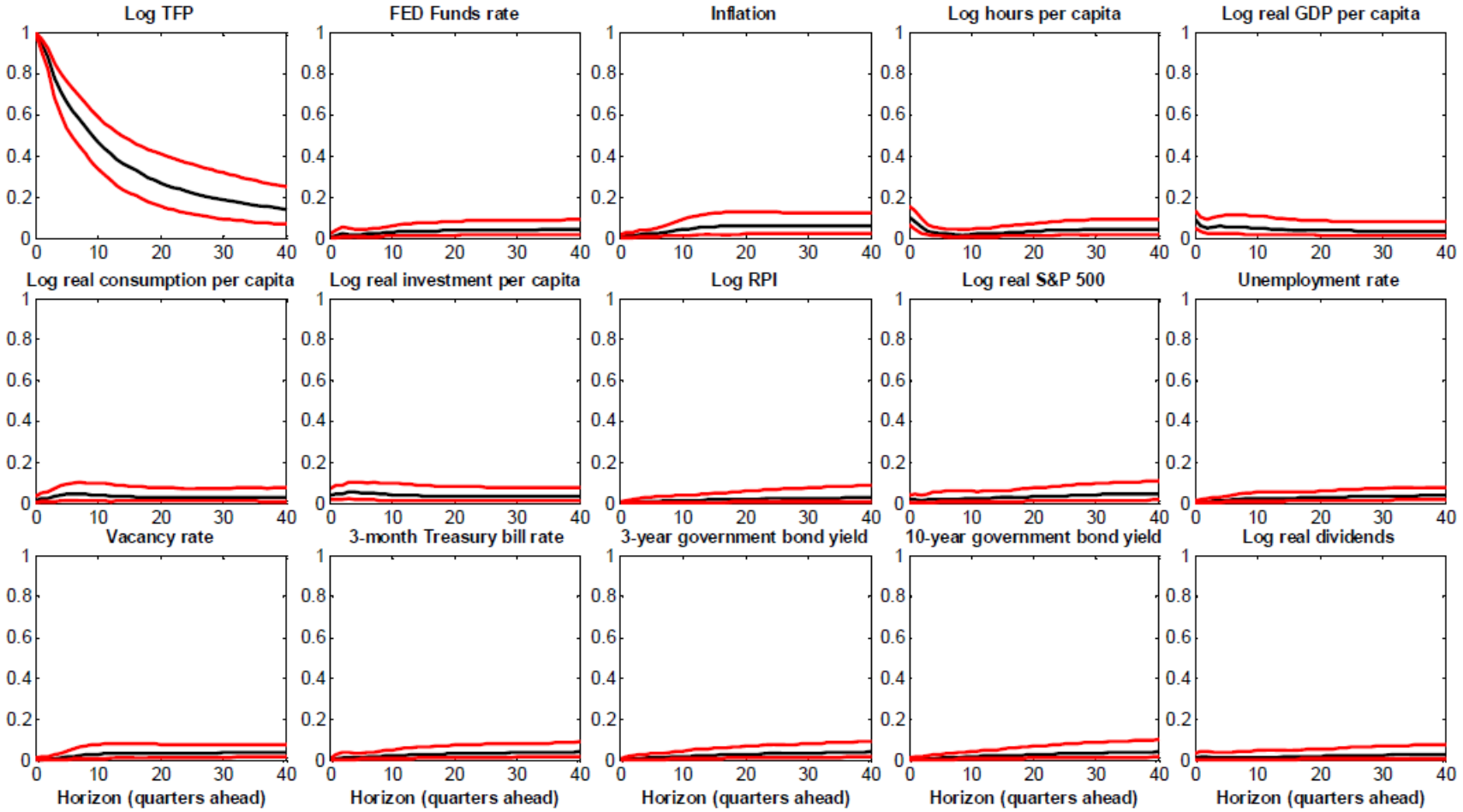

Figure H.1 Fractions of forecast error variance explained by non-news shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

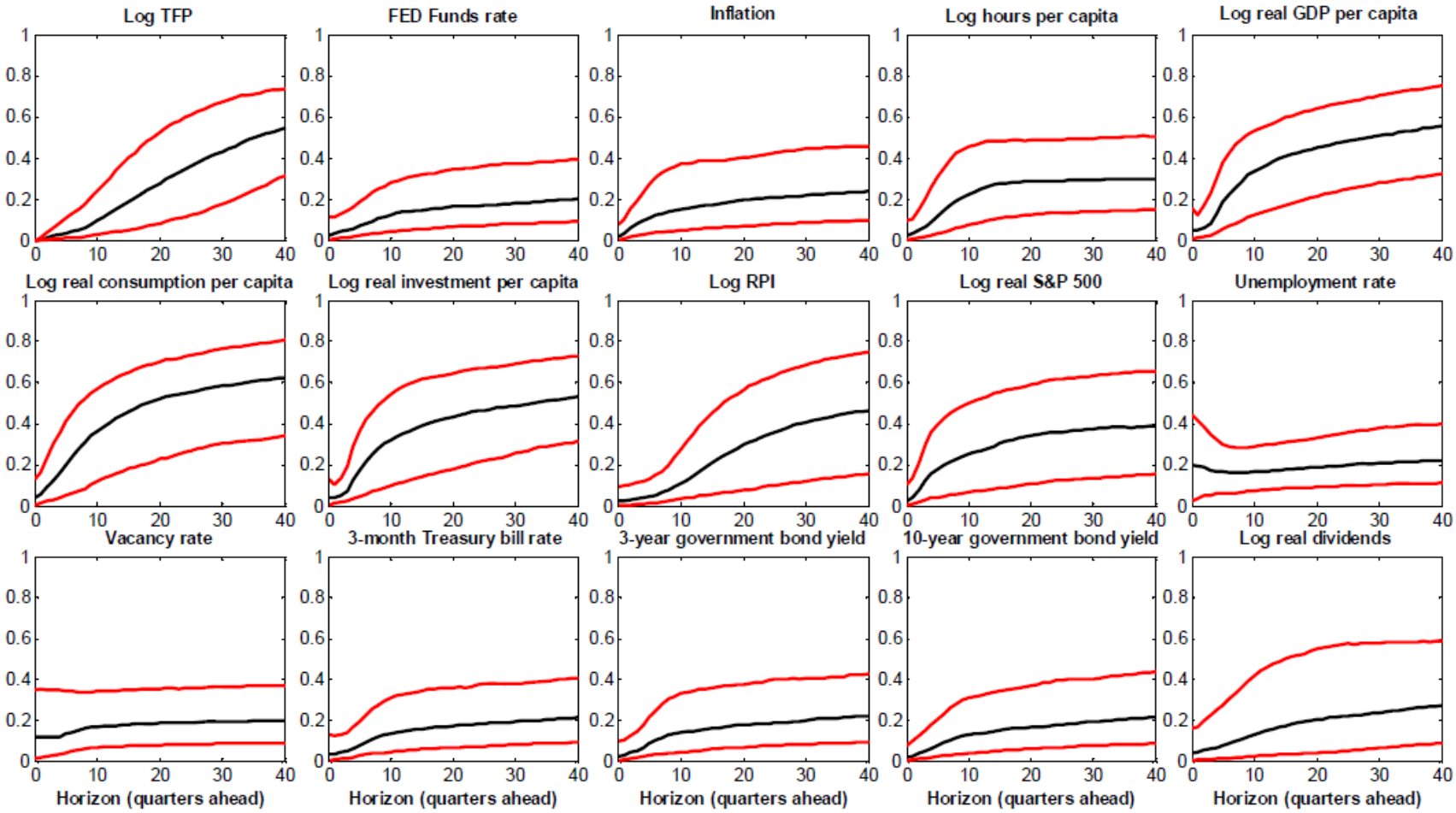

Figure H.2 Fractions of forecast error variance explained by news shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

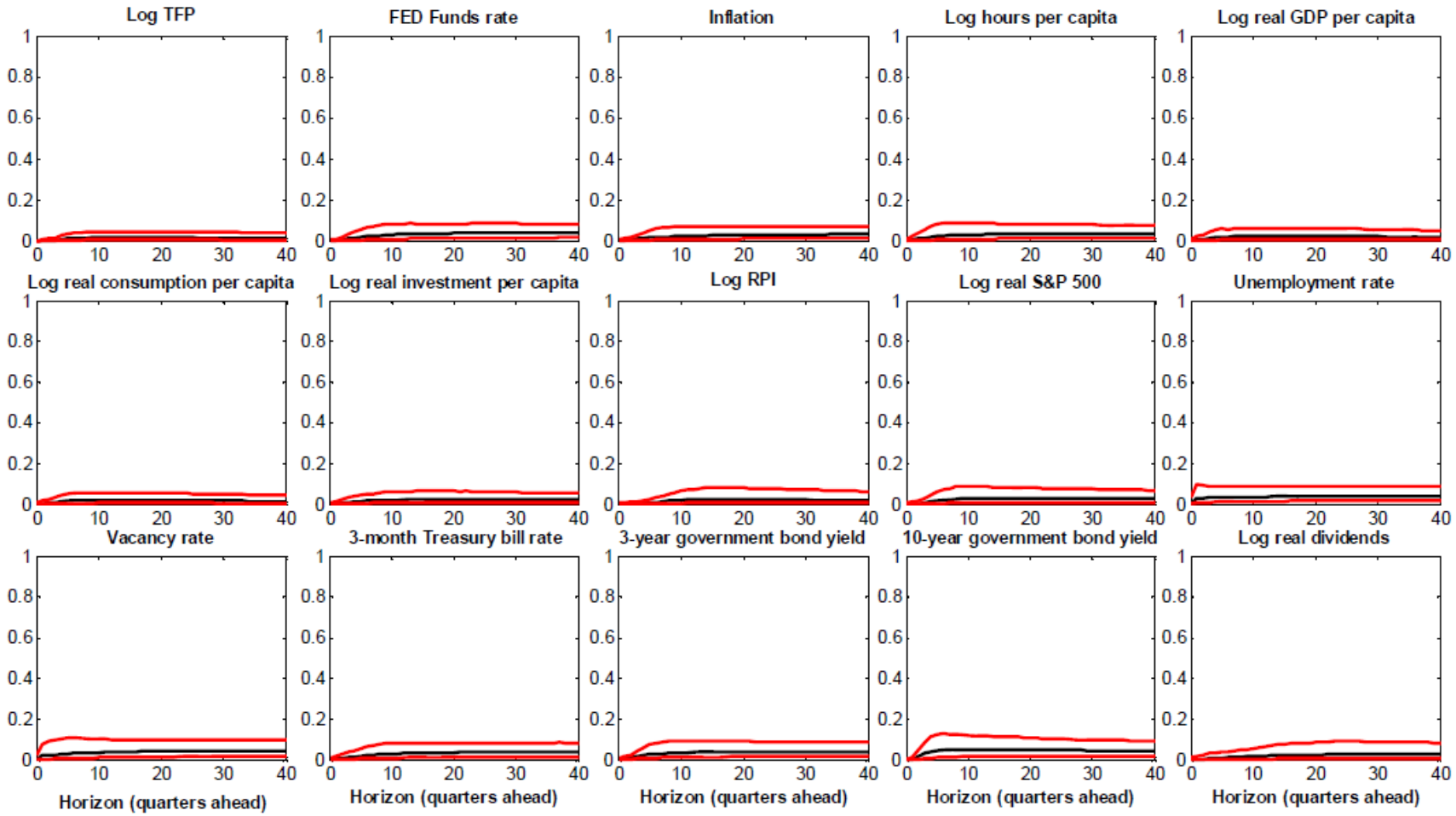

Figure H.3 Fractions of forecast error variance explained by noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

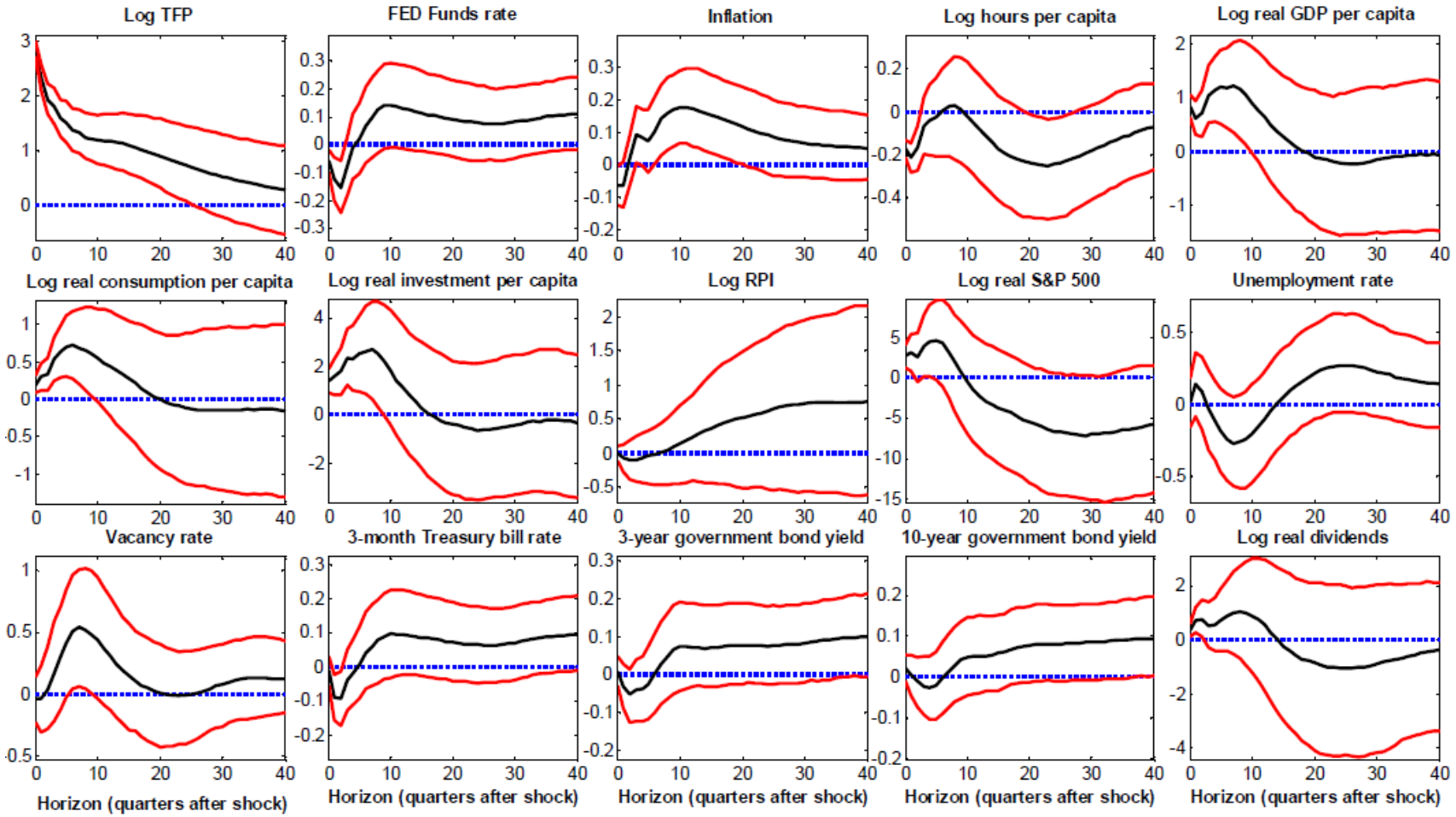

Figure H.4 Impulse-response functions non-news shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

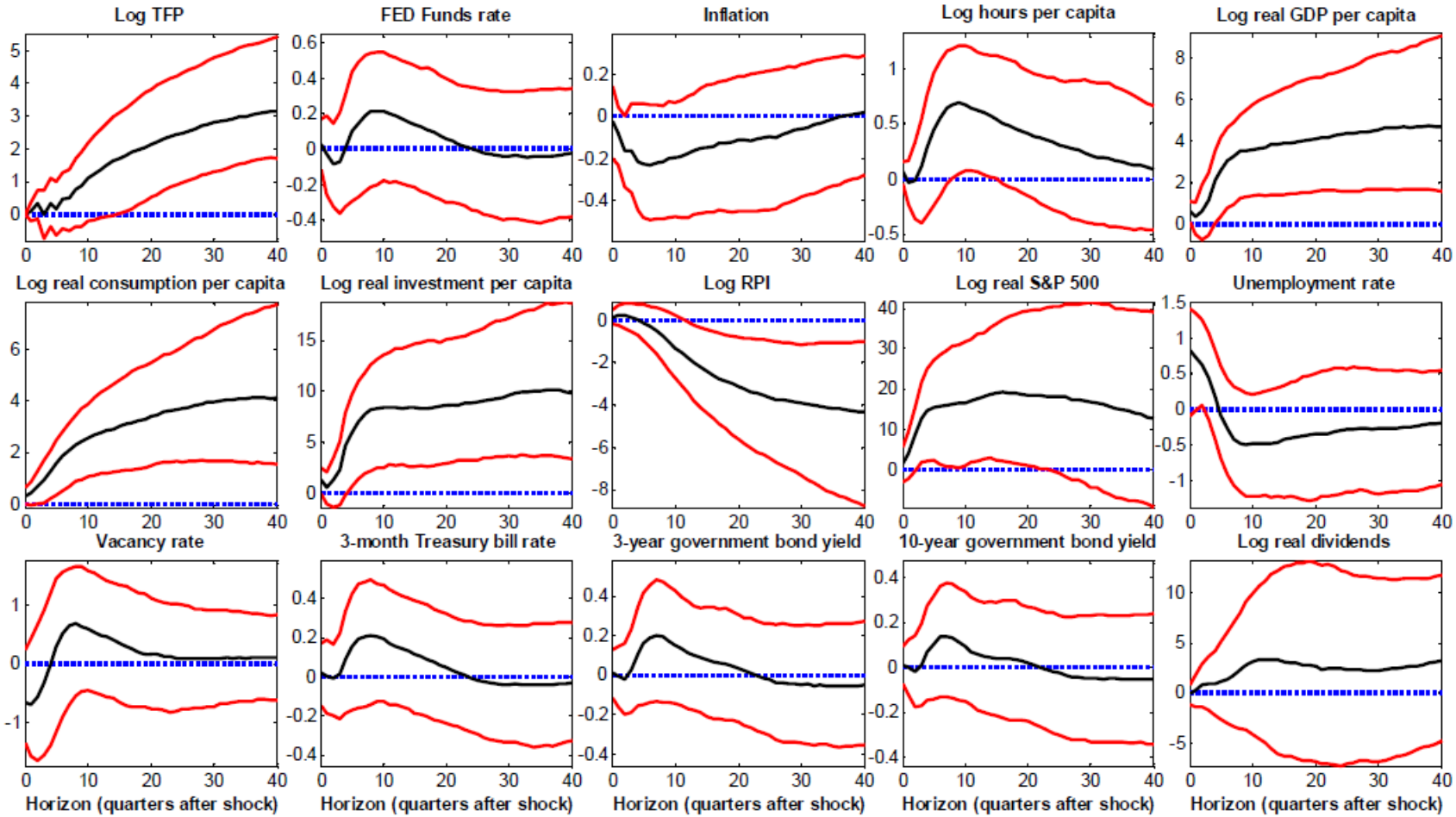

Figure H.5 Impulse-response functions news shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

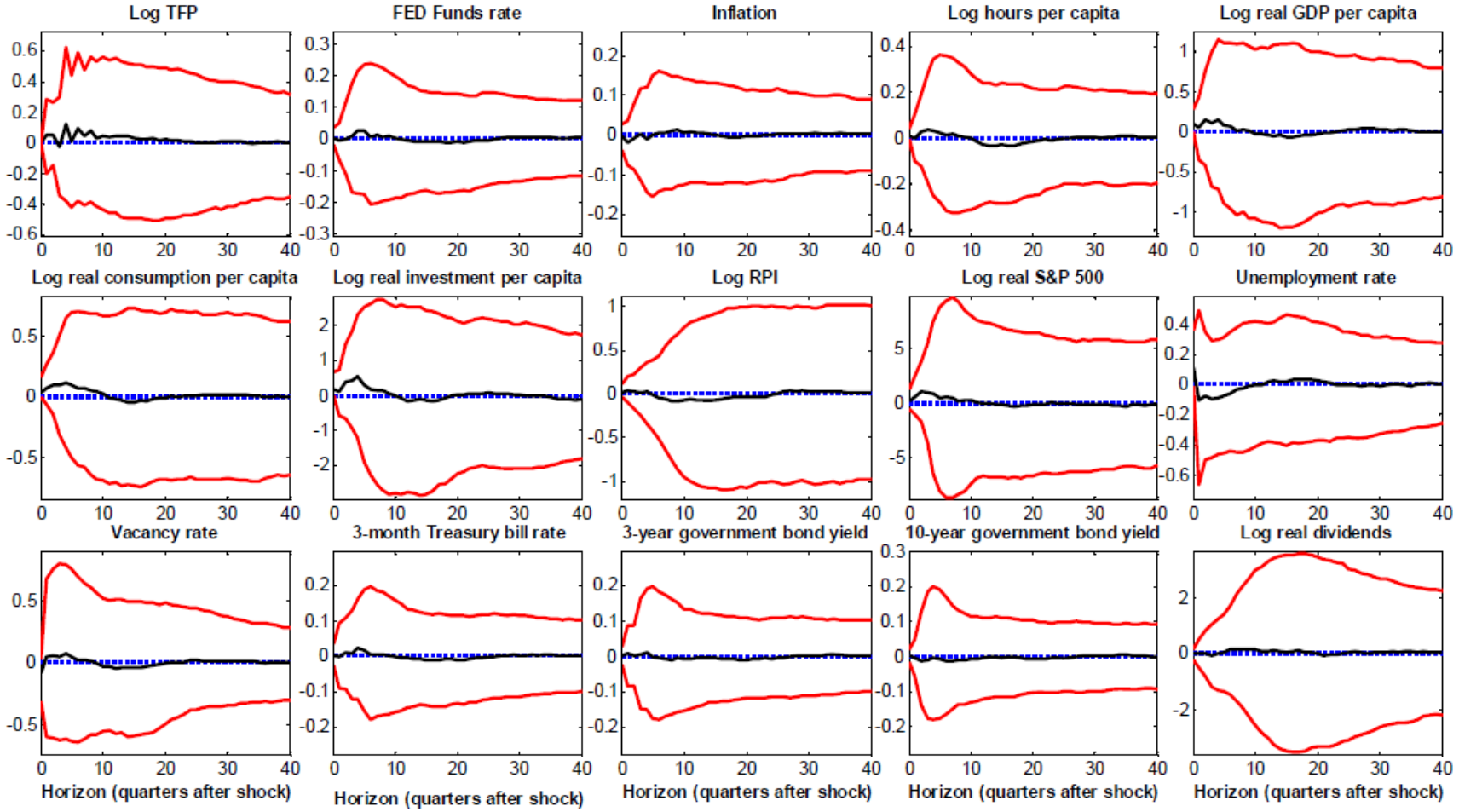

Figure H.6 Impulse-response functions noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact