

# Choosing Between Identification Schemes in Noisy-News Models\*

Joshua Chan  
Purdue University<sup>†</sup>

Eric Eisenstat  
University of Queensland<sup>‡</sup>

Gary Koop  
University of Strathclyde<sup>§</sup>

## Abstract

This paper is about identifying structural shocks in noisy-news models using structural vector autoregressive moving average (SVARMA) models. We develop a new identification scheme and efficient Bayesian methods for estimating the resulting SVARMA. We discuss how our identification scheme differs from the one which is used in existing theoretical and empirical models. Our main contributions lies in the development of methods for choosing between identification schemes. We estimate specifications with up to 20 variables using US macroeconomic data. We find that our identification scheme is preferred by the data, particularly as the size of the system is increased and that noise shocks generally play a negligible role. However, small models may overstate the importance of noise shocks.

---

\*We would like to thank Luca Benati who, in earlier collaborations, offered many insightful comments about noisy-news models which helped inspire the present paper.

<sup>†</sup>Department of Economics, 100 Grant St, West Lafayette, IN 47907, USA. Email: joshuacc.chan@gmail.com

<sup>‡</sup>School of Economics, University of Queensland, St Lucia, QLD 4072, Australia. Email: e.eisenstat@uq.edu.au

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

# 1 Introduction

This paper considers econometric estimation and model selection for noisy-news models. Macroeconomic models such as this involve the use of identifying assumptions so as to give shocks and parameters an economic interpretation. These identifying assumptions have implications for the econometric model being estimated. In earlier work with noisy-news models (Benati, Chan, Eisenstat, and Koop, 2020), we developed a new way of identifying news and noise shocks and showed how these can be estimated using an SVARMA. That paper further explored the relationship between this identification scheme and the robust features of Dynamic Stochastic General Equilibrium (DSGE) models, such as the Real Business Cycle (RBC) models, which are typically employed in the noisy-news literature. However, not all assumptions underlying the SVARMA methodology are aligned with assumptions made in theoretical models.

In this paper, we consider two alternative assumptions regarding agents' information sets with strong implications for empirical work. The first is a standard assumption used in the theoretical noisy-news literature, which as we show, places strong restrictions on the time series process that characterises macroeconomic data. Such restrictions imply standard time series methods such as VARs or VARMA cannot be used in general, and this has led previous empirical work to with more restrictive specifications (e.g. Forni, Gambetti, Lippi, and Sala, 2017). However, a key point in this paper is that such a theoretical assumption can be tested empirically: it is possible to quantify the extent to which multivariate time series data support it.

To this end, we derive an identification scheme based on alternative assumptions regarding agents' information sets. Consequently, the resulting econometric specification lacks the strong restrictions imposed by theoretical macroeconomic models. We further show that news and noise shocks can be identified in either framework (albeit under different assumptions about observables), but relaxing the strong theoretical restrictions leads to a straightforward SVARMA model that can be easily implemented in practice. We then develop methods for empirically comparing the two specifications defined by the alternative theoretical assumptions. In particular, we derive methods for calculating the Bayes factor comparing the two models. In an empirical application, we find strong evidence in favour of our identifying assumptions.

The paper is organized as follows. Section 2 discusses the existing literature in relation to identification and econometric estimation and outlines our contribution relative to it. Section 3 illustrates the issues involved in the context of a simple theoretical model. Section 4 provides formal details of our identification scheme and a discussion of how it relates to assumptions made in the theoretical literature. Section 5 describes our econometric methods and, in particular, derives methods for calculating the Bayes factor comparing different identification schemes. Section 6 is an empirical exercise which investigates the role of news and noise shocks in an application involving a large US macroeconomic data set. Section 7 concludes.

## 2 Identification and Econometric Estimation of Noisy-News Models

Following the seminal contribution of Beaudry and Portier (2006), the notion that a fundamental variable (e.g. productivity) may be driven not only by the familiar ‘surprise’ (i.e., non-news) shocks with an immediate impact, but also by anticipated (i.e., news) shocks is at the forefront of the macroeconomics research agenda (see Barsky, Basu and Lee, 2014). News shocks are defined as those which change expectations about future productivity, but have no immediate impact on current productivity. In recent years, several authors—see in particular Barsky and Sims (2012), Blanchard, L’Huillier, and Lorenzoni (2013) and Forni, Gambetti, Lippi, and Sala (2017)—have started to explore the implications of a straightforward departure from the basic ‘news versus non-news’ dichotomy: 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?

Developing appropriate econometric methods for estimating the effects of news and noise shocks has proven to be challenging. Barsky and Sims (2012) and Blanchard, L’Huillier, and Lorenzoni (2013) both estimate DSGE models directly, but under different theoretical settings they obtain diametrically opposing results regarding the role of noise shocks in driving economic business cycles. Blanchard, L’Huillier, and Lorenzoni (2013) further comment that semi-structural time series methods such as SVARs cannot be used to identify news and noise shocks. Forni, Gambetti, Lippi, and Sala (2017) build on this and provide an in-depth analysis of the issues, which can be summarized as follows.

There are four key features of including non-news, news, and noise shocks in the same econometric model that make this setting particularly difficult to work with:

1. News and noise shocks affect all variables identically on impact (since agents cannot initially distinguish between them), which implies the infinite moving average (MA) representation is *non-fundamental*.
2. Different non-fundamental MA representations imply different speeds with which agents learn about news and noise shocks from future observations.
3. DSGE models featuring news and noise shocks *assume* information structures that result in systems where non-news, news and noise shocks *cannot* be computed from past, current, and future observations of economic variables.
4. The role of news and noise shocks is likely to be related to the amount of information observed by agents, and therefore, the number of variables considered by the econometrician is important.

Issue 1 means that a traditional SVAR approach cannot be used. Issue 2 adds to the difficulty because non-fundamental representations are observationally equivalent when Gaussian errors are assumed. Hence, an approach is needed to obtain inference without arbitrarily imposing restrictions on how quickly agents learn about the nature of the shocks.

Forni, Gambetti, Lippi, and Sala (2017) deal with issues 1 and 2 by employing a particular set of *dynamic rotations* of news and noise shocks. Consequently, they

assume the transformed system is fundamental and proceed to estimate a standard VAR. As it turns out, the dynamic rotations can be recovered from the estimated VAR coefficients and the transformation reversed. The implicit structural model that results is a VARMA, which Forni, Gambetti, Lippi, and Sala (2017) use to compute impulse response functions (IRFs) and forecast error variance decompositions (FEVDs). They find noise shocks to play a substantial role in explaining economic fluctuations.

An important feature of the approach in Forni, Gambetti, Lippi, and Sala (2017) is that non-news shocks are excluded from the model. This is a consequence of issue 3, which implies that under the typical *theoretical assumptions*, certain restrictions are imposed on agents' information sets such that the three shocks (non-news, news and noise) are not recoverable from the data at any horizon. Hence, a VAR representation cannot be obtained. However, we stress that these restrictions (and, hence, the non-recoverability of the shocks) is a result of assumptions, which as we show below, can be tested empirically.<sup>1</sup>

The primary aim of this paper is to test the theoretical assumption underlying issue 3 that is utilized by the theoretical literature. We do so by assuming *per a contrario* that the restrictions on agents' information sets in issue 3 do not hold, and subsequently, develop econometric methods that successfully address issues 1, 2 and 4.

This is motivated by a number of considerations. Firstly, excluding non-news shocks outright (as Forni, Gambetti, Lippi, and Sala, 2017, do) may not always be desirable as it implies *all* exogenous variation in, e.g., productivity is anticipated. Secondly, to our knowledge, no empirical investigation to date has considered systems with more than six variables, and therefore, issue 4 has not been thoroughly addressed. On the other hand, the plausibility of restrictions on agents' information sets may be related to the amount of information that is available to agents, and hence, it is sensible to consider empirical tests of such restrictions, particularly as more variables are added to the econometric specification.

Consequently, the first contribution of this paper is to introduce a set of identifying assumptions on a SVARMA and prove that they allow for identification of impulse responses to non-news, news and noise shocks. We do this for both cases—when agents' information sets are restricted (as in DSGE models) as well as when they are not. The second contribution of this paper lies in the development of Bayesian econometric methods for estimating the resulting SVARMA in the unrestricted case. This is not a straightforward development since a common finding is that VARMA can be over-parameterized and difficult to estimate, particularly in high dimensional models such as those used in this paper. Building on previous work with reduced form VARMA (see Chan, Eisenstat, and Koop, 2016; Chan and Eisenstat, 2017), we derive methods for estimating the SVARMA implied by our identification scheme. We use stochastic search variable selection (SSVS) methods (see George, Sun, and Ni, 2008) to pick out restrictions and ensure parsimony in an otherwise over-parameterized model. In our empirical section, we successfully work with VARMA with up to 20 dependent variables. For reasons discussed above, and given the general desire of macroeconomists to work with larger data sets (e.g. due to concerns about whether an insufficiency of

---

<sup>1</sup>Indeed, in the presence of these restrictions, Chahrour and Jurado (2018a) prove that under additional restrictions news and noise representations can be observationally equivalent.

information could lead to a lack of identification of structural VARs expressed in Forni and Gambetti, 2014), the ability of our econometric methods to work with such large dimensional SVARMA is a very useful feature.

Given our departure from the restrictions in issue 3, the third contribution of this paper lies in exploring the differences between our set of assumptions and those used in the theoretical literature. The key distinction is formalized in our Assumption 1 (defined below). This assumption implies that all structural shocks are a function of past current and future observations of the variables in the model. In contrast, the restrictions in the theoretical literature arise from the assumption that each agent's information set contains only the current and past values of a fundamental variable as well as a signal about the news shock. Such an assumption contradicts our Assumption 1: whereas Assumption 1 implies the transfer function of a linear system is full rank (almost everywhere), the theoretical literature assumes that it is singular.

We show that IRFs may be uniquely identified under either of these assumptions, but is our Assumption 1 better than the assumption employed in the theoretical literature? Our attempt to answer this question constitutes the fourth contribution of this paper. We develop a method for empirically testing which assumption is supported by the data. In particular, we use the Savage-Dickey density ratio (SDDR, see Verdinelli and Wasserman, 1995) to obtain a Bayes factor comparing a model identified using our assumptions to one in which restrictions from the theoretical literature hold.

The fifth contribution lies in a substantive empirical application involving up to 20 variables. We find our econometric methods to work well. In particular, we are able to disentangle non-news, news and noise shocks, and the Bayes factors indicate support for our identifying scheme. Our main empirical result is that noise shocks play a minor role in macroeconomic fluctuations, explaining small, or even negligible, fractions of the forecast error variance of main macroeconomic variables. However, small models may overstate the importance of noise shocks. As more variables are added, news shocks emerge as the only ones that substantially explain macroeconomic fluctuations.

We also note that the SVARMA-based econometric methods we develop in this paper are for a specific class of macroeconomic models involving a particular identification scheme. This class of macroeconomic models is an important and popular one, but (with appropriate alternations) our econometric methods can be used with other non-fundamental SVARMA identified in different ways.

### 3 Motivating Example

Suppose a fundamental process  $\{a_t : t \in \mathbb{Z}\}$  is defined by:

$$a_t = \epsilon_t^{\text{NN}} + \epsilon_{t-1}^{\text{NE}} + \theta \epsilon_{t-2}^{\text{NE}}. \quad (1)$$

We will refer to  $\epsilon_t^{\text{NE}}$  as the *news* shock because it does not impact the fundamental process immediately, but rather with a lag of one or more periods. In (1), the news shock affects the fundamental a minimum of one period and a maximum of two periods (if  $\theta \neq 0$ ) after its realization. In contrast,  $\epsilon_t^{\text{NN}}$  is a shock that affects the fundamental process in the same period in which it occurs. We refer to this as the *non-news* shock

to convey the idea that it captures everything that impacts the fundamental but is not anticipated. It can also be regarded as the conventional *surprise* shock in the news-driven business cycles literature (e.g., Barsky and Sims, 2011).

We emphasize that the fundamental process defined by (1) is *not* affected by any other shock besides non-news and news. Indeed, this is a general feature of the type of fundamental processes defined in the literature: non-news and noise shocks explain 100% of the forecast error variance of the fundamental process. We formalize the concept of a news-driven fundamental process as Assumption 3 in subsequent sections.

In the *noisy* news literature, the key assumption is that agents do not observe  $\epsilon_t^{\text{NE}}$  at time  $t$ , but rather a signal of the form:

$$s_t = \epsilon_t^{\text{NE}} + u_t. \quad (2)$$

The shock  $u_t$  captures the *noise* in the signal: it prevents agents from observing  $\epsilon_t^{\text{NE}}$  completely, but it does not impact the fundamental process at any period. The three shocks are assumed to be independent white noise processes, namely  $(\epsilon_t^{\text{NN}}, \epsilon_t^{\text{NE}}, u_t)' \sim \mathcal{WN}(0, \text{diag}\{\sigma_{\text{NN}}^2, \sigma_{\text{NE}}^2, \sigma_u^2\})$ . These distributional assumptions, in turn, imply the following key moment conditions for the fundamental and signal processes:

$$\begin{aligned} \text{Var}(a_t) &= \sigma_{\text{NN}}^2 + (1 + \theta^2)\sigma_{\text{NE}}^2, & \text{Cov}(a_t, s_{t-1}) &= \sigma_{\text{NE}}^2, \\ \text{Var}(s_t) &= \sigma_{\text{NE}}^2 + \sigma_u^2, & \text{Cov}(a_t, s_{t-2}) &= \text{Cov}(a_t, a_{t-1}) = \theta\sigma_{\text{NE}}^2. \end{aligned}$$

Assume that  $\sigma_{\text{NN}}^2 > 0$ ,  $\sigma_{\text{NE}}^2 > 0$  and  $\sigma_u^2 > 0$  so that the above defines a *mixed news-noise* representation of the fundamental and signal (i.e. with non-trivial non-news, news, and noise shocks). However,  $\theta$  may or may not be equal to zero. As will be shown below, having a fundamental process with  $\theta \neq 0$  plays an important role in the identification of an econometric model based on the mixed news-noise representation specified above.

To this end, assume the econometrician observes the fundamental  $a_t$  along with two variables  $y_t$  and  $z_t$  given by:

$$y_t = K_{ya}a_t + K_{ys}s_t, \quad (3)$$

$$z_t = K_{za}a_t + K_{zs}s_t + K_{z\epsilon}\epsilon_t^{\text{NN}}, \quad (4)$$

where  $K_{ys} \neq 0$  and  $K_{zs} \neq 0$ . The  $(3 \times 1)$  observed data vector  $(a_t, y_t, z_t)'$  admits the structural moving average representation:

$$\begin{aligned} \begin{pmatrix} a_t \\ y_t \\ z_t \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ K_{ya} & K_{ys} & K_{ys} \\ K_{za} + K_{z\epsilon} & K_{zs} & K_{zs} \end{pmatrix} \begin{pmatrix} \epsilon_t^{\text{NN}} \\ \epsilon_t^{\text{NE}} \\ u_t \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & 1 & 0 \\ 0 & K_{ys} & 0 \\ 0 & K_{zs} & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{t-1}^{\text{NN}} \\ \epsilon_{t-1}^{\text{NE}} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & \theta & 0 \\ 0 & K_{ys}\theta & 0 \\ 0 & K_{zs}\theta & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{t-2}^{\text{NN}} \\ \epsilon_{t-2}^{\text{NE}} \\ u_{t-2} \end{pmatrix} \\ &= \begin{pmatrix} 1 & L + \theta L^2 & 0 \\ K_{ya} & K_{ys} + K_{ya}(L + \theta L^2) & K_{ys} \\ K_{za} + K_{z\epsilon} & K_{zs} + K_{za}(L + \theta L^2) & K_{zs} \end{pmatrix} \begin{pmatrix} \epsilon_t^{\text{NN}} \\ \epsilon_t^{\text{NE}} \\ u_t \end{pmatrix} = \mathbf{K}(L)\epsilon_t. \quad (5) \end{aligned}$$

Two important features of the impulse response functions embedded in (5) are noteworthy. Firstly, the responses of  $y_t$  and  $z_t$  to news and noise shocks at time  $t$  are identical. As noted by Forni, Gambetti, Lippi, and Sala (2017) and discussed in detail by Benati, Chan, Eisenstat, and Koop (2020), this is a general feature of the signal extraction problem faced by agents in noisy news settings because they are not able to distinguish between news and noise immediately when receiving a signal  $s_t$  about the fundamental. Secondly, all three variables continue to respond to news shocks in  $t + 1$  and  $t + 2$  but none of the variables respond to the noise shock for  $t > 0$ . That is, the impulse response functions to news and noise shocks diverge in periods following initial impact. Our Assumption 4 below captures these features for the general case.

Another important observation related to (5) is the role of  $K_{z\epsilon}$ . In particular, it is straightforward to verify that  $\det \mathbf{K}(L) = (L + \theta L^2)K_{ys}K_{z\epsilon}$ , so that  $\det \mathbf{K}(L) = 0$  if and only if  $K_{z\epsilon} = 0$ . Importantly, this means that if  $K_{z\epsilon} \neq 0$ , the *three* observed variables  $(a_t, y_t, z_t)$  *must be* driven by *at least three* shocks. This highlights issue 3 listed in the introduction. It implies that econometric models involving non-news, news and noise shocks have certain properties, which in the context of the present example can be characterized as follows.

Theoretical models that assume agents' information sets consist solely of the fundamental and signal systematically impose  $K_{z\epsilon} = 0$ , since every endogenous variable (such as  $y_t$  and  $z_t$ ) must lie *strictly* in the space spanned by  $\{a_{t-l}, s_{t-l} : l \geq 0\}$ . Under this assumption, however,  $\det \mathbf{K}(L) = 0$  in the present example, implying that (i) the spectral density for the observed data  $(a_t, y_t, z_t)$  is *singular* and (ii) the three shocks  $(\epsilon_t^{\text{NN}}, \epsilon_t^{\text{NE}}, u_t)$  are *not recoverable* from present, past and future observations of  $(a_t, y_t, z_t)$ . Essentially,  $K_{z\epsilon} = 0$  if and only if the three observables are driven by *two shocks or less*.

In contrast, time series econometric methods generally rely on the assumption that  $\det \mathbf{K}(L) \neq 0$ , with SVAR methods further requiring  $\det \mathbf{K}(z) \neq 0$  for all  $|z| \leq 1, z \in \mathbb{C}$ . Consequently, with  $K_{z\epsilon} = 0$ , to estimate the system in (5) consisting of parameters  $K_{ya}, K_{ys}, K_{za}, K_{zs}, \theta, \sigma_{\text{NN}}^2, \sigma_{\text{NE}}^2$  and  $\sigma_u^2$ , would require an entirely new set of tools: neither VAR nor VARMA methods are directly applicable.

The approach in this paper builds on the observation that with  $K_{z\epsilon} \neq 0$  the system in (5) can be efficiently estimated and the restriction  $K_{z\epsilon} = 0$  can be tested empirically. To this end, we propose a Bayesian VARMA methodology and test based on the Savage-Dickey density ratio. Identification of  $K_{ya}, K_{ys}, K_{za}, K_{zs}, K_{z\epsilon}, \theta, \sigma_{\text{NN}}^2, \sigma_{\text{NE}}^2$  and  $\sigma_u^2$ ,

when  $K_{z\epsilon} \neq 0$ , follows from  $\theta \neq 0$  and the moment conditions:

$$\begin{aligned}
\text{Var}(a_t) &= \sigma_{\text{NN}}^2 + (1 + \theta^2)\sigma_{\text{NE}}^2, \\
\text{Cov}(a_t, a_{t-1}) &= \theta\sigma_{\text{NE}}^2, \\
\text{Cov}(a_t, y_{t-1}) &= K_{ya}\text{Cov}(a_t, a_{t-1}) + K_{ys}\sigma_{\text{NE}}^2, \\
\text{Cov}(a_t, z_{t-1}) &= K_{za}\text{Cov}(a_t, a_{t-1}) + K_{ys}\sigma_{\text{NE}}^2 + K_{z\epsilon}\sigma_{\text{NN}}^2, \\
\text{Cov}(a_t, y_{t-2}) &= K_{ys}\text{Cov}(a_t, a_{t-1}), \\
\text{Cov}(a_t, z_{t-2}) &= K_{zs}\text{Cov}(a_t, a_{t-1}), \\
\text{Cov}(a_{t-1}, y_t) &= K_{ya}\text{Cov}(a_t, a_{t-1}), \\
\text{Cov}(a_{t-1}, z_t) &= K_{za}\text{Cov}(a_t, a_{t-1}), \\
\text{Cov}(y_t, z_t) &= K_{ya}\text{Cov}(a_t, z_t) + K_{ys}K_{zs}\sigma_{\text{NE}}^2 + K_{ys}K_{zs}\sigma_u^2.
\end{aligned}$$

However, setting  $\theta = 0$  forces  $\text{Cov}(a_t, a_{t-1}) = 0$ , and with it, obliterates the moments  $\text{Cov}(a_t, y_{t-2})$ ,  $\text{Cov}(a_t, z_{t-2})$ ,  $\text{Cov}(a_{t-1}, y_t)$ , and  $\text{Cov}(a_{t-1}, z_t)$ . As a result, it can be easily shown that the remaining unrestricted moments are not sufficient to recover  $K_{ya}$ ,  $K_{ys}$ ,  $K_{za}$ ,  $K_{zs}$ ,  $K_{z\epsilon}$ ,  $\sigma_{\text{NN}}^2$ ,  $\sigma_{\text{NE}}^2$  and  $\sigma_u^2$ . Intuitively, identification of (5) requires that the fundamental process embodies sufficiently rich dynamics to exploit the information in inter-temporal moments of the observed variables.

Observe further that with  $\theta \neq 0$ , all parameters in (5) are identified even when  $K_{z\epsilon} = 0$  and thus  $\det \mathbf{K}(L) = 0$ . In particular, this yields a case where the shocks are *not recoverable* from (present, past, and future) observables, but the impulse response functions are nevertheless *identified*. Benati, Chan, Eisenstat, and Koop (2020, Appendix I) demonstrate that if the fundamental  $a_t$  is observed along with its rational expectations  $E_t(a_{t+1}), E_t(a_{t+2}), \dots$  (as assumed in, e.g., Chahrour and Jurado, 2018a), then *mixed representation* (involving the three shocks: non-news, news, and noise) based on (1) and (2), with  $\theta \neq 0$ , *cannot* be observationally equivalent to either a *pure news representation* (involving only news and non-news shocks) or a *pure noise representation* (involving only non-news and noise shocks).<sup>2</sup> The key to this result is that with  $\theta \neq 0$ , the observation  $E_t(a_{t+2})$  is not trivial and generates enough information to distinguish the mixed representation from the other two. These derivations are reproduced (for convenience) in Appendix C.

Although such simple assumptions about fundamental processes do not clearly extend to the general case in identifying IRFs under non-recoverability, the basic concept that IRFs may be identified does prevail even when the shocks are not recoverable. What is needed to identify IRFs in the more general setting is a set of observables that *recover* the signal.

Specifically, if the signal  $s_t$  is observed by the econometrician, then even with  $\theta = 0$ , the parameters  $\sigma_{\text{NN}}^2$ ,  $\sigma_{\text{NE}}^2$  and  $\sigma_u^2$  are identified from the moments  $\text{Var}(a_t)$ ,  $\text{Var}(s_t)$  and  $\text{Cov}(a_t, s_{t-1})$ ; with  $K_{z\epsilon} = 0$ , the remaining parameters in (5) follow from  $\text{Cov}(a_t, y_t)$ ,  $\text{Cov}(a_t, y_{t-1})$ ,  $\text{Cov}(a_t, z_t)$  and  $\text{Cov}(a_t, z_{t-1})$ . When  $s_t$  is not observed directly, there may exist a set of variables that reveal it. For example, this will be the case if either

---

<sup>2</sup>Note, however, that the pure news and pure noise representations are indeed observationally equivalent (as shown by Chahrour and Jurado, 2018a, Proposition 1); but neither of these is observationally equivalent to the mixed representation as long as  $\theta \neq 0$ .

$K_{ys} = 1$ ,  $K_{zs} = 1$  or  $K_{ys} + K_{zs} = 1$ . The concept of observed variables revealing the signal was proposed by Forni, Gambetti, Lippi, and Sala (2017), who suggested stock prices or survey measures of consumer sentiment as suitable candidates. We formalize this idea in subsection 4.3 and use it to provide a general identification result.

## 4 Identification Theory

In this section, we extend the ideas introduced in Section 3 to a general class of models. Following the noisy-news literature, we are going to assume there are at least three structural shocks:  $\epsilon_{1,t}$  is a conventional (normalized to have standard deviation one) non-news shock to a fundamental variable such as productivity,  $\epsilon_{2,t}$  is the (normalized) news shock and  $\epsilon_{3,t}$  is the (normalized) noise shock.

The theory which underlies the separate identification of these shocks involves two aspects. Firstly, news shocks, by definition, are about the future. They have no immediate effect on the fundamental variable, but reflect information, e.g., about future total factor productivity (TFP). As a simple example, consider a news shock as an announcement by a major company such as Tesla about a future project or technological development. This has no impact on Tesla's production now, and will only have an impact in the future. The second aspect is that news may be noisy. It is only as time passes that we will know whether the announced project or technological development actually occurs (i.e. whether the announcement is news or noise). In a structural model, these theoretical considerations place restrictions on impulse responses at impact. They imply that  $\epsilon_{2,t}$  and  $\epsilon_{3,t}$  have no immediate effect on the fundamental process and that they have the same impact on all variables in the model for some period (i.e. news and noise can only be disentangled as time passes).

We now define our set of identifying assumptions and explain why they are consistent with these considerations. Suppose that an  $n \times 1$  vector of observable macroeconomic variables in levels admits a structural moving average representation of the form

$$\mathbf{y}_t = \mathbf{K}(L)\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I}_n). \quad (6)$$

We assume that  $\mathbf{K}(L)$  satisfies the following.

**Assumption 1.**  $\mathbf{K}(L) = \mathbf{K}_0 + \mathbf{K}_1L + \dots + \mathbf{K}_lL^l + \dots$  is a  $n \times n$  matrix with entries  $K_{ij}(L)$  being rational functions in  $L$  over the reals. In addition,  $\mathbf{K}(L)$  has normal rank  $n$ , i.e.  $\text{rank } \mathbf{K}(z) = n$  almost everywhere for  $z \in \mathbb{C}$ .

**Assumption 2.** Let  $\kappa_p(L)$  denote the poles polynomial of  $\mathbf{K}(L)$  and  $\kappa_z(L)$  the zeros polynomial.  $\kappa_p(L)$  and  $\kappa_z(L)$  have the following properties:

1.  $\kappa_p(z) \neq 0$  for all  $|z| < 1, z \in \mathbb{C}$ ;
2.  $\kappa_z(L) = L^\tau \kappa_\omega(L)$ , where  $\tau > 0$  is an integer and  $\omega = (\omega_1, \dots, \omega_m)$  is a multi-index with  $\omega_h \in \{0, 1\}$ . Let  $\mathbb{A}_\omega = \{\alpha_1, \dots, \alpha_m\}$  with  $\alpha_h \in \{z \in \mathbb{C} : z \neq 0\}$  such that  $|\alpha_h| \geq 1$  iff  $\omega_h = 0$ ,  $|\alpha_h| < 1$  iff  $\omega_h = 1$ , and  $\alpha_h \neq \alpha_k$  for all  $h \neq k$ . Then,  $\kappa_\omega(z) = 0$  iff  $z \in \mathbb{A}_\omega$ .

**Assumption 3.**  $K_{l,12} = 0$  for  $l = 0, \dots, \tau - 1$  and  $K_{0,11} \neq 0$ ;  $K_{1j}(L) = 0$  for all  $j = 3, \dots, n$ .

**Assumption 4.** *there exists a constant  $c > 0$  such that  $K_{l,i3} = cK_{l,i2}$  for  $l = 0, \dots, \tau-1$  and  $i = 2, \dots, n$ .*

**Assumption 5.** *Let  $\chi(L)$  be the least common multiple of the denominators of  $K_{11}(L)$  and  $K_{12}(L)$ , such that  $\hat{K}_{11}(L) = \chi(L)K_{11}(L)$  and  $\hat{K}_{12}(L) = \chi(L)K_{12}(L)$  are polynomials. Then, there exists a constant  $\zeta \neq 0$  such that  $\hat{K}_{12}(\zeta) = 0$  and  $\hat{K}_{11}(\zeta) \neq 0$ .*

**Proposition 1.** *Let  $\mathbf{K}(L)\boldsymbol{\epsilon}_t$  and  $\tilde{\mathbf{K}}(L)\tilde{\boldsymbol{\epsilon}}_t$  be two observationally equivalent representations and let  $\boldsymbol{\Gamma}$  be a  $n \times n$  constant orthogonal matrix partitioned as*

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} \end{pmatrix},$$

*where  $\boldsymbol{\Gamma}_{11}$  is  $3 \times 3$  and  $\boldsymbol{\Gamma}_{22}$  is  $(n-3) \times (n-3)$ . If  $\mathbf{K}(L)$  and  $\tilde{\mathbf{K}}(L)$  both satisfy Assumptions 1-5 for a given  $\omega$  and  $\tau$ , then  $\tilde{\mathbf{K}}(L) = \mathbf{K}(L)\boldsymbol{\Gamma}$ ,  $\boldsymbol{\Gamma}_{12} = \boldsymbol{\Gamma}_{21} = 0$  and  $\boldsymbol{\Gamma}_{11}$  is diagonal, almost everywhere.*

The proof is in Appendix A.

## 4.1 Discussion

Proposition 1 states that under Assumptions 1-5, impulse responses to  $\epsilon_{1,t}$ ,  $\epsilon_{2,t}$ , and  $\epsilon_{3,t}$  are uniquely determined (up to sign normalization) by the observable process  $\{\mathbf{y}_t; t \in \mathbb{Z}\}$ . Each of these assumptions warrants further discussion.

### 4.1.1 Assumptions 3-5

Assumptions 3-5 are theory-driven restrictions that define and identify  $\epsilon_{1,t}$  as the (normalized) non-news shock,  $\epsilon_{2,t}$  as the (normalized) news shock, and  $\epsilon_{3,t}$  as the (normalized) noise shock. In this respect, Assumption 4 reflects the idea that agents respond to (unnormalized) news and noise shocks identically on impact and for the duration of the anticipation horizon (i.e. periods  $0, \dots, \tau-1$ ), which we have argued is a general feature of theoretical models that include noisy signals about future realizations of a fundamental process.

Assumptions 3 and 5 define a class of models for the fundamental process (e.g. TFP) itself. The restrictions in Assumption 3 state that only the non-news shock  $\epsilon_{1,t}$  can influence the fundamental on impact, while only non-news and news shocks ( $\epsilon_{1,t}$  and  $\epsilon_{2,t}$ ) can influence the fundamental at all horizons after impact. These restrictions are standard in the literature; they are consistent with, e.g., the Barsky and Sims (2011) restrictions that non-news and news explain all the forecast error variance of TFP at all horizons, but only non-news affects TFP on impact. In practice, we do not impose all exclusion restrictions exactly, but (similar to Barsky and Sims, 2011) rely on the approximate restriction that non-news and news explain the maximum fraction of the forecast error variance of the fundamental at some large, but finite horizon. This is discussed in more detail in Section 5.

Assumption 5, however, constrains the fundamental process in a way that requires some justification. To highlight the intuition, observe first that the rationality of  $\mathbf{K}(L)$

implies the fundamental process is an ARMA of the form:

$$\chi(L)a_t = \widehat{K}_{11}(L)\epsilon_{1,t} + \widehat{K}_{12}(L)\epsilon_{2,t}, \quad (7)$$

where  $\chi(L)$  is the least common multiple of the denominators of  $K_{11}(L)$  and  $K_{12}(L)$ .

Assumption 5 then requires that (i)  $\widehat{K}_{12}(L)$  has at least one nonzero root  $\zeta$ , and (ii) that this root is not also a root of  $\widehat{K}_{11}(L)$ . Intuitively, it means that the news shock  $\epsilon_{2,t}$  affects the fundamental in a way that is sufficiently distinct from the effect of the non-news shock  $\epsilon_{1,t}$ . Accordingly, the observed fundamental exhibits sufficiently rich dynamics to disentangle non-news, news, and noise shocks.

Following the discussion in Section 3, a simple specification such as

$$a_t = a_{t-1} + \sigma_{\text{NN}}\epsilon_{1,t} + \sigma_{\text{NE}}\epsilon_{2,t-1}$$

will not satisfy Assumption 5, and this will undermine the identification of the desired impulse response functions in a general setting. Allowing for an additional lag of the news shock to affect the fundamental, however, rectifies this deficiency:

$$a_t = a_{t-1} + \sigma_{\text{NN}}\epsilon_{1,t} + \sigma_{\text{NE}}\epsilon_{2,t-1} + \theta\sigma_{\text{NE}}\epsilon_{2,t-2}.$$

In particular, the extended process is characterized by  $\widehat{K}_{11}(L) = \sigma_{\text{NN}}$  and  $\widehat{K}_{12}(L) = \sigma_{\text{NE}}(1 + \theta L)L$ , so that  $\widehat{K}_{12}(-1/\theta) = 0$  but  $\widehat{K}_{11}(-1/\theta) = \sigma_{\text{NN}} \neq 0$ .

Another example of a fundamental process that satisfies Assumptions 3 and 5 is the following:

$$\begin{aligned} a_t &= a_{t-1} + \eta_t + \mu_{t-1}, \\ \eta_t &= \rho_\eta \eta_{t-1} + \sigma_{\text{NN}}\epsilon_{1,t}, \\ \mu_t &= \rho_\mu \mu_{t-1} + \sigma_{\text{NE}}\epsilon_{2,t}, \end{aligned}$$

where  $0 < \rho_\eta < 1$ ,  $0 < \rho_\mu < 1$  and  $\rho_\eta \neq \rho_\mu$ . This leads to

$$(1 - \rho_\mu L)(1 - \rho_\eta L)(1 - L)a_t = \sigma_{\text{NN}}(1 - \rho_\mu L)\epsilon_{1,t} + \sigma_{\text{NE}}(1 - \rho_\eta L)L\epsilon_{2,t},$$

where  $\widehat{K}_{12}(1/\rho_\eta) = 0$  but  $\widehat{K}_{11}(1/\rho_\eta) \neq 0$  as long as  $\rho_\eta \neq \rho_\mu$ . Therefore, Assumption 5 is not restrictive in the sense that it will be satisfied by many reasonable specifications for the fundamental process, whereas only a small group of overly-simplistic processes is excluded.

#### 4.1.2 Assumptions 1-2

Assumption 1 establishes the basic structural framework in which we draw inference about non-news, news, and noise shocks from macroeconomic time-series data. In most macroeconomic applications, this type of assumption is taken as standard. The full rank assumption for  $\mathbf{K}(L)$  implies that all structural shocks can be computed from past, current and future observations of the data (see Chan and Ho, 2004, Lemma 2 or Chahrour and Jurado, 2018b, Theorem 1). However, it is at odds with typical assumptions made in the theoretical noisy-news literature which imply particular reduced rank

forms for  $\mathbf{K}(L)$ . In sub-section 4.2 we develop a simple procedure to test such reduced rank assumptions in the context of our identified, full rank system.

Assumption 2 establishes the stability and fundamentalness of the system in terms of the roots characterizing the poles and zeros polynomials. In particular, the assumption that the poles polynomial  $\kappa_p(L)$  has no roots of modulus less than unity ensures that the observed process is stable (though possibly co-integrated).

The zeros polynomial  $\kappa_z(L)$ , on the other hand, has one root at zero (with multiplicity  $\tau$ ) and all finite, non-zero roots located inside / outside the unit circle according to the ones and zeros pattern in  $\omega$ . The root at zero reflects the fact that a news shock only impacts the fundamental with a delay of  $\tau$  periods, so that in the structural representation, the first  $\tau - 1$  impulse response matrices  $\mathbf{K}_0, \dots, \mathbf{K}_{\tau-1}$  are singular.

The location of the nonzero roots carries important implications for the class of theoretical models our empirical framework identifies. Before discussing these in more detail, it is important to emphasize that the nonzero roots (and therefore the structural representation) are only unique up to a fixed  $\omega$ .

In particular, an alternative system corresponding to  $\tilde{\omega}$  with  $|\tilde{\omega}_h - \omega_h| = 1$  will have a root of  $\kappa_z(L)$  at

$$\tilde{\alpha}_h = \frac{1}{\alpha_h}.$$

More generally, a representation corresponding to  $\omega$  may be transformed to obtain an alternative representation corresponding to  $\tilde{\omega}$ , with some combination of roots of the zeros polynomial “flipped.” Because this transformation can always be achieved without changing the second-order properties of the data, different structural representations corresponding to different  $\omega$  are observationally equivalent in any Gaussian setting.

In terms of the underlying theoretical models, one can regard  $\omega$  as being related to the speed with which agents learn about various shocks in the system. For example, specifying  $\omega = \{0, \dots, 0\}$  implies (6) can be rewritten as

$$\mathbf{D}(L)\mathbf{y}_{t+1} = \boldsymbol{\epsilon}_t,$$

where  $\mathbf{D}(L) = \mathbf{D}_0 + \mathbf{D}_1L + \dots + \mathbf{D}_lL^l + \dots$ . Therefore, when all roots of  $\kappa_\omega(L)$  are outside (or on) the unit circle, all shocks at time  $t$  are recovered from linear combinations of  $\mathbf{y}_{t+1}, \mathbf{y}_t, \dots$ .

While this does not imply  $\boldsymbol{\epsilon}_t$  is included in agents’ time- $(t + 1)$  information set, as some variables in  $\mathbf{y}_{t+1}$  (such as consumption, investment, etc.) result from actions taken by agents at time  $t + 1$ , it does mean that  $\boldsymbol{\epsilon}_t$  potentially enters agents’ time- $(t + 2)$  information set. Then, agents would be fully informed about all shocks including news and noise shocks that occurred two periods prior.

At the opposite end of the spectrum, specifying  $\omega = \{1, \dots, 1\}$  yields

$$\mathbf{D}(L)\mathbf{y}_{t+1} = \boldsymbol{\epsilon}_t,$$

where  $\mathbf{D}(L) = \mathbf{D}_0 + \mathbf{D}_1L^{-1} + \dots + \mathbf{D}_lL^{-l} + \dots$  and  $L^{-1}$  denotes the forward operator. In this case, when all roots of  $\kappa_\omega(L)$  are inside the unit circle, time- $t$  structural shocks can be recovered only from future observations of the variables. Thus, agents learn about news and noise shocks slowly and become fully informed only at the infinite horizon.

For all intermediate cases, linear combinations of both past and future observations of economic variables reveal time- $t$  shocks, and the speed with which agents learn about them depends on how many finite, non-zero roots of  $\kappa_z(L)$  are inside / outside the unit circle.

The complication lies in the choice of  $\omega$ . For any process  $\{\mathbf{y}_t; t \in \mathbb{Z}\}$ , we can obtain a countable, finite set of  $2^m$  observationally equivalent representations satisfying Assumption 2. As  $m$  grows linearly with  $n$ , this set grows with the number of variables in the model. Point identification requires selecting one of these  $2^m$  representations by fixing  $\omega$ , but without clear justification, any particular choice runs the risk of inducing arbitrary inference. In this paper, we deal with this issue as follows.

For point identification, we consider the two extreme cases in the context of agent learning. The first, labelled  $\mathcal{R}_2$ , is based on  $\omega = \{0, \dots, 0\}$  and corresponds to all shocks being revealed to agents after at most two periods. Hence, this scenario assumes the fastest possible learning. The second, labelled  $\mathcal{R}_\infty$ , is based on  $\omega = \{1, \dots, 1\}$  and corresponds to all shocks being revealed to agents only at the infinite horizon. Accordingly, this scenario assumes slowest possible learning.

In addition, we obtain inference from set-identified IRFs and FEVDs. In particular, our Bayesian approach provides a straightforward method to compute posterior distributions of IRFs and FEVDs of interest without the need to fix  $\omega$ . Instead, we parameterize the set of possible representations with a parameter whose support is a set of integers, i.e.  $\varrho \in \{1, \dots, 2^m\}$ , such that  $\omega_\varrho$  identifies a particular combination of “flipped roots.” We then assign a uniform prior to  $\varrho$  subject to the restriction that  $\omega_\varrho$  leads to a valid representation (i.e., all flipped roots are finite and either real or complex conjugate pairs). In Appendix B, we provide the details of a Markov Chain Monte Carlo (MCMC) algorithm for generating draws of all the parameters, including  $\varrho$ . In essence, these draws are from a posterior that is averaged over all possible representations, and therefore, yields valid set-identified inference.

Note that this approach involves drawing  $\varrho$  as an integer from  $1, \dots, 2^m$  and then computing IRFs / FEVDs for a particular (valid) representation corresponding to  $\omega_\varrho$ . This is straightforward computationally and, thus, inference from set-identified IRFs and FEVDs is feasible even for large models. In our empirical application we work with VARMA(2, 2) models of up to  $n = 20$  variables.

## 4.2 Testing Assumptions About Agents’ Information

In this sub-section we derive an alternative SVARMA representation which is based on the theoretical noisy-news literature. We show that it does not allow for the recovery of news and noise shocks. However, because this SVARMA is a restricted version of our earlier SVARMA we can calculate the Bayes factor comparing the two of them using the Savage-Dickey density ratio (SDDR) thus allowing for a test of which sets of assumptions is preferable.

The theoretical literature on news / noise shocks typically assumes that agents’ information at time  $t$  consists *only* of the space spanned by  $\{a_{t-l}, s_{t-l}\}_{l=0}^\infty$ , that is the current and past observations of the fundamental ( $a_t$ ) and the signal ( $s_t$ ) about the future fundamental which is a combination of news and noise shocks. This implies

that any other observable variable, to the extent that it results from actions taken by agents, must be a function of fundamentals and signals, but not the individual non-news, news, and noise shocks. Therefore, if we decompose  $\mathbf{y}_t = (y_{1,t}, \mathbf{y}'_{2,t})'$  such that  $y_{1,t} \equiv a_t$  is the fundamental and  $\mathbf{y}_{2,t}$  are all other variables observed by the econometrician, then denoting the signal by  $s_t$  we obtain

$$y_{1,t} = K_{11}(L)\epsilon_{1,t} + K_{12}(L)\epsilon_{2,t}, \quad (8)$$

$$\mathbf{y}_{2,t} = \mathbf{d}(L)y_{1,t} + \mathbf{e}(L)s_t + \mathbf{K}_{24}(L)\boldsymbol{\epsilon}_{4,t}, \quad (9)$$

$$s_t = \sigma_{\text{NE}}(L)\epsilon_{2,t} + \sigma_u(L)\epsilon_{3,t} \quad (10)$$

where  $\mathbf{d}(L)$ ,  $\mathbf{e}(L)$ ,  $\sigma_{\text{NE}}(L)$  and  $\sigma_u(L)$  are rational functions in  $L$ .

In consequence, we obtain

$$\begin{aligned} \begin{pmatrix} y_{1,t} \\ \mathbf{y}_{2,t} \end{pmatrix} &= \begin{pmatrix} K_{11}(L) & K_{12}(L) & 0 & 0 \\ \mathbf{K}_{21}(L) & \mathbf{K}_{22}(L) & \mathbf{K}_{23}(L) & \mathbf{K}_{24}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \boldsymbol{\epsilon}_{4,t} \end{pmatrix} \\ &= \begin{pmatrix} K_{11}(L) & K_{12}(L) & 0 & 0 \\ \mathbf{d}(L)K_{11}(L) & \mathbf{d}(L)K_{12}(L) + \mathbf{e}(L)\sigma_{\text{NE}}(L) & \mathbf{e}(L)\sigma_u(L) & \mathbf{K}_{24}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \boldsymbol{\epsilon}_{4,t} \end{pmatrix}. \end{aligned}$$

We shall assume that

$$\text{rank} \begin{pmatrix} \mathbf{d}(L) & \mathbf{e}(L) & \mathbf{K}_{24}(L) \end{pmatrix} = n - 1, \quad (11)$$

i.e., the  $(n - 1) \times (n - 1)$  rational transfer matrix formed by concatenating  $\mathbf{d}(L)$ ,  $\mathbf{e}(L)$  and  $\mathbf{K}_{24}(L)$  is of full rank almost everywhere. Note that for the moment we do not impose any further restrictions on  $\mathbf{d}(L)$ ,  $\mathbf{e}(L)$ ,  $\sigma_{\text{NE}}(L)$  or  $\sigma_u(L)$ , and in addition,  $\mathbf{y}_{2,t}$  may or may not contain the signal itself  $s_t$ . This is because no further structure is necessary for the purposes of the test described in this sub-section. We will return to this issue below.

Given (8)-(10), it is easily seen that with only two observed variables ( $a_t$  and  $s_t$ ) the agents can never fully recover the shocks  $\epsilon_{1,t}$ ,  $\epsilon_{2,t}$  and  $\epsilon_{3,t}$  at any horizon (even at infinity). Moreover, as long as all other variables  $\mathbf{y}_{2,t}$  satisfy (9), observing them perfectly in the future does not help the agents in this respect either. This is at odds with the full rank condition in Assumption 1. Indeed, by assuming  $\text{rank } \mathbf{K}(L) = n$ , we are effectively assuming that some variables in  $\mathbf{y}_{2,t}$  respond to non-news, news, and/or noise shocks above and beyond their effect on signals and fundamentals. This immediately implies that at some point in time, agents are able to use all observable data to recover exactly the shocks that occurred in the past. They may learn slowly or quickly, according to the structure of zeros in  $\mathbf{K}(L)$ , as discussed in Section 4.1.2; but  $\epsilon_{1,t}$ ,  $\epsilon_{2,t}$  and  $\epsilon_{3,t}$  are fully revealed at the infinite horizon.

Within our framework, the hypothesis that agents never fully recover the non-news, news, and noise shocks is a testable one. In particular, Proposition 1 establishes that under the alternative hypothesis, all coefficients in  $\mathbf{K}(L)$  are identified (up to sign),

while non-recoverability can be represented by restrictions on the elements of  $\mathbf{K}(L)$  as follows.

Since  $\mathbf{d}(L) = \mathbf{K}_{21}(L)/K_{11}(L)$  and  $\mathbf{e}(L) = \mathbf{K}_{23}(L)/\sigma_u(L)$  under non-recoverability, we obtain

$$\mathbf{K}_{22}(L) - \mathbf{K}_{21}(L) \frac{K_{12}(L)}{K_{11}(L)} = \mathbf{K}_{23}(L) \frac{\sigma_{\text{NE}}(L)}{\sigma_u(L)}. \quad (12)$$

This imposes  $n - 2$  equality constraints on the elements of  $\mathbf{K}(L)$ . Specifically, let

$$\mathbf{l}(L) = \mathbf{K}_{22}(L)K_{11}(L) - \mathbf{K}_{21}(L)K_{12}(L)$$

and let  $\mathbf{V}(z)$  be the  $(n - 1) \times (n - 2)$  matrix satisfying

$$\mathbf{V}(z)' \mathbf{K}_{23}(z) = 0, \quad \mathbf{V}(z)' \mathbf{V}(z) = \mathbf{I}_{n-2},$$

for  $z \in \mathbb{C}$ . Then defining

$$\mathbf{m}(z) = \frac{\mathbf{V}(z)' \mathbf{l}(z)}{\|\mathbf{l}(z)\|}$$

it follows that (12) can be equivalently stated as<sup>3</sup>

$$\mathbf{m}(z) = 0 \text{ for all } z \in \mathbb{C}. \quad (13)$$

A Bayesian approach to testing for (13) can be developed using the SDDR. Suppose the researcher is interested in a restriction such as (13). The SDDR is the ratio of the posterior of the unrestricted model evaluated at the restriction to the prior evaluated at the same restriction. Under weak conditions, this can be shown to be equal to the Bayes factor comparing the unrestricted and restricted models. The main condition, which we adopt in this paper, is that, for parameters that appear in both restricted and unrestricted models, their priors agree in the sense that the prior in the restricted model equals the prior in the unrestricted model conditional on the restriction holding. See Verdinelli and Wasserman (1995) for details. The SDDR can be calculated using a posterior and prior simulator for the unrestricted model. It is also conceptually appealing in this case because it does not require parameters to be identified in the restricted model (e.g., Koop and Potter, 1999).

In terms of model selection, the Bayes factor in favor of the unrestricted model 1 compared to the restricted model 0, denoted as  $\text{BF}_{10}$ , has a straightforward interpretation. Suppose  $\text{BF}_{10} > 1$ . Then, we infer that the data generally prefers the unrestricted model over *any* model satisfying (13). Taking this one step further, the actual value of  $\text{BF}_{10}$  maybe interpreted as the prior odds one would need to place in favor of a restricted model in order to equate the two types of models from a decision-theoretic point of view. For example, if  $\text{BF}_{10} = 10$ , we would conclude that a researcher must prefer *a priori* some model satisfying (13) to the unrestricted model with a ratio of 10 to 1 in order to be indifferent between the two *a posteriori*. In general, a larger  $\text{BF}_{10}$  is directly related to the amount of evidence in favor of the unrestricted model that is embodied by the data. Conversely,  $\text{BF}_{10} < 1$  may be interpreted as evidence

---

<sup>3</sup>Dividing by  $\|\mathbf{l}(z)\|$  normalizes  $\mathbf{m}(z)$  such that each element is bounded:  $-1 \leq m_i(z) \leq 1$ ; this is useful for numerical stability of the simulation algorithm.

that there exists some model satisfying the non-recoverability hypothesis, which fits the data better than the unrestricted model. In this case, however, the SDDR based approach does not provide any insight regarding the type of restricted model that may be more appropriate. To fully ascertain this alternative, one would need to also estimate a particular model satisfying (13).

To that end, we conclude this section by describing one unique representation for a class of models satisfying (13). We postulate that our estimation algorithm for our full rank SVARMA can be extended to a reduced-rank setting. Thus, it provides a practical methodology for drawing inference from this representation as well. However, we do not pursue such a methodology in this paper, focusing instead on full rank systems satisfying Assumptions 1-5. In our empirical work, we find that US macroeconomic data strongly favors full rank systems. Hence, we leave the details of methodological extensions to reduced-rank systems satisfying (13) for future research.

### 4.3 Uniqueness Under Non-Recoverability

Following Forni, Gambetti, Lippi, and Sala (2017), let us now assume that the signal  $s_t$  observed by the agents takes the form

$$s_t = \sigma_{\text{NE}} \epsilon_{2,t} + \sigma_u \epsilon_{3,t}, \quad (14)$$

along with the normalization  $|K_{\tau^*,12}| = \sigma_{\text{NE}}$  for some  $\tau^* \geq \tau$ . In addition, a subset of the observed variables  $\mathbf{y}_{2,t}$  contain variables that *reveal* the signal  $s_t$ .

Forni, Gambetti, Lippi, and Sala (2017) considered the existence of one such variable in their empirical work, and suggested stock prices or the University of Michigan sentiment index to be reasonable candidates. In principle, we may regard a number of financial and policy variables as exhibiting this property, which can be exploited for identification. Doing so, however, requires a more formal treatment of the idea that certain variables reveal the signal.

Consequently, let  $\mathbb{R}(z)^n$  denote the  $n$ -dimensional space of rational functions (with argument  $z \in \mathbb{C}$ ) over the reals.

**Definition 1.** *The  $n \times 1$  vector of variables  $\mathbf{y}_t$  reveals the signal  $s_t$  if and only if  $\mathbf{y}_t = \mathbf{a}_t + \mathbf{b}(L)s_t$  and there exists a functional  $\mathbf{g} : \mathbb{R}(z)^n \rightarrow \mathbb{R}$  such that*

1.  *$\mathbf{g}$  is homogeneous of degree 1, i.e.  $\mathbf{g}(\mathbf{f}(z)c) = \mathbf{g}(\mathbf{f}(z))c$  for all  $\mathbf{f}(z) \in \mathbb{R}(z)^n$  and  $c \in \mathbb{R}$ ;*
2.  *$\mathbf{g}(\mathbf{b}(z)) = 1$ .*

In the present setting, the  $(n-1) \times 1$  vector  $\mathbf{y}_{2,t}$  is related to the signal  $s_t$  by (9), and therefore, these variables reveal the signal if  $\mathbf{g}(\mathbf{e}(z)) = 1$  for some  $\mathbf{g}$  satisfying the definition. Of course, if the signal is observed directly (say  $y_{i,t} \equiv s_t$ ), then  $e_i(L) = 1$  and so  $e_i(0) = 1$ . Hence, this can be represented by  $\mathbf{g}(\mathbf{e}(z)) = e_i(0)$ . Another example is

$$\mathbf{g}(\mathbf{e}(z)) = \frac{e_i^{(k)}(0)}{k!} = 1,$$

which corresponds to the  $k$ th coefficient of the power series expansion of  $e_i(z)$  being equal to one. Intuitively, such an assertion states that the signal  $s_t$  is revealed through  $y_{i,t}$  after  $k$  periods, i.e.

$$y_{i,t} = d_i(L)y_{1,t} + \sum_{l=1}^{k-1} e_{l,i}s_{t-l} + s_{t-k} + \sum_{l=k+1}^{\infty} e_{l,i}s_{t-l} + \mathbf{K}_{i4}(L)\epsilon_{4,t}.$$

In this empirical setting, Forni, Gambetti, Lippi, and Sala (2017) force agents to learn fully about shocks by assuming  $K_{11}(L) = 0$  so that the fundamental process reduces to  $a_t = K_{12}(L)\epsilon_{2,t}$ . With only two shocks remaining, agents are able to recover them from the past and future observations of the fundamental and signal, with the speed of learning determined by the structure of non-zero roots in  $K_{12}(L)$ .

However, eliminating non-news by forcing  $K_{11}(L) = 0$  may have important implications on the inference about news and noise shocks obtained from time-series data. That is, without a non-news shock, economic variables cannot be affected by *unanticipated* sources of variation in productivity, and consequently, such effects (if present) are attributed to other shocks in the system, including noise. This may partly explain the substantial fraction of forecast error variance in economic data being explained by noise shocks that Forni, Gambetti, Lippi, and Sala (2017) find.

In departure from Forni, Gambetti, Lippi, and Sala (2017), we assume (as in Barsky and Sims, 2012) that the fundamental is driven by three shocks. In particular, we will assume the fundamental is an ARMA process given by (7), with the additional restrictions that  $\hat{K}_{11}(L)$  and  $\hat{K}_{12}(L)$  are *coprime* polynomials and  $\hat{K}_{11}(L)$  has no roots of modulus less than unity.<sup>4</sup>

As pointed out by Forni, Gambetti, Lippi, and Sala (2017), any system satisfying (7), (9) and (14) implies that agents never fully learn about the true nature of structural shocks, even at the infinite horizon. Since operationally this means shocks cannot be represented as linear combinations of past, present and future observations, regardless of how many endogenous variables are observed, VAR methods are altogether not suitable for quantifying the effects of non-news, news and noise shocks in this setting.

However, one can still obtain a unique structural representation for an observed data process  $\mathbf{y}_t$ . The key is that we can translate the idea of certain variables revealing the signal into identifying restrictions on  $\mathbf{K}(L)$ . Then, combining this with restrictions (9), (14) and (7) gives rise to the following assumptions on  $\mathbf{K}(L)$ , which can be viewed as modified versions of Assumptions 1 and 5.

**Assumption 6.**  $\mathbf{K}(L) = \mathbf{K}_0 + \mathbf{K}_1L + \dots + \mathbf{K}_lL^l + \dots$  is a  $n \times n$  matrix with entries  $K_{ij}(L)$  being rational functions in  $L$  over the reals. In addition,  $\mathbf{K}(L)$  has normal rank  $n - 1$ , i.e.  $\text{rank } \mathbf{K}(z) = n - 1$  almost everywhere for  $z \in \mathbb{C}$ .

**Assumption 7.**  $|K_{\tau^*,12}| = \sigma_{NE}$  for some  $\tau^* \geq \tau$  and the following hold:

1.  $\hat{K}_{11}(L)$  and  $\hat{K}_{12}(L)$  defined in (7) are coprime polynomials and  $\hat{K}_{11}(L)$  has no roots of modulus less than unity.

---

<sup>4</sup>Consistent with Forni, Gambetti, Lippi, and Sala (2017), however, we do not make any assumption regarding the non-zero roots of  $\hat{K}_{12}(L)$ .

2.  $\frac{1}{c}K_{i3}(L) = K_{i2}(L) - K_{i1}(L)\frac{K_{12}(L)}{K_{11}(L)}$  for all  $i \geq 2$ .
3.  $\mathbf{y}_{2,t}$  reveals the signal  $s_t \equiv |K_{\tau^*,12}|(\epsilon_{2,t} + c\epsilon_{3,t})$  such that  $\mathbf{g}(\mathbf{K}_{23}(z)) = c|K_{\tau^*,12}|$  for some  $\mathbf{g}$  satisfying Definition 1, and where  $\mathbf{K}_{23}(z)$  is a  $(n-1) \times 1$  rational vector with entries  $K_{i3}(z)$  for  $i = 2, \dots, n$ .

With these assumptions, the following result holds.

**Proposition 2.** *Let  $\mathbf{K}(L)\epsilon_t$ ,  $\tilde{\mathbf{K}}(L)\tilde{\epsilon}_t$  and  $\mathbf{\Gamma}$  be defined as in Proposition 1. If  $\mathbf{K}(L)$  and  $\tilde{\mathbf{K}}(L)$  both satisfy Assumptions 2-4, 6 and 7 for a given  $\omega$ ,  $\tau$  and  $\tau^*$ , then  $\tilde{\mathbf{K}}(L) = \mathbf{K}(L)\mathbf{\Gamma}$ ,  $\mathbf{\Gamma}_{12} = \mathbf{\Gamma}_{21} = 0$  and  $\mathbf{\Gamma}_{11}$  is diagonal. Moreover,  $\omega = \emptyset$  almost everywhere.*

The proof is in Appendix A. The fact that  $\omega = \emptyset$  for almost all elements of  $\mathbf{K}(L)$  satisfying the assumptions of Proposition 2 implies that identification in this case is *not* subject to the “root flipping” issues discussed sub-section 4.1.2. It follows from the fact that  $\mathbf{K}(L)$  can be written as a product of a  $n \times (n-1)$  matrix, which is zero-free almost everywhere (Anderson and Deistler, 2008), and a  $(n-1) \times n$  matrix which is zero-free by construction.

We further note in passing that it is possible to obtain similar uniqueness results under the non-recoverability hypothesis using alternative assumptions. For instance, the assumption that  $\mathbf{y}_{2,t}$  reveals the signal is sufficient, but not necessary. If it is removed, one would need to formulate different assumptions about what is observable to the econometrician with respect to the agent’s information set. However, a more general development in this direction is beyond the scope of the present paper.

## 5 Methodology

### 5.1 Justification for SVARMA

In this paper we use SVARMAs, instead of the SVARs that dominate the empirical macroeconomic literature. In general, several authors (Cooley and Dwyer, 1998, e.g.) have argued 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. The fundamental concept underlying this argument was formalized by Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson (2006), who characterize the equilibrium of a general class of DSGE models with linear transition laws in a state-space framework, and show how this can be represented by a finite-order VARMA. The structural VAR approach can then be justified as a truncated approximation to the infinite-order VAR obtained by inverting the MA part of the finite-order VARMA.

The inversion/truncation step, however, entails a number of difficulties. Some researchers—see, e.g., Chari, Kehoe, and McGrattan (2008) and Poskitt and Yao (2017)—point out that the errors which arise due to this can potentially be substantial, because the typical lag and series lengths used in practice (e.g.  $p < 5, T < 250$  for quarterly data) are much too short.

Moreover, in many cases such as the permanent-income model of Hansen, Roberds, and Sargent (1991), the resulting VARMA representation entails an MA component that

is not invertible, and therefore, a VAR approximation to such a model does not exist.<sup>5</sup> VARMA representations admitting an infinite-order VAR representation are commonly known as fundamental, while those that do not are referred to as non-fundamental.

From Assumption 1, the systems we work with admit a structural VARMA( $p, q$ ) representation of the form<sup>6</sup>

$$\mathbf{B}(L)\mathbf{y}_t = \mathbf{A}(L)\boldsymbol{\epsilon}_t, \quad (15)$$

where  $\mathbf{A}(L) = \mathbf{A}_0 + \mathbf{A}_1L + \dots + \mathbf{A}_qL^q$ ,  $\mathbf{B}(L) = \mathbf{I}_n + \mathbf{B}_1L + \dots + \mathbf{B}_pL^p$  and  $\mathbf{K}(L) = \mathbf{B}(L)^{-1}\mathbf{A}(L)$ . Observe that the roots of  $\det \mathbf{A}(L)$  are equivalent to the roots of  $\kappa_z(L)$ . However, Assumption 4 implies that (15) is inherently *not fundamental*. This is because the proportionality restriction induces a reduced rank impact matrix  $\mathbf{A}_0 \equiv \mathbf{A}(0)$ . Since  $\det \mathbf{A}(0) = 0$ , it is clear that  $\det \mathbf{A}(z)$ , for  $z \in \mathbb{C}$ , has a root inside the unit circle. This means that  $\mathbf{A}(L)$  is not invertible on the closed unit disk, and therefore, it is not possible to obtain a *structural* VAR representation.

Nevertheless, one may still obtain a fundamental *reduced-form* VARMA representation by an appropriate application of Blaschke filters. Under additional restrictions, one can then derive a *reduced-form* VAR. This is, in fact, the approach undertaken by Forni, Gambetti, Lippi, and Sala (2017). In their setup, a proportionality restriction identical to Assumption 4 is imposed and IRFs / FEVDs are computed from a structural VARMA( $p, \tau$ ), similar to (15). Consequently, their methodology involves first estimating a reduced-form VAR, then transforming it to the structural VARMA representation *ex post*.

The method described below is also based on a similar concept: estimate a fundamental reduced-form representation as a first step, then transform it *ex post* to the desired non-fundamental structural representation. In contrast to Forni, Gambetti, Lippi, and Sala (2017), however, we work directly with VARMA specifications throughout.

As will be discussed in more detail shortly, working with VARs at any stage in our setting is neither appropriate nor necessary. The reduced-form VARMA specification we propose yields both efficient and practically simple algorithms in a Bayesian framework. We use MCMC methods to obtain draws from the posterior of the reduced-form VARMA, then transform each draw to obtain a sample from the posterior of the structural VARMA. In the following subsections we outline basic ideas underpinning this methodology, with technical details provided in Appendix B.

---

<sup>5</sup>Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson (2006) provide an easily verifiable condition under which the VARMA admits an infinite-order VAR representation.

<sup>6</sup>For clarity of exposition, we suppress all deterministic terms as they do not affect any of the theory or methodology discussed in this paper. In practice, we include an intercept in all estimated models.

## 5.2 Econometric Framework

To design an efficient estimation algorithm, it is more practical to work with a reduced-form VARMA specification rather than the structural one in (15). To this end, define

$$\mathbf{C}_1 = \begin{pmatrix} 0 & 0 & 1 \\ \frac{c}{\sqrt{1+c^2}} & \frac{1}{\sqrt{1+c^2}} & 0 \\ -\frac{1}{\sqrt{1+c^2}} & \frac{c}{\sqrt{1+c^2}} & 0 \\ & & \mathbf{I}_{n-3} \end{pmatrix}$$

and let

$$\Phi(L) = \mathbf{K}(L)\mathbf{C}_1 \begin{pmatrix} L^{-\tau} \\ \mathbf{I}_{n-1} \end{pmatrix},$$

where  $L^{-1}$  denotes the forward operator and  $L^{-\tau} = (L^{-1})^\tau$ . Observe that the *zeros* polynomial of  $\Phi(L)$  has no roots at zero and at least one root at infinity. The remaining roots lie inside / outside the unit circle according to  $\omega$  and follow the same pattern as the roots of  $\kappa_\omega(L)$ .

Since  $\Phi_0 \equiv \Phi(0)$  is invertible, we derive a reduced-form VARMA corresponding to (15) as

$$\mathbf{B}(L)\mathbf{y}_t = \Theta(L)\mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(0, \Sigma), \quad (16)$$

with  $\Sigma = \Phi_0\Phi_0'$  and  $\Phi(L) = \mathbf{B}(L)^{-1}\Theta(L)\Phi_0$ . In (16),  $\mathbf{B}(L)$  is the same as in (15) and  $\Theta(L) = \mathbf{I}_n + \Theta_1 L + \dots + \Theta_q L^q$ , where  $q$  is the same polynomial degree as in  $\mathbf{A}(L)$ . Likewise,  $\det \Theta(L)$  has roots that are identical to those of the zeros polynomial of  $\Phi(L)$ , and therefore, the same combination of finite roots inside / outside the unit circle as the roots of  $\kappa_\omega(L)$ .

Hence, we can build an estimation algorithm directly on the reduced form, then transform estimated parameters  $\mathbf{B}_1, \dots, \mathbf{B}_p, \Theta_1, \dots, \Theta_q, \Sigma$  to obtain the structural representation as

$$\mathbf{A}(L) = \Theta(L)\Phi_0 \begin{pmatrix} L^\tau \\ \mathbf{I}_{n-1} \end{pmatrix} \mathbf{C}_1'.$$

In direct parallel to standard SVAR approaches, this step requires  $\Phi_0$  and  $\mathbf{C}_1$  which are not estimated in the reduced-form specification. To obtain these quantities, we derive a set of empirical restrictions that correspond to Assumptions 3-5, as follows:

- R1. the first column of the product  $\Theta_j\Phi_0$  is zero for all  $j = q - \tau + 1, \dots, q$ ;
- R2. the fundamental only responds to non-news shocks on impact;
- R3. the news shock explains the maximum fraction of the forecast error variance of the fundamental at some large (but finite) horizon;
- R4. on impact, the response of all variables to noise shocks is proportional to the response of all variables to news shocks.

The general approach can then be outlined as follows:

- Step 1. Estimate a reduced form VARMA with the restriction  $\text{rank } \Theta_j \leq n - 1$  for all  $j = q - \tau + 1, \dots, q$  enforced.

- Step 2. Compute an arbitrary decomposition (e.g. Cholesky)  $\Theta_0 \Theta_0' = \Sigma$ .
- Step 3. If  $r_\star = \max_{j=q-\tau+1, \dots, q} \{\text{rank } \Theta_j\}$ , let  $\Delta$  be the  $n \times r_\star$  matrix such that  $\Theta_j \Theta_0 \Delta = 0$  and  $\Delta' \Delta = I_{r_\star}$ . Note that  $\Delta$  is unique (up to sign) if  $r_\star = 1$ ; otherwise numerical optimisation may be used to determine a unique  $\Delta$  as described in Appendix B. Set  $\Gamma_0 = (\Delta, \Delta_\perp)$ , such that  $\Gamma_0' \Gamma_0 = \Gamma \Gamma_0' = I_n$  (i.e. a constant orthogonal matrix).
- Step 4. Compute  $\tilde{\Theta}_j = \Theta_j \Gamma_0$  and observe that as long as  $\Delta$  is unique, the first column of  $\tilde{\Theta}_j$  is equivalent to the first column of  $\Theta_j \Phi_0$  for all  $j = 1, \dots, q$ , with the first column of  $\tilde{\Theta}_j$  being zero for all  $j = q - \tau + 1, \dots, q$ . Hence, we now apply the simple Blaschke transformation that shifts in time the first column of each  $\tilde{\Theta}_1, \dots, \tilde{\Theta}_q$ , i.e. for all  $i = 1, \dots, n$ , set

$$\begin{aligned} \tilde{A}_{j,i1} &= 0, & j &= 0, \dots, \tau - 1, \\ \tilde{A}_{j,i1} &= \tilde{\Theta}_{j-\tau, i1}, & j &= \tau, \dots, q. \end{aligned}$$

The resulting polynomial matrix  $\tilde{\mathbf{A}}(L) = \tilde{\mathbf{A}}_0 + \tilde{\mathbf{A}}_1 L + \dots + \tilde{\mathbf{A}}_q L^q$  is such that  $\mathbf{A}(L) = \tilde{\mathbf{A}}(L) \Gamma$ , where  $\Gamma$  is a constant orthogonal matrix.

- Step 5. Compute  $\Gamma$  according to the empirical restrictions R2-R4 above. The details of this procedure are provided in Appendix B.

We conclude this sub-section by noting that it is theoretically possible to obtain a VAR from (16) by setting  $q = \tau$  and  $\Theta_1 = \dots = \Theta_\tau = 0$ . The corresponding SVARMA would have  $\mathbf{A}(L) = \mathbf{A}_0 + \mathbf{A}_1 L + \dots + \mathbf{A}_\tau L^\tau$ , with  $\mathbf{A}_j$ ,  $j \geq 1$ , having all columns equivalently zero except the second and third columns, which are proportional. To the extent that there is no *a priori* reason to justify such restrictions, we prefer to work with the general VARMA specification.

In addition, starting with a VAR would restrict our SVARMA to have all non-zero roots of  $\det \mathbf{A}(L)$  outside the unit circle, which is equivalent to our  $\mathcal{B}_2$  representation discussed in sub-section 4.1.2. Thus, it would only allow us to analyze the effects non-news, news and noise shocks assuming that these three shocks are fully revealed to agents after, at most, two periods. On the other hand, working with a full VARMA allows us to select from the fundamental or one of the basic non-fundamental representations of (16), determined by the location of the roots of  $\det \Theta(L)$ . The number of non-zero roots of  $\det \mathbf{A}(L)$  that are inside the unit circle are consequently determined by the total number of roots of  $\det \Theta(L)$  that are inside the unit circle.

We use Bayesian methods to simulate from the posterior of the reduced-form parameters, then transform these draws to obtain a posterior representation for the IRFs and FEVDs. As outlined in the following sub-section and discussed in extensive detail in Appendix B, efficient simulation methods are readily available for this purpose. Fundamental or non-fundamental representations are determined in the post-processing of draws obtained from the reduced-form VARMA. Alternatively, set identified impulse responses are easily constructed by randomly sampling one of these representations with uniform probabilities, given the fact that fundamental and non-fundamental representations are observationally equivalent under Gaussian errors.

### 5.3 Estimating the VARMA

Despite having many attractive statistical properties, VARMA 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  $\text{VAR}(p)$  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, Justiniano, and Melosi, 2014, use a VAR with  $n = 9$ ).

Adding a VMA component increases this to  $(p + q)n$  making over-parameterization concerns even more relevant. In preceding sections we have focused on the structural identification restrictions needed for  $\mathbf{K}(L) = \mathbf{B}(L)^{-1}\mathbf{A}(L)$  to be unique, but even the reduced form VARMA in (16) suffers from a lack of identification which arises from the possibility of common factors in the VAR and VMA parts of the model. Additional restrictions are typically needed to eliminate these common factors, and both these restrictions as well as the structural identification restrictions are required for the SVARMA in (15) to be unique.

These issues are discussed in detail in Chan, Eisenstat, and Koop (2016). A crucial aspect of this approach is the use of prior shrinkage to mitigate over-parameterization concerns. One important insight is that in a Bayesian setting, unique identification of both  $\mathbf{B}(L)$  and  $\mathbf{A}(L)$  in (15) is not necessary if the primary interest lies only in identifying IRFs (and FEVDs), which are obtained from  $\mathbf{K}(L)$ . In essence, draws from the posterior of structural parameters  $\mathbf{B}_1, \dots, \mathbf{B}_p$  and  $\mathbf{A}_0, \dots, \mathbf{A}_q$  may be obtained even if this joint posterior is not unimodal. Subsequently, these draws may be transformed to draws of the IRFs, whose posterior mode is indeed unique. Therefore, a sensible Bayesian approach to working with structural VARMA imposes structural identification restrictions *exactly*, while using priors to *shrink* the common factors in VAR and VMA parts of the model (rather than eliminating them exactly).

Using hierarchical shrinkage priors, Chan, Eisenstat, and Koop (2016) also develop Bayesian methods for estimating VARMA which work even with large values of  $n$ . We adapt these methods to the VARMA 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 the Appendix B. Here we provide a basic outline of our strategy.

A unique, reduced-form VARMA may be obtained by rewriting (16) in recursive form:

$$\begin{aligned} \tilde{\mathbf{B}}_0 \mathbf{y}_t = & \tilde{\mathbf{B}}_1 \mathbf{y}_{t-1} + \dots + \tilde{\mathbf{B}}_{p^*} \mathbf{y}_{t-p^*} \\ & + \tilde{\mathbf{B}}_0 \mathbf{u}_t + \mathbf{M}_1 \mathbf{u}_{t-1} + \dots + \mathbf{M}_{p^*} \mathbf{u}_{t-p^*}, \quad \mathbf{u}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \end{aligned} \quad (17)$$

where  $p^* = \max\{p, q\}$  and  $\tilde{\mathbf{B}}_0$  is lower triangular with ones on the diagonal. The coefficients in (16) are related to the recursive VARMA in (17) by:

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

where  $\mathbf{M}_{q+1} = \dots = \mathbf{M}_{p^*} = 0$  if  $p > q$  (and similarly for  $\tilde{\mathbf{B}}_j$  if  $p < q$ ). By imposing additional restrictions on the coefficients of this recursive form, the VARMA specification

is rendered unique (Luetkepohl, 2005); in this case, it is referred to as the canonical *echelon* form.

However, estimating the exact echelon form is computationally challenging because the restrictions depend on the lag structure of each equation in the recursive VARMA, which itself needs to be estimated (please refer to Appendix B for more details). In a Bayesian approach, Chan, Eisenstat, and Koop (2016) have shown that approximating the echelon form with shrinkage priors on the coefficients of the recursive form is both efficient and effective. In addition, they showed that 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.

In Appendix B, we show how our VARMA written in (17) can be written equivalently as an expanded-form VARMA. We then use a prior designed to ensure shrinkage and parsimony in this potentially over-parameterized specification. Again, details of the hierarchical shrinkage priors are provided in Appendix B. However, since our structural identification strategy requires  $\text{rank } \Theta_j \leq n - 1$  for all  $j = q - \tau + 1, \dots, q$ , we achieve this by also imposing

$$M_{p^*-\tau+1,ni} = \dots = M_{p^*,ni} = 0, \quad i = 1, \dots, n. \quad (18)$$

Therefore, the last row of  $\mathbf{M}_{p^*-\tau+1}, \dots, \mathbf{M}_{p^*}$  is zero, indicating that the variable ordered last is estimated with  $\tau$  fewer MA lags than the remaining variables. On the other hand, the mean of the  $n$ th variable in the system depends contemporaneously on all other variables through  $\mathbf{B}_0$ .

In our empirical application, we always set *stock price* to be the last variable in the system. The implication of this is that stock price depends on less MA lags than all other variables, once its contemporaneous dependence on the remaining variables is accounted for. Of course, it is always possible to expand the MA lag order of the  $n$ th equation by specifying a large  $p^*$  and allowing the shrinkage priors to address any potential over-parameterisation that results in the other equations.

For the results presented in Section 6, we use  $p^* = 2$  in all estimated models. This approach leads to draws from the SVARMA in (15) with  $p = q = 2$ , but with possibly many coefficients posteriors concentrated around zero. Our evidence also suggests that shrinkage priors are adequate in mitigating over-parameterisation and lead to accurate inference on IRFs and FEVDs, even with models involving as many as 20 variables.

## 5.4 Computing the SDDR

To compute the SDDRs, we need draws of  $\mathbf{B}_1, \dots, \mathbf{B}_p$  and  $\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_q$  from both the prior and posterior distributions. The posterior draws are obtained using the approach outlined in sub-sections 5.2 and 5.3. Draws from the prior (which is implied by the prior on reduced-form parameters detailed in Appendix B) are obtained by sampling directly from the prior specified on the reduced-form parameters, then transforming them to draws of  $\mathbf{B}_1, \dots, \mathbf{B}_p$  and  $\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_q$  using the approach in sub-section 5.2.

For a given  $z \in \mathbb{C}$  and a draw of the structural parameters (from either the prior or

posterior) we compute

$$\mathbf{K}(z) = (\mathbf{I}_n + \mathbf{B}_1 z + \cdots + \mathbf{B}_p z^p)^{-1} (\mathbf{A}_0 + \mathbf{A}_1 z + \cdots + \mathbf{A}_q z^q),$$

and subsequently,  $\mathbf{m}(z)$  in (13) using directly the formulae in sub-section 4.2. Thus, we obtain draws of  $\mathbf{m}(z)$  from either a conditional distribution with density  $p(\mathbf{m}(z) | \mathbf{y})$ , if the structural parameter draws are from the posterior, or a marginal distribution with density  $p(\mathbf{m}(z))$ , if the structural parameter draws are from the prior.

With these draws of  $\mathbf{m}(z)$ , we can evaluate both  $p(\mathbf{m}(z) = 0 | \mathbf{y})$  and  $p(\mathbf{m}(z) | 0)$ . As noted in sub-section 4.2,  $|m_i(z)| \leq 1$  so the density function is guaranteed to exist. Consequently, we may use a non-parametric density estimator to obtain

$$\text{BF}_{10}(z) = \frac{p(\mathbf{m}(z) = 0 | \mathbf{y})}{p(\mathbf{m}(z) = 0)}.$$

Note that the restriction in (13) must hold *for all*  $z \in \mathbb{C}$ . Therefore, evidence in favor of the unrestricted model may be taken as the maximum Bayes factor

$$\text{BF}_{10} = \max_{z \in \mathbb{C}} \text{BF}_{10}(z).$$

In practice, we make two simplifications to facilitate computation. First, we consider a subset  $\mathbb{S} \in \mathbb{C}$  over which to search for the maximum  $\text{BF}_{10}(z)$ ; in our empirical application we use  $\mathbb{S} = (0, 1)$ . Second,  $\mathbf{m}(z)$  is in general a  $(n-2) \times 1$  vector, implying an  $n-2$  dimensional distribution whose density needs to be estimated non-parametrically. For the  $n = 6$  model, this implies four restrictions, and we use a multivariate kernel density estimator to evaluate the four dimensional density functions.<sup>7</sup> For models with  $n > 6$ , to mitigate the curse of dimensionality, we select four of the  $n-2$  restrictions to test.<sup>8</sup> In this respect, the evidence in favor of unrestricted models reported in Table 2 should be regarded as conservative.

To generate the results in Table 2, we proceed by first obtaining a sample of 200 draws of  $z$ , distributed uniformly on the interval  $(0, 1)$ . Then for each  $z$ , we compute  $\text{BF}_{10}(z)$  as described above. Finally, we take  $\text{BF}_{10}$  to be the maximum value of  $\text{BF}_{10}(z)$  attained in this random sample.

## 6 Empirical Application

Our data set includes TFP plus several other macroeconomic variables. We consider SVARMAs of several dimensions. Each SVARMA includes a core set of three variables

---

<sup>7</sup>Specifically, we use Matlab's `mvkdensity` function, which we have found to perform well in this setting. The bandwidths are computed using the formula

$$\text{bw}_i(z) = \frac{1}{0.0675G} \sum_{g=1}^G |m_i^{(g)}(z) - \bar{m}_i(z)| \left( \frac{4}{(n_r + 2)G} \right)^{\frac{1}{n_r + 4}},$$

where  $m_i^{(g)}(z)$  is the  $g$ th draw of the  $i$ th element of  $\mathbf{m}(z)$ ,  $\bar{m}_i(z)$  is the average of the draws,  $G$  is the total number of draws, and  $n_r$  is the number of restrictions that are tested.

<sup>8</sup>For the  $n = 10$  and  $n = 15$  models we use  $(m_2(z), m_4(z), m_5(z), m_6(z))$ ; for the  $n = 20$  model we use  $(m_3(z), m_4(z), m_5(z), m_6(z))$ .

used in most theoretical and empirical models in this literature. These are: TFP, consumption and stock prices. The 6-variable model matches the VAR estimated by Forni, Gambetti, Lippi, and Sala (2017). The remaining models augment these specifications with additional economic information representing agents' actions. Table 1 describes the variables used in each of the models we estimate, where the "X" indicates that the variable is included in the corresponding model. The variables in the table are ordered in the same way that they are included in the estimation algorithm. Notice that S&P 500 Index is ordered last in all models we estimate. The sample period goes from 1959Q4 (when the University of Michigan consumer sentiment index first becomes available) to 2008Q3 (so that we exclude the period during which the Federal Funds rate has been at the zero lower bound). Further details regarding the data are provided in Appendix D.

With the exception of variables that are already indices or rates (i.e. interest rates, inflation rates or indices), all variables in each model are logged. 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)). With regards to lag lengths, we make the relatively conservative choices of  $p = q = 2$ , trusting to the SSVS prior to remove unnecessary lags from the model.

For each SVARMA dimension we calculate three sets of IRFs based on different treatments of the root flipping issue discussed in Sub-section 4.1.2: i)  $\mathcal{R}_2$ , which assumes the fastest possible learning about news and noise shocks; ii)  $\mathcal{R}_\infty$  which assumes the slowest learning; and iii) set identification by averaging over different root configurations.

Before presenting IRFs and FEVDs, we present Table 2 which gives Bayes factors in favor of our identification approach (using Assumption 1) as opposed to the identification restriction used in much of the theoretical literature (the reduced rank assumption) for the two extreme cases  $\mathcal{R}_2$  and  $\mathcal{R}_\infty$  for VARMA of different dimension. It can be seen that, in every single case, the Bayes factors support our Assumption 1. There is also a tendency for the strength of support for our Assumption 1 to increase with dimension. The strongest evidence is found in models with  $n = 20$ .

Table 2 also presents evidence on which VARMA dimension is preferred. In particular, we compute the deviance information criterion (DIC) for each model (see Appendix B for further discussion). All the DICs are computed using the marginal distribution of either the three variables in the  $n = 3$  model or the six variables in the  $n = 6$  case as the likelihood. A model with a smaller DIC value is preferred.<sup>9</sup> It can be seen that there is strong evidence in favor of  $n = 20$ .

---

<sup>9</sup>The results are based on 10 parallel MCMC chains, each of which is of length 50000 (with a burnin of 5000).

Variable Description	$n = 3$	$n = 6$	$n = 10$	$n = 15$	$n = 20$
Total Factor Productivity	X	X	X	X	X
Federal Funds Rate		X	X	X	X
Inflation			X	X	X
Nonfarm Business Hours			X	X	X
Consumption	X	X	X	X	X
Gross Domestic Product		X	X	X	X
Investment		X	X	X	X
Relative Price of Investment			X	X	X
Unemployment Rate				X	X
3M Treasury Rate				X	X
5YR Treasury Rate			X	X	X
3YR Treasury Rate					X
10YR Treasury Rate					X
Vacancy Rate					X
Dividends				X	X
Industrial Production Index				X	X
Real M2					X
USD/GBP Exchange Rate					X
UM Consumer Sentiment Index				X	X
S&P 500 Index	X	X	X	X	X

Table 1: Variables included in each model of varying dimensions. The variables are ordered in the same way that they are included in the estimation algorithm.

FEVDs for the three representations ( $\mathcal{R}_2$ ,  $\mathcal{R}_\infty$  and set identification) are presented in Figures 1, 2 and 3. IRFs are presented in Figures 4, 5 and 6. For the sake of brevity, we only present results for  $n = 20$ , which is the dimension chosen by DIC, and  $n = 6$ , which contains the same variables as Forni, Gambetti, Lippi, and Sala (2017). Estimates for other SVARMA dimensions are given in Appendix E. Estimates are based on 100000 draws obtained from the MCMC algorithm following a burn-in of 10000 draws. Highest posterior density regions are constructed using the algorithm in Hyndman (1996). The main finding emerging from these figures is the negligible role played by noise shocks for U.S. macroeconomic fluctuations. With one exception, it can be seen that noise shocks explain very small fractions of the FEV of all series at all horizons and IRFs to noise shocks have credible intervals that include zero at all horizons. The one exception occurs for the stock price variable using  $\mathcal{R}_\infty$  which assumes very slow learning. For this variable, at short horizons there is some evidence that noise shocks have a small impact. But even this small impact vanishes after two years.

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 to emphasize in our results. However, the figures show other interesting patterns relating to the other shocks which we discuss in

$n$	SDDR		DIC	
	$\mathcal{R}_2$	$\mathcal{R}_\infty$	3 variables	6 variables
3	1.6	1.8	1600.81 (0.17)	-
6	21.0	12.0	1604.86 (0.25)	2907.89 (0.36)
10	20.6	15.9	1578.64 (0.30)	2907.18 (0.29)
15	16.6	14.9	1568.87 (0.21)	2860.72 (0.40)
20	49.5	34.7	1553.17 (0.33)	2820.94 (0.44)

Table 2: SDDRs and DICs for models of different dimensions. SDDRs represent Bayes factors in favor of the unrestricted specification for models of varying dimensions.  $\mathcal{R}_2$  denotes the representation with all non-zero, finite roots outside the unit circle (fastest learning);  $\mathcal{R}_\infty$  denotes the representation with all finite roots inside the unit circle (slowest learning). DICs compare unrestricted models only. Those labelled “3 variables” are computed using the marginal distribution of the three variables in the  $n = 3$  case as the likelihood; those labelled “6 variables” are computed using the marginal distribution of the six variables in the  $n = 6$  case as the likelihood. Numerical standard errors are in parentheses.

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 50% 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. using set identification the point estimates indicate they tend to explain around 20% of the FEV at long horizons).

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 for TFP. As for other series, non-news shocks tend to play a small role in 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 two point identification schemes ( $\mathcal{R}_2$  and  $\mathcal{R}_\infty$ ) and set identification tend to yield very similar results. There are a few exceptions to this such as the one noted above for stock prices and  $\mathcal{R}_\infty$ . For the fast learning  $\mathcal{R}_2$  representation, the FEVDs of

non-news shocks falls faster over time than for other representations. But overall, our findings are qualitatively similar for all schemes, indicating that the root flipping issue is not a major concern in this application.

Previously, we presented evidence in favor of the  $n = 20$  model and found that the strongest support for our Assumption 1 can be found in this case. Comparing IRFs and FEVDs for  $n = 20$  to  $n = 6$  we find them to be quite similar. A point worth noting in the  $n = 20$  results is that estimation precision is quite good. Credible intervals are not substantially wider than in more parsimonious lower dimensional models. This is due to our SSVS prior which is effectively shrinking unimportant coefficients in the non-parsimonious VARMA to zero.

## 7 Conclusions

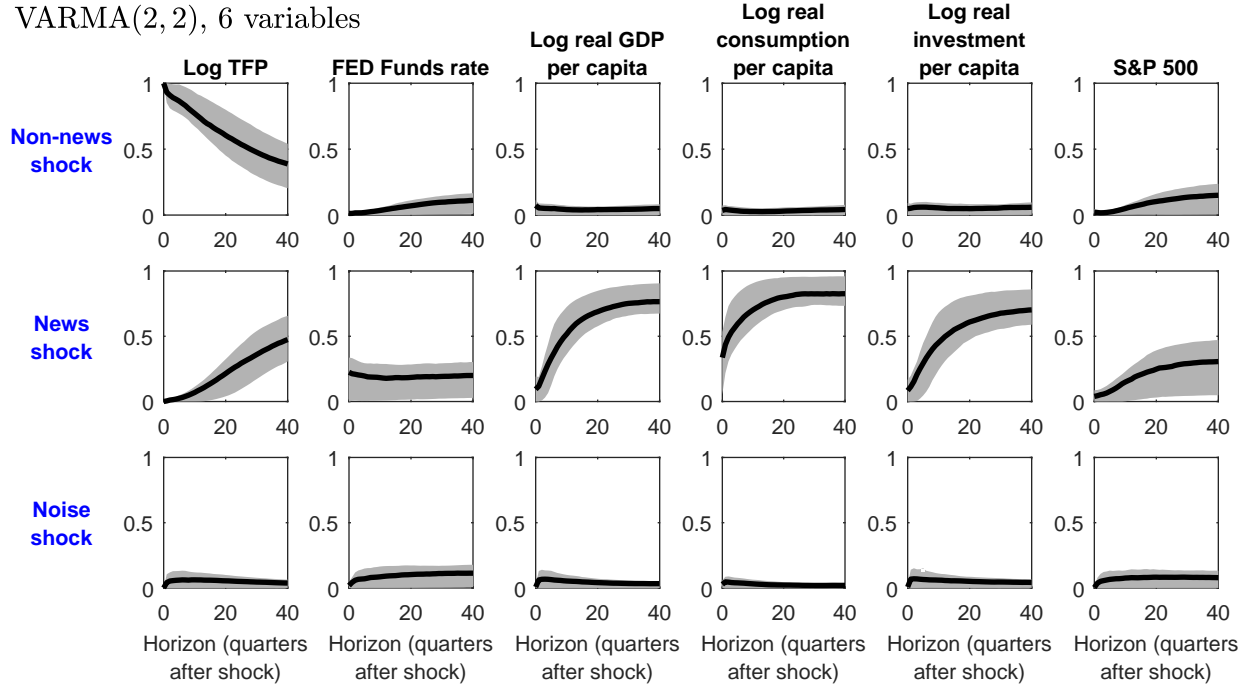
In this paper, we have established conditions under which impulse responses to non-news, news and noise shocks are identified, with particular focus on restrictions in the information structures commonly assumed in the theoretical literature. We further provided a full econometric methodology for estimating IRFs and FEVDs assuming such restrictions do not hold, and along with this, we developed a method for calculating the Bayes factor comparing models which embed these two sets of assumptions. Thus, we obtain a way of testing which set of assumptions is supported by the data.

Our approach involves the use of an SVARMA, which overcomes a number of deficiencies that make SVARs impractical in this setting. Especially when working with a large number of dependent variables, VARMA models can be over-parameterized and we have developed a Bayesian econometric method involving a hierarchical SSVS shrinkage prior which allows for parsimonious and efficient estimation in high dimensions. Hence, we established that SVARMA models of the large dimensions increasingly used in macroeconomics can be easily and robustly estimated.

The methodology introduced in this paper is potentially useful in a wide variety of contexts. For example, it can be directly applied to problems with noisy news about fundamentals other than productivity, such as dividends, exchange rates, fiscal policy or defence spending. With straightforward modifications, it can also be used to investigate the effects of news and noise shocks at alternative anticipation horizons. Likewise, our methodology can be extended in a number of interesting directions, such as specifications involving multiple fundamental processes.

In an empirical application involving up to 20 variables, we have presented evidence suggesting that our econometric methods can produce accurate estimates of news and noise shocks. Our key empirical finding is that noise shocks play a minor role in macroeconomic fluctuations, explaining negligible fractions of the forecast error variance of the main macroeconomic variables. In addition, we find strong evidence in favour of the assumption that agents learn about non-news, news and noise shocks at some point in the future, particularly as more variables are added in the system. To the extent that existing DSGEs ubiquitously assume agents cannot fully recover these three shocks from what they observe, our results suggest that further development of DSGEs in this direction is warranted.

VARMA(2, 2), 6 variables



VARMA(2, 2), 20 variables

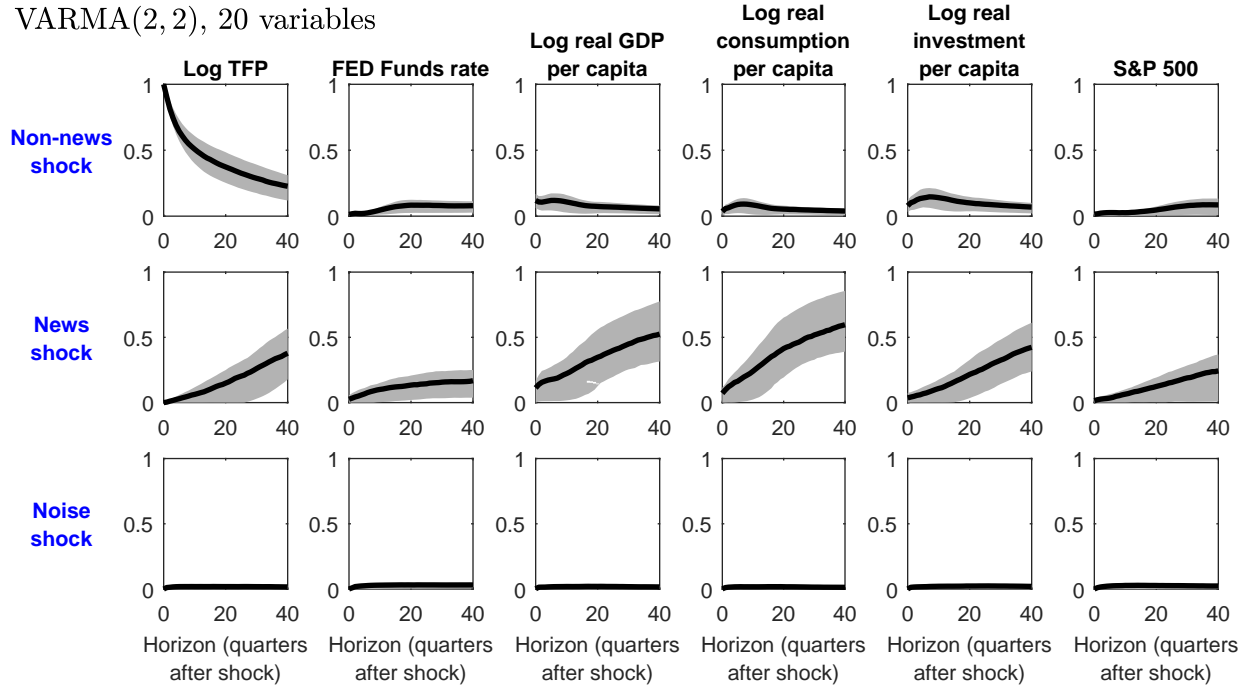
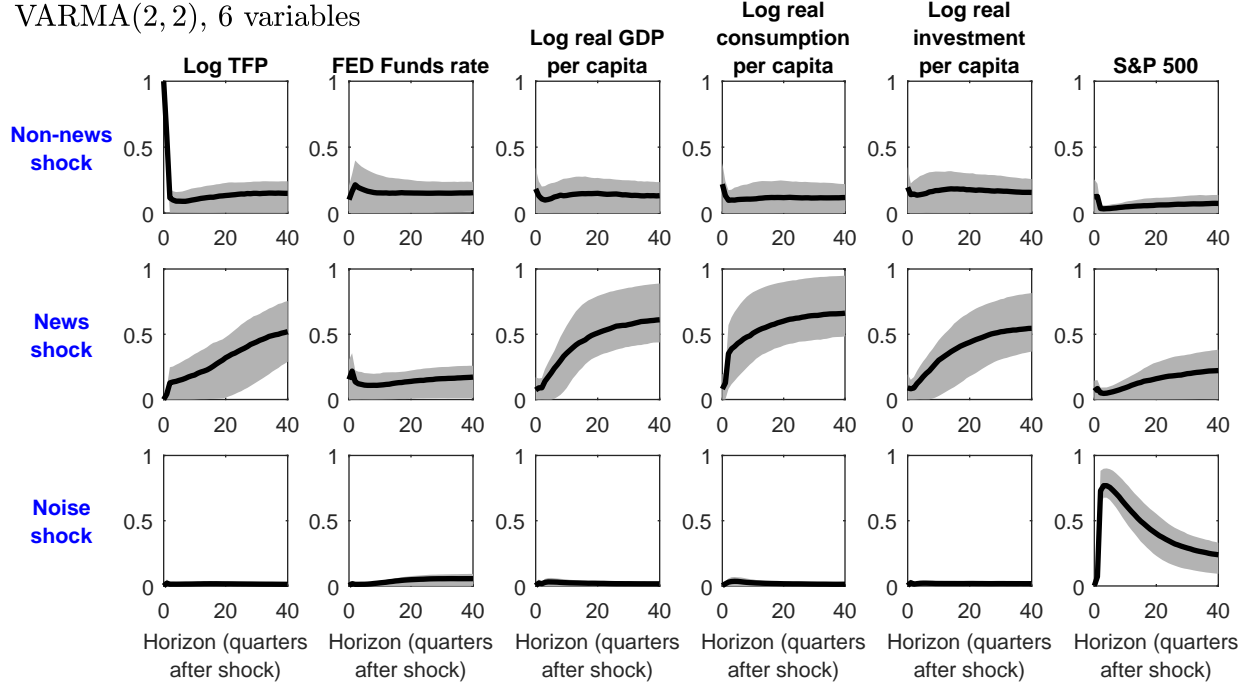


Figure 1: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_2$  (all non-zero, finite roots outside the unit circle). Top panel:  $n = 6$ , bottom panel:  $n = 20$ .

VARMA(2, 2), 6 variables



VARMA(2, 2), 20 variables

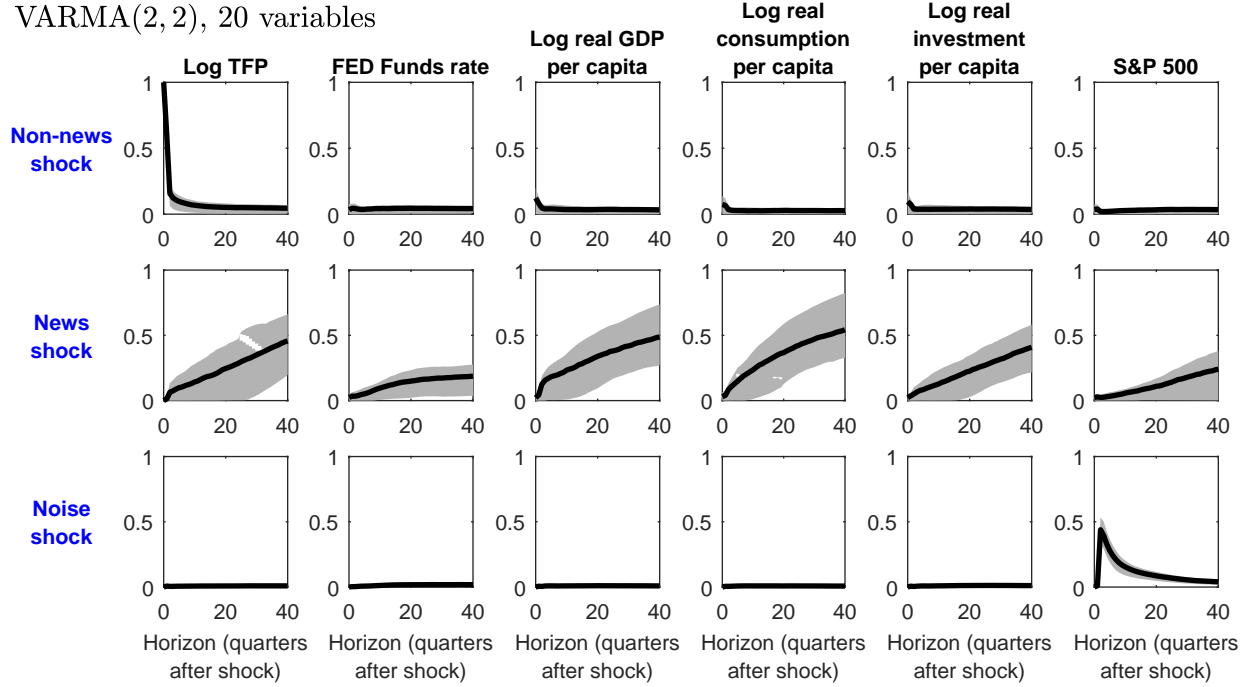
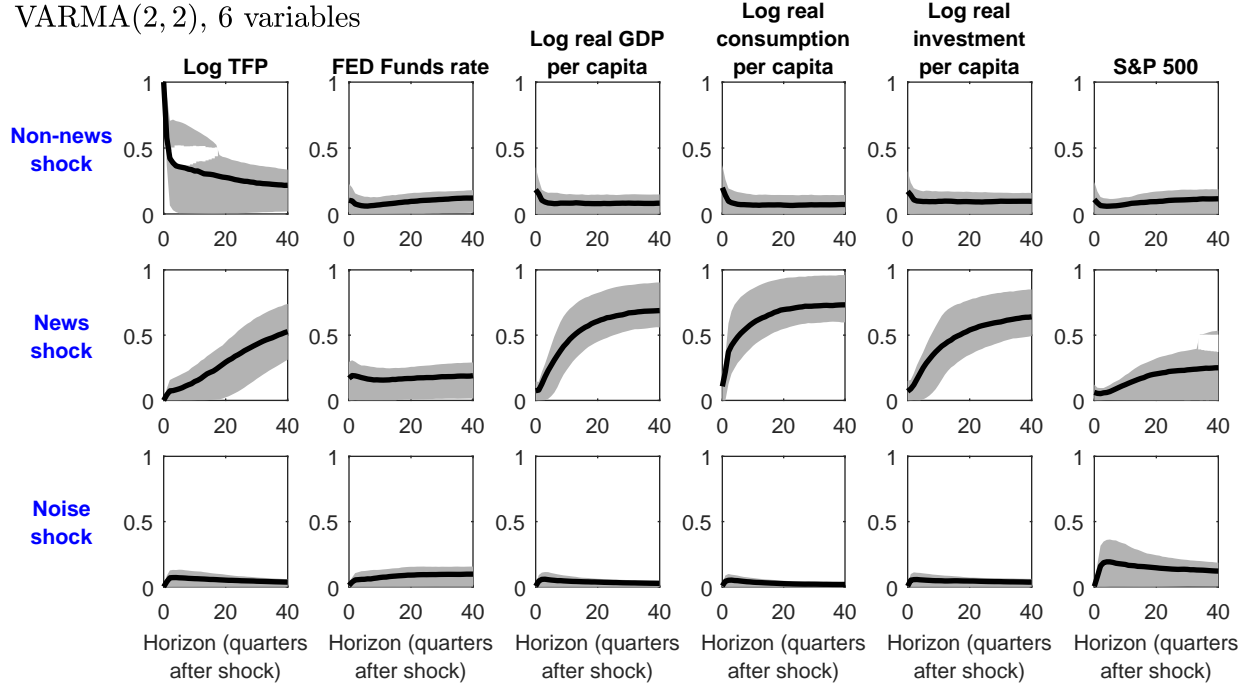


Figure 2: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_\infty$  (all finite roots inside the unit circle). Top panel:  $n = 6$ , bottom panel:  $n = 20$ .

VARMA(2, 2), 6 variables



VARMA(2, 2), 20 variables

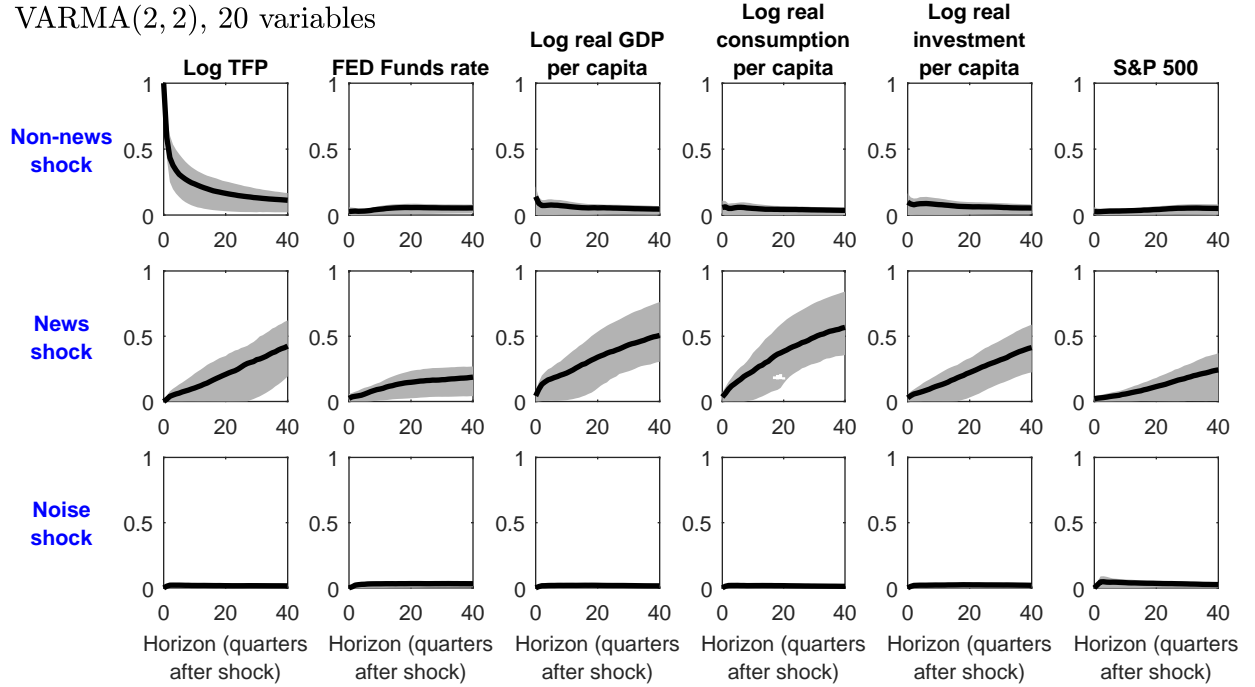
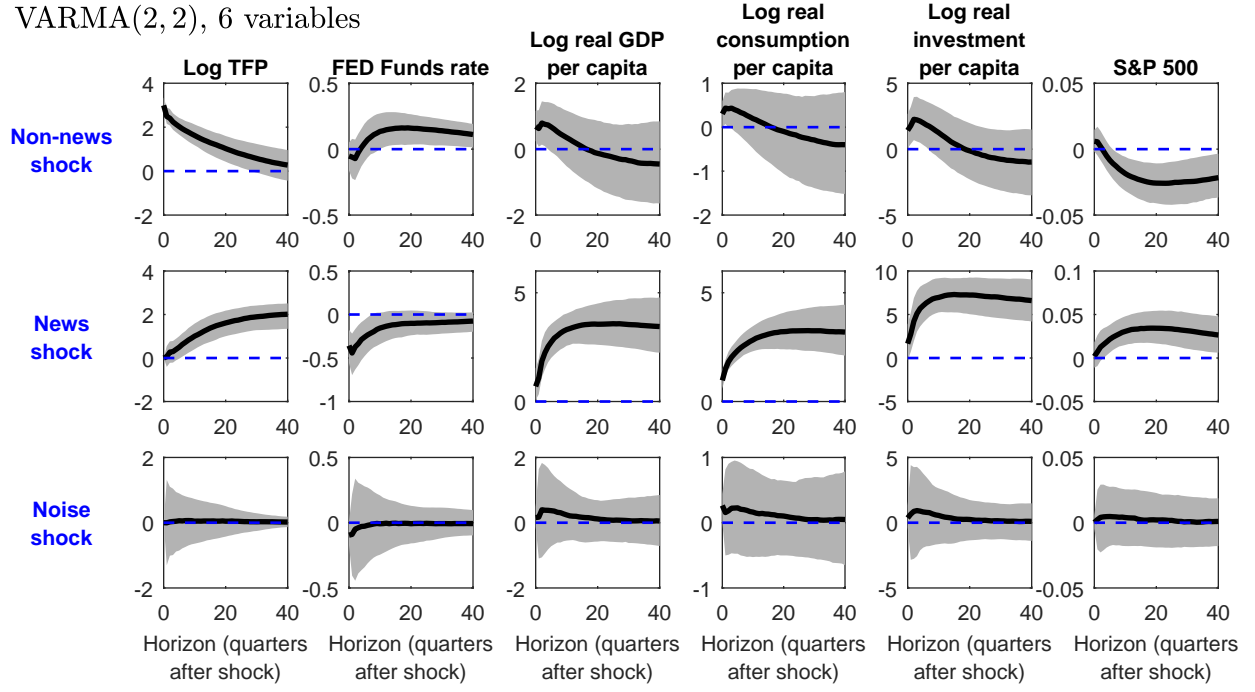


Figure 3: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on set identification over all possible representations. Top panel:  $n = 6$ , bottom panel:  $n = 20$ .

VARMA(2, 2), 6 variables



VARMA(2, 2), 20 variables

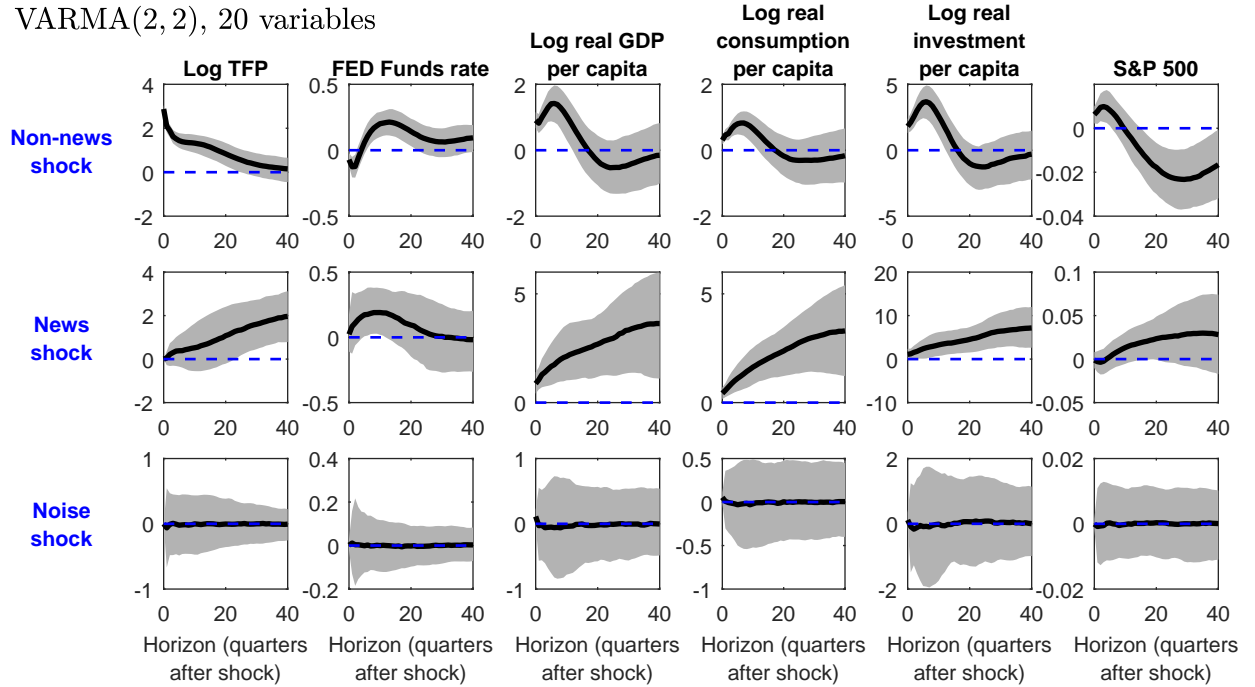
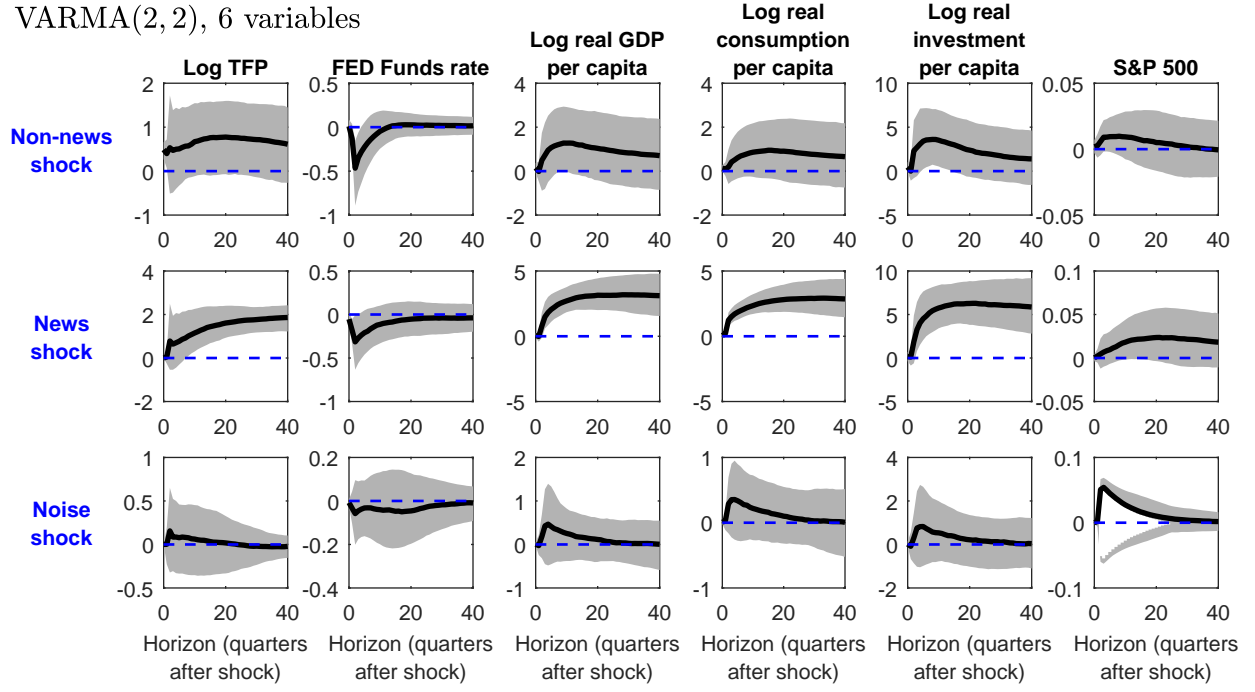


Figure 4: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_2$  (all non-zero, finite roots outside the unit circle). Top panel:  $n = 6$ , bottom panel:  $n = 20$ .

VARMA(2, 2), 6 variables



VARMA(2, 2), 20 variables

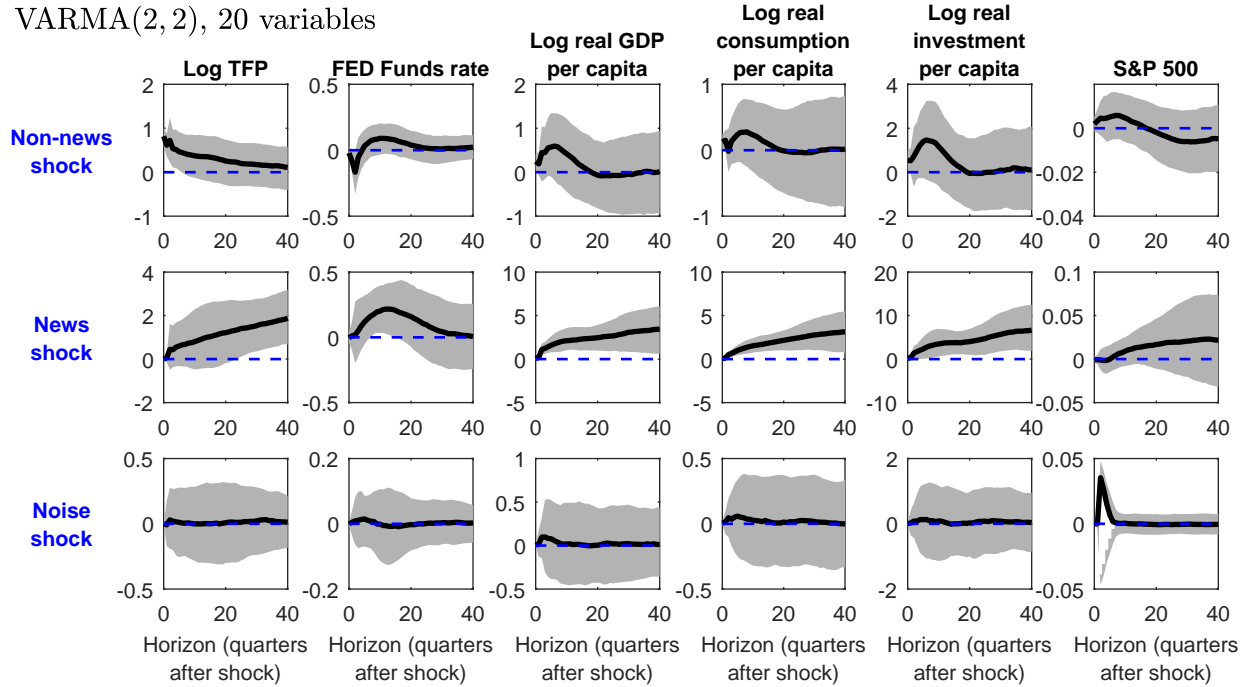
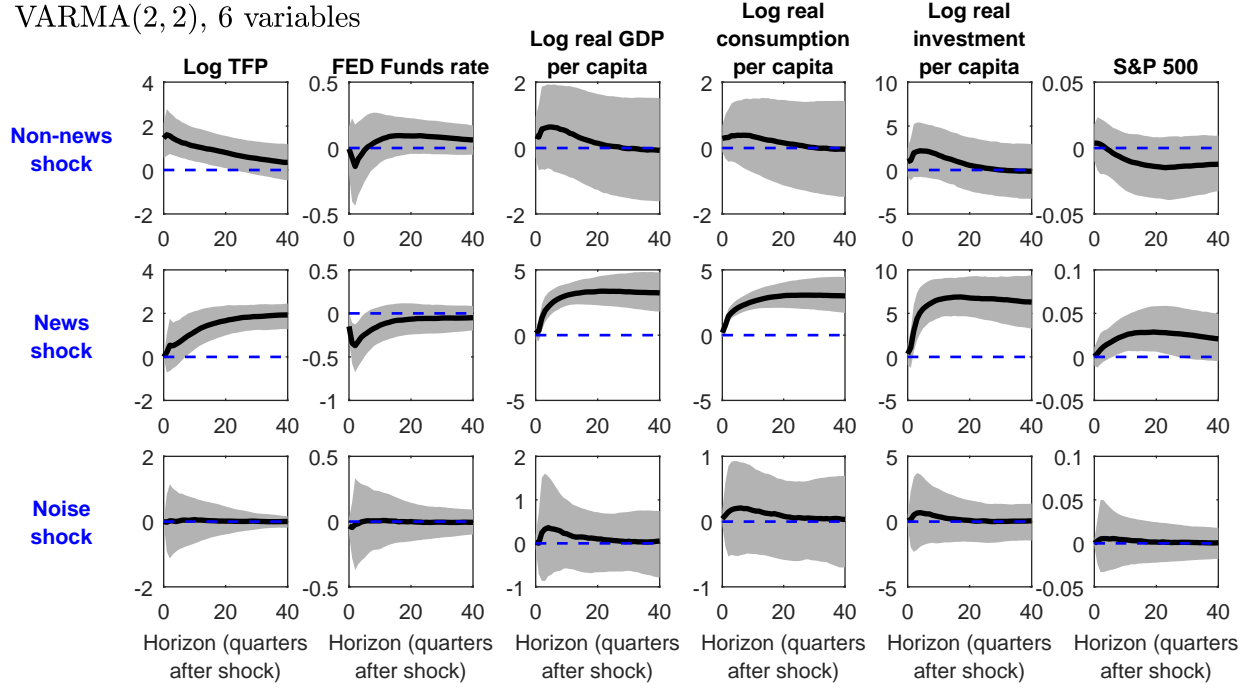


Figure 5: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_\infty$  (all finite roots inside the unit circle). Top panel:  $n = 6$ , bottom panel:  $n = 20$ .

VARMA(2, 2), 6 variables



VARMA(2, 2), 20 variables

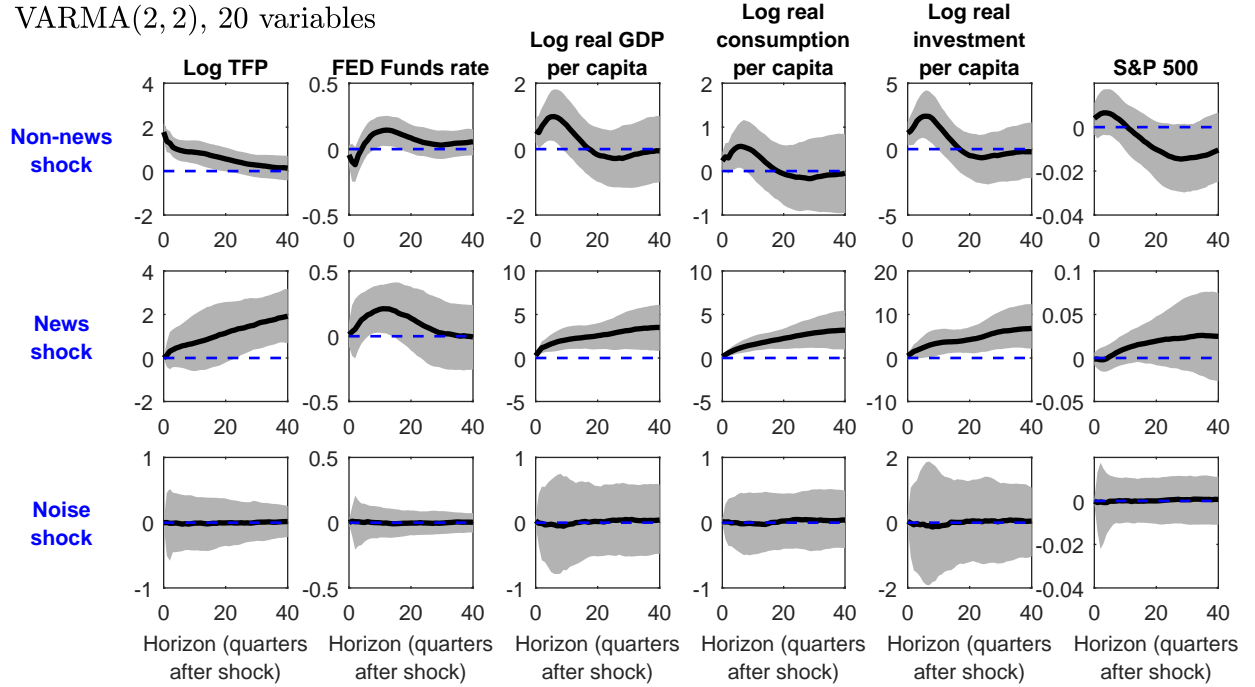


Figure 6: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on set identification over all possible representations. Top panel:  $n = 6$ , bottom panel:  $n = 20$ .

## References

- ANDERSON, B. D. O., AND M. DEISTLER (2008): “Properties of Zero-Free Transfer Function Matrices,” *SICE Journal of Control, Measurement, and System Integration*, 1(4), 284–292.
- BAGGIO, G., AND A. FERRANTE (2016): “On the Factorization of Rational Discrete-Time Spectral Densities,” *IEEE Transactions on Automatic Control*, 61, 969–981.
- BARSKY, R., A. JUSTINIANO, AND L. MELOSI (2014): “The Natural Rate of Interest and Its Usefulness for Monetary Policy,” *American Economic Review*, 104(5), 37–43.
- BARSKY, R., AND E. SIMS (2011): “News Shocks and Business Cycles,” *Journal of Monetary Economics*, 58(3), 273–289.
- (2012): “Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence,” *American Economic Review*, 102(4), 1343–1377.
- BENATI, L., J. C. C. CHAN, E. EISENSTAT, AND G. KOOP (2020): “Identifying Noise Shocks,” *Journal of Economic Dynamics and Control*, forthcoming.
- BLANCHARD, O., J.-P. L’HUILIER, AND G. LORENZONI (2013): “News, Noise, and Fluctuations: An Empirical Exploration,” *American Economic Review*, 103(7), 3045–70.
- CELEUX, G., F. FORBES, C. ROBERT, AND D. TITTERINGTON (2006): “Deviance Information Criteria for Missing Data Models,” *Bayesian Analysis*, 1(4), 651–674.
- CHAHROUR, R., AND K. JURADO (2018a): “News or Noise? The Missing Link,” *American Economic Review*, 108, 1702–1736.
- (2018b): “Recoverability,” Working Paper.
- CHAN, J., AND E. EISENSTAT (2017): “Efficient Estimation of Bayesian VARMA with Time-Varying Coefficients,” *Journal of Applied Econometrics*, 32(7), 1277–1297.
- CHAN, J., E. EISENSTAT, AND G. KOOP (2016): “Large Bayesian VARMA,” *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.
- CHAN, K.-S., AND L.-H. HO (2004): “On the Unique Representation of Non-Gaussian Multivariate Linear Processes,” Technical Report 341, Department of Statistics and Actuarial Science, University of Iowa.
- CHARI, V., P. KEHOE, AND E. MCGRATTAN (2008): “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.
- FERNANDEZ-VILLAYERDE, J., J. RUBIO-RAMIREZ, T. SARGENT, AND M. WATSON (2006): “A’s, B’s, C’s (and D’s) for Understanding VARs,” mimeo.
- FORNI, M., AND L. GAMBETTI (2014): “Testing for Sufficient Information in Structural VARs,” *Journal of Monetary Economics*, 66(September), 124–136.
- FORNI, M., L. GAMBETTI, M. LIPPI, AND L. SALA (2017): “Noisy News in Business Cycles,” *American Economic Journal: Macroeconomics*, 9(4), 122–152.
- GEORGE, E., D. SUN, AND S. NI (2008): “Bayesian Stochastic Search for VAR Model Restrictions,” *Journal of Econometrics*, 142, 553–580.
- 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.
- HANSEN, L. P., W. T. ROBERDS, AND T. J. SARGENT (1991): “Time Series Implications of Present Value Budget Balance and of Martingale Models of Consumption and Taxes,” in *Rational Expectations Econometrics*, ed. by L. P. Hansen, and T. J. Sargent. Westview Press, Boulder.
- HYNDMAN, R. J. (1996): “Computing and Graphing Highest Density Regions,” *The American Statistician*, 50(2), 120–126.
- 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), 311–334.
- 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.
- KOOP, G., AND S. M. POTTER (1999): “Bayes Factors and Nonlinearity: Evidence from Economic Time Series,” *Journal of Econometrics*, 88(2), 251–281.

- 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, 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.
- LUETKEPOHL, 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.
- MOL, C. D., D. GIANNONE, AND L. REICHLIN (2008): “Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?,” *Journal of Econometrics*, 146(2), 318 – 328.
- PLAGBORG-MØLLER, M. (2016): “Bayesian Inference on Structural Impulse Response Functions,” available at: [http://scholar.harvard.edu/files/plagborg/files/irf\\_bayes.pdf](http://scholar.harvard.edu/files/plagborg/files/irf_bayes.pdf).
- POSKITT, D., AND W. YAO (2017): “Vector Autoregressions and Macroeconomic Modeling: An Error Taxonomy,” *Journal of Business & Economic Statistics*, 35(3), 407–419.
- 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.
- VERDINELLI, I., AND L. WASSERMAN (1995): “Computing Bayes Factors using a Generalization of the Savage-Dickey density ratio,” *Journal of the American Statistical Association*, 90(430), 614–618.

## A Proofs of Propositions

To prove Proposition 1, we first state a lemma characterizing uniqueness when observationally equivalent representations are restricted to be equivalent up to constant orthogonal rotations.

**Lemma 1.** *Let  $\mathbf{K}(L)\epsilon_{\underline{l}}$ ,  $\tilde{\mathbf{K}}(L)\tilde{\epsilon}_t$  and  $\mathbf{\Gamma}$  be defined as in Proposition 1. Assume  $\tilde{\mathbf{K}}(L) = \mathbf{K}(L)\mathbf{\Gamma}$ . If  $\mathbf{K}(L)$  and  $\tilde{\mathbf{K}}(L)$  both satisfy Assumptions 1-5, then  $\mathbf{\Gamma}_{12} = \mathbf{\Gamma}_{21} = 0$  and  $\mathbf{\Gamma}_{11}$  is diagonal.*

*Proof.* By Assumptions 3 and 5,  $\tilde{\mathbf{K}}(L)$  satisfies

$$\begin{aligned}\tilde{\mathbf{K}}_{0,12} &= \mathbf{K}_{0,11}\mathbf{\Gamma}_{12} + \mathbf{K}_{0,12}\mathbf{\Gamma}_{22} = 0, \\ \tilde{\mathbf{K}}_{1j}(L) &= \mathbf{K}_{11}(L)\mathbf{\Gamma}_{1j} + \mathbf{K}_{12}(L)\mathbf{\Gamma}_{2j} = 0, \quad j \geq 3.\end{aligned}$$

Since  $K_{0,12} = 0$ , but  $K_{0,11} \neq 0$  and  $K_{l,12} \neq 0$  for some  $l \geq 1$  (as implied by the rank conditions in Assumption 5), we immediately obtain  $\mathbf{\Gamma}_{12} = \dots = \mathbf{\Gamma}_{1n} = 0$  and  $\mathbf{\Gamma}_{23} = \dots = \mathbf{\Gamma}_{2n} = 0$ . Orthogonality of  $\mathbf{\Gamma}$  then implies  $\mathbf{\Gamma}_{21} = \dots = \mathbf{\Gamma}_{n1} = 0$  and  $\mathbf{\Gamma}_{32} = \dots = \mathbf{\Gamma}_{n2} = 0$  as well, with  $|\mathbf{\Gamma}_{11}| = 1$  and  $|\mathbf{\Gamma}_{22}| = 1$ .

By Assumption 4, we obtain for some constant  $\tilde{c} > 0$

$$\tilde{K}_{0,i3} - \tilde{c}\tilde{K}_{0,i2} = \sum_{j=1}^n K_{0,ij}(\mathbf{\Gamma}_{j3} - \tilde{c}\mathbf{\Gamma}_{j2}) = 0, \quad i = 2, \dots, n. \quad (19)$$

Taking into account the restrictions on  $\mathbf{\Gamma}$  established above and the fact that

$$K_{0,i2} = \frac{1}{c}K_{0,i3},$$

the restrictions in (19) can be represented by

$$\begin{pmatrix} K_{0,23} & \dots & K_{0,2n} \\ \vdots & \ddots & \vdots \\ K_{0,n3} & \dots & K_{0,nn} \end{pmatrix} \begin{pmatrix} \mathbf{\Gamma}_{33} \pm \tilde{c}/c \\ \mathbf{\Gamma}_{43} \\ \vdots \\ \mathbf{\Gamma}_{n3} \end{pmatrix} = 0.$$

Since the first matrix in this expression is  $n-1 \times n-2$ , and Assumption 4 implies it is of rank  $n-2$ , we obtain  $\mathbf{\Gamma}_{i3} = 0$  for all  $i \geq 4$ ,  $|\mathbf{\Gamma}_{33}| = \tilde{c}/c$ , and by orthogonality of  $\mathbf{\Gamma}$ ,  $\mathbf{\Gamma}_{3j} = 0$  for all  $j \geq 4$ ,  $|\mathbf{\Gamma}_{33}| = 1$ .  $\square$

*Proof of Proposition 1.* Following the derivations in Section 5, define

$$\mathbf{C}_1 = \begin{pmatrix} 0 & 0 & 1 \\ \frac{c}{\sqrt{1+c^2}} & \frac{1}{\sqrt{1+c^2}} & 0 \\ -\frac{1}{\sqrt{1+c^2}} & \frac{c}{\sqrt{1+c^2}} & 0 \\ & & \mathbf{I}_{n-3} \end{pmatrix}, \quad \mathbf{C}_2 = \begin{pmatrix} 0 & 0 & 1 \\ \frac{\tilde{c}}{\sqrt{1+\tilde{c}^2}} & \frac{1}{\sqrt{1+\tilde{c}^2}} & 0 \\ -\frac{1}{\sqrt{1+\tilde{c}^2}} & \frac{\tilde{c}}{\sqrt{1+\tilde{c}^2}} & 0 \\ & & \mathbf{I}_{n-3} \end{pmatrix},$$

and let

$$\Phi(L) = \mathbf{K}(L)\mathbf{C}_1 \begin{pmatrix} L^{-\tau} & \\ & \mathbf{I}_{n-1} \end{pmatrix}, \quad \tilde{\Phi}(L) = \tilde{\mathbf{K}}(L)\mathbf{C}_2 \begin{pmatrix} L^{-\tau} & \\ & \mathbf{I}_{n-1} \end{pmatrix},$$

where  $L^{-1}$  denotes the forward operator and  $L^{-\tau} = (L^{-1})^\tau$ .

Observe that if  $\omega = 0$  in Assumption 2, then  $\Phi(L)$  and  $\tilde{\Phi}(L)$  are *fundamental* representations; otherwise they are *basic nonfundamental* representations (Lippi and Reichlin, 1994, Definition 3) corresponding to the fundamental one associated with  $\omega = 0$ . Accordingly, Lippi and Reichlin (1994, Theorem 3) yields  $\tilde{\Phi}(L) = \Phi(L)\mathbf{C}_0$ , where  $\mathbf{C}_0$  is a constant orthogonal matrix.

Moreover, if  $\mathbf{K}(L)$  satisfies Assumptions 3-5, then  $\Phi(L)$  satisfies

$$\Phi_{14}(L) = \dots = \Phi_{1n}(L) = 0, \quad (20)$$

$$\Phi_{11}(L)L^\tau - c\Phi_{12}(L) = 0, \quad \Phi_{11}(0) \neq 0, \quad (21)$$

$$\hat{\Phi}_{11}(\zeta) = \hat{\Phi}_{12}(\zeta) = 0, \quad \hat{\Phi}_{13}(\zeta) \neq 0, \quad \text{for some } \zeta \neq 0, \quad (22)$$

where  $\hat{\Phi}_{1i} = \chi(L)\Phi_{1i}$  for  $i = 1, 2, 3$ . The same holds for  $\tilde{\Phi}(L)$  if  $\tilde{\mathbf{K}}(L)$  satisfies Assumptions 3-5 as well, with  $\tilde{\chi}(L)$  being the least common multiple of the the denominators of  $\tilde{K}_{11}(L)$  and  $\tilde{K}_{12}(L)$ .

The restrictions in (20) imply

$$\frac{c}{\sqrt{1+c^2}}\hat{K}_{12}(L)L^{-\tau}C_{0,1i} + \frac{1}{\sqrt{1+c^2}}\hat{K}_{12}(L)C_{0,2i} + \hat{K}_{11}(L)C_{0,3i} = 0, \quad i = 4, \dots, n.$$

But by Assumption 5,  $\hat{K}_{11}(\zeta)C_{0,3i} = 0$ , for some  $\zeta \neq 0$  and  $\hat{K}_{11}(\zeta) \neq 0$ . Hence,  $C_{0,3i} = 0$  for all  $i = 4, \dots, n$ . At the same time  $\hat{K}_{12}(0) = 0$  by Assumption 3, which together with  $C_{0,3i} = 0$  yields  $\hat{K}_{\tau,12}C_{0,1i} = 0$ , and consequently,  $C_{0,1i} = 0$ . Finally, it follows from the remaining term,  $\frac{1}{\sqrt{1+c^2}}\hat{K}_{12}(L)C_{0,2i} = 0$ , that  $C_{0,2i} = 0$ .

Hence, it holds that  $C_{0,1i} = C_{0,2i} = C_{0,3i} = 0$  for all  $i \geq 4$ , and by orthogonality of  $\mathbf{C}_0$ , this implies  $C_{0,i1} = C_{0,i2} = C_{0,i3} = 0$  for all  $i \geq 4$  as well. Another important implication is that the functions  $\hat{K}_{12}(L)L^{-\tau}$ ,  $\hat{K}_{12}(L)$  and  $\hat{K}_{11}(L)$  are *linearly independent*.

The restrictions in (21) further imply

$$\begin{aligned} & \frac{1}{\sqrt{1+c^2}}\hat{K}_{12}(L) \left( C_{0,11} - \frac{\tilde{c}}{c}C_{0,22} \right) + \frac{1}{\sqrt{1+c^2}}\hat{K}_{12}(L)L^\tau C_{0,21} \\ & + \hat{K}_{11}(L)L^\tau C_{0,31} - \frac{\tilde{c}c}{\sqrt{1+c^2}}\hat{K}_{12}(L)L^{-\tau}C_{0,21} - \tilde{c}\hat{K}_{11}(L)C_{0,32} = 0. \end{aligned} \quad (23)$$

If the three functions  $\hat{K}_{12}(L)L^{-\tau}$ ,  $\hat{K}_{12}(L)$  and  $\hat{K}_{11}(L)$  are linearly independent, then so are the four functions  $\hat{K}_{12}(L)L^{-\tau}$ ,  $\hat{K}_{12}(L)$ ,  $\hat{K}_{12}(L)L^\tau$ , and  $\hat{K}_{11}(L)L^\tau$ . Hence, if  $C_{0,32} = 0$  in (23), it follows immediately that  $C_{0,21} = C_{0,31} = C_{0,12} = 0$  and  $|C_{0,11}| = |C_{0,22}| = 1$ .

Likewise, if all five functions  $\hat{K}_{11}(L)$ ,  $\hat{K}_{12}(L)L^{-\tau}$ ,  $\hat{K}_{12}(L)$ ,  $\hat{K}_{12}(L)L^\tau$ , and  $\hat{K}_{11}(L)L^\tau$  in (23) are linearly independent, then we again trivially obtain  $C_{0,21} = C_{0,31} = C_{0,12} =$

$C_{0,32} = 0$  and  $|C_{0,11}| = |C_{0,22}| = 1$ . Otherwise, there exist constants  $\xi_1, \xi_2, \xi_3$  and  $\xi_4$  such that

$$\widehat{K}_{11}(L) = \xi_1 \widehat{K}_{12}(L) + \xi_2 \widehat{K}_{12}(L)L^\tau + \xi_3 \widehat{K}_{11}(L)L^\tau + \xi_4 \widehat{K}_{12}(L)L^{-\tau}.$$

We aim to show that in this case,  $C_{0,32} \neq 0$  can only hold for coefficients of  $\mathbf{K}(L)$  in a set with Lebesgue measure zero, so that  $C_{0,21} = C_{0,31} = C_{0,12} = C_{0,32} = 0$  and  $|C_{0,11}| = |C_{0,22}| = 1$  holds almost everywhere.

To this end, observe that  $\xi_3 = 1/\zeta \neq 0$  and  $\xi_4 = K_{0,11}/K_{\tau,12} \neq 0$ . Substituting into (23) yields

$$\begin{aligned} & \frac{1}{\sqrt{1+c^2}} \widehat{K}_{12}(L) \left( C_{0,11} - \frac{\tilde{c}}{c} C_{0,22} - \sqrt{1+c^2} \xi_1 \tilde{c} C_{0,32} \right) \\ & + \frac{1}{\sqrt{1+c^2}} \widehat{K}_{12}(L)L^\tau \left( C_{0,21} - \sqrt{1+c^2} \xi_2 \tilde{c} C_{0,32} \right) + \widehat{K}_{11}(L)L^\tau (C_{0,31} - \xi_3 \tilde{c} C_{0,32}) \\ & - \frac{\tilde{c}c}{\sqrt{1+c^2}} \widehat{K}_{12}(L)L^{-\tau} \left( C_{0,21} + \frac{\sqrt{1+c^2}}{c} \xi_4 C_{0,32} \right) = 0. \end{aligned} \quad (24)$$

Since,  $\widehat{K}_{12}(L)L^{-\tau}$ ,  $\widehat{K}_{12}(L)$ ,  $\widehat{K}_{12}(L)L^\tau$ , and  $\widehat{K}_{11}(L)L^\tau$  are linearly independent, however, this results in:

$$\begin{aligned} C_{0,11} &= \left( \frac{C_{0,22}}{C_{0,32}} + \sqrt{1+c^2} \xi_1 \right) \tilde{c} C_{0,32}, \\ C_{0,21} &= \sqrt{1+c^2} \xi_2 \tilde{c} C_{0,32}, \\ C_{0,31} &= \xi_3 \tilde{c} C_{0,32}, \\ C_{0,12} &= -\frac{\sqrt{1+c^2}}{c} \xi_4 C_{0,32}, \end{aligned}$$

whereas orthogonality of  $\mathbf{C}_0$  further requires:

$$\begin{aligned} 1 &= C_{0,11}^2 + C_{0,21}^2 + C_{0,31}^2, \\ 1 &= C_{0,12}^2 + C_{0,22}^2 + C_{0,32}^2, \\ 0 &= C_{0,11}C_{0,12} + C_{0,21}C_{0,22} + C_{0,31}C_{0,32}. \end{aligned}$$

This system of seven equations in seven unknowns ( $C_{0,11}$ ,  $C_{0,21}$ ,  $C_{0,31}$ ,  $C_{0,12}$ ,  $C_{0,22}$ ,  $C_{0,32}$  and  $\tilde{c}$ ) admits a unique solution, up to sign normalization, of the form:

$$\begin{aligned} C_{0,11} &= \frac{\left( \frac{\sqrt{1+c^2}}{c} \right) \xi_1 + \xi_5}{\sqrt{\left( \left( \frac{\sqrt{1+c^2}}{c} \right) \xi_1 + \xi_5 \right)^2 + (1+c^2)\xi_2^2 + \xi_3^2}}, & C_{0,12} &= \frac{-\left( \frac{\sqrt{1+c^2}}{c} \right) \xi_4}{\sqrt{1 + \left( \frac{1+c^2}{c^2} \right) \xi_4^2 + \xi_5^2}}, \\ C_{0,21} &= \frac{\sqrt{1+c^2} \xi_2}{\sqrt{\left( \left( \frac{\sqrt{1+c^2}}{c} \right) \xi_1 + \xi_5 \right)^2 + (1+c^2)\xi_2^2 + \xi_3^2}}, & C_{0,22} &= \frac{\xi_5}{\sqrt{1 + \left( \frac{1+c^2}{c^2} \right) \xi_4^2 + \xi_5^2}}, \\ C_{0,31} &= \frac{\xi_3}{\sqrt{\left( \left( \frac{\sqrt{1+c^2}}{c} \right) \xi_1 + \xi_5 \right)^2 + (1+c^2)\xi_2^2 + \xi_3^2}}, & C_{0,32} &= \frac{1}{\sqrt{1 + \left( \frac{1+c^2}{c^2} \right) \xi_4^2 + \xi_5^2}}, \end{aligned}$$

where

$$\xi_5 = \frac{\xi_3 - \left(\frac{1+c^2}{c^2}\right) \xi_1 \xi_4}{\left(\frac{\sqrt{1+c^2}}{c}\right) \xi_4 - \sqrt{1+c^2} \xi^2}.$$

In addition, there exist unique (up to sign normalization) constants  $C_{0,13}$ ,  $C_{0,23}$  and  $C_{0,33}$  satisfying

$$\begin{aligned} 0 &= C_{0,13}C_{0,11} + C_{0,23}C_{0,21} + C_{0,33}C_{0,31}, \\ 0 &= C_{0,13}C_{0,12} + C_{0,23}C_{0,22} + C_{0,33}C_{0,32}, \\ 1 &= C_{0,13}^2 + C_{0,23}^2 + C_{0,33}^2. \end{aligned}$$

Clearly,  $C_{0,13} \neq 0$ ,  $C_{0,23} \neq 0$  and  $C_{0,33} \neq 0$  almost everywhere.

The restrictions in (22), however, require that there exists a constant  $\tilde{\zeta} \neq 0$  such that  $\widehat{\Phi}_{11}(\tilde{\zeta}) = \widehat{\Phi}_{12}(\tilde{\zeta}) = 0$  but  $\widehat{\Phi}_{13}(\tilde{\zeta}) \neq 0$ , where  $\widehat{\Phi}_{1i}(L) = \tilde{\chi}(L)\widehat{\Phi}_{1i}(L)$  for  $i = 1, 2, 3$ . This, in turn, implies

$$\begin{aligned} 0 &= \frac{c}{\sqrt{1+c^2}} \check{K}_{12}(\tilde{\zeta}) \tilde{\zeta}^{-\tau} C_{0,11} + \frac{1}{\sqrt{1+c^2}} \check{K}_{12}(\tilde{\zeta}) C_{0,21} + \check{K}_{11}(\tilde{\zeta}) C_{0,31}, \\ 0 &= \frac{c}{\sqrt{1+c^2}} \check{K}_{12}(\tilde{\zeta}) \tilde{\zeta}^{-\tau} C_{0,12} + \frac{1}{\sqrt{1+c^2}} \check{K}_{12}(\tilde{\zeta}) C_{0,22} + \check{K}_{11}(\tilde{\zeta}) C_{0,32}, \\ 0 &\neq \frac{c}{\sqrt{1+c^2}} \check{K}_{12}(\tilde{\zeta}) \tilde{\zeta}^{-\tau} C_{0,13} + \frac{1}{\sqrt{1+c^2}} \check{K}_{12}(\tilde{\zeta}) C_{0,23} + \check{K}_{11}(\tilde{\zeta}) C_{0,33}, \end{aligned}$$

where  $\check{K}_{1i}(L) = \tilde{\chi}(L)K_{1i}(L) = \tilde{\chi}(L)/\chi(L)\widehat{K}_{1i}(L)$  for  $i = 1, 2$ . Since  $\widehat{\Phi}_{11}(L)$ ,  $\widehat{\Phi}_{12}(L)$  and  $\widehat{\Phi}_{13}(L)$  are polynomials,  $\tilde{\chi}(L) = \tilde{\chi}(L)/\chi(L)$  is a polynomial with  $\tilde{\chi}(z) < \infty$  for all  $z \in \mathbb{C}$ ; since  $\check{K}_{12}(\tilde{\zeta})$  and  $\check{K}_{11}(\tilde{\zeta})$  cannot vanish simultaneously,  $\tilde{\chi}(\tilde{\zeta}) \neq 0$  either. It follows that

$$\begin{aligned} \frac{c}{\sqrt{1+c^2}} \tilde{\chi}(\tilde{\zeta}) \widehat{K}_{12}(\tilde{\zeta}) \tilde{\zeta}^{-\tau} &= C_{0,13} \bar{\xi}, \\ \frac{1}{\sqrt{1+c^2}} \tilde{\chi}(\tilde{\zeta}) \widehat{K}_{12}(\tilde{\zeta}) &= C_{0,23} \bar{\xi}, \\ \tilde{\chi}(\tilde{\zeta}) \widehat{K}_{11}(\tilde{\zeta}) &= C_{0,33} \bar{\xi}, \end{aligned}$$

for some  $\bar{\xi} \neq 0$ . But these equations can only be satisfied for

$$\widehat{K}_{12} \left( \left( \frac{cC_{0,23}}{C_{0,13}} \right)^{\frac{1}{\tau}} \right) = \frac{\sqrt{1+c^2}C_{0,23}}{C_{0,33}} \widehat{K}_{11} \left( \left( \frac{cC_{0,23}}{C_{0,13}} \right)^{\frac{1}{\tau}} \right).$$

As this condition does not hold for almost all  $\widehat{K}_{11}(L)$ ,  $\widehat{K}_{12}(L)$  and  $c$ , we conclude that  $C_{0,32} \neq 0$  almost everywhere, and in this case,  $C_{0,21} = C_{0,31} = C_{0,12} = 0$  as well. Orthogonality of  $\mathbf{C}_0$  then yields  $C_{0,13} = 0$ , and together with the previously established restrictions, we obtain

$$\begin{aligned} C_{0,12} &= \cdots = C_{0,1n} = 0, \\ C_{0,21} &= \cdots = C_{0,n1} = 0, \end{aligned}$$

such that  $\mathbf{C}_0$  commutes with  $\begin{pmatrix} L^{-\tau} & \\ & \mathbf{I}_{n-1} \end{pmatrix}$ . Therefore,

$$\begin{aligned}\tilde{\mathbf{K}}(L) &= \tilde{\Phi}(L) \begin{pmatrix} L^\tau & \\ & \mathbf{I}_{n-1} \end{pmatrix} \mathbf{C}'_2 \\ &= \Phi(L) \mathbf{C}_0 \begin{pmatrix} L^\tau & \\ & \mathbf{I}_{n-1} \end{pmatrix} \mathbf{C}'_2 \\ &= \Phi(L) \begin{pmatrix} L^\tau & \\ & \mathbf{I}_{n-1} \end{pmatrix} \mathbf{C}_0 \mathbf{C}'_2 \\ &= \Phi(L) \begin{pmatrix} L^\tau & \\ & \mathbf{I}_{n-1} \end{pmatrix} \mathbf{C}'_1 \mathbf{C}_1 \mathbf{C}_0 \mathbf{C}'_2 \\ &= \mathbf{K}(L) \mathbf{\Gamma},\end{aligned}$$

almost everywhere, with  $\mathbf{\Gamma} = \mathbf{C}_1 \mathbf{C}_0 \mathbf{C}'_2$ . Lemma 1 confirms the remainder of the proposition.  $\square$

*Proof of Proposition 2.* Define

$$\mathbf{C}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{c}{\sqrt{1+c^2}} & \frac{1}{\sqrt{1+c^2}} \\ 0 & -\frac{1}{\sqrt{1+c^2}} & \frac{c}{\sqrt{1+c^2}} \\ & & & \mathbf{I}_{n-3} \end{pmatrix}, \quad \mathbf{C}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\tilde{c}}{\sqrt{1+\tilde{c}^2}} & \frac{1}{\sqrt{1+\tilde{c}^2}} \\ 0 & -\frac{1}{\sqrt{1+\tilde{c}^2}} & \frac{\tilde{c}}{\sqrt{1+\tilde{c}^2}} \\ & & & \mathbf{I}_{n-3} \end{pmatrix},$$

and let

$$\Phi(L) = \mathbf{K}(L) \mathbf{C}_1 \begin{pmatrix} 1 & & \\ & L^{-\tau} & \\ & & \mathbf{I}_{n-2} \end{pmatrix}, \quad \tilde{\Phi}(L) = \tilde{\mathbf{K}}(L) \mathbf{C}_2 \begin{pmatrix} 1 & & \\ & L^{-\tau} & \\ & & \mathbf{I}_{n-2} \end{pmatrix}.$$

Focusing for the moment on

$$\Phi(L) = \begin{pmatrix} K_{11}(L) & \frac{cL^{-\tau}K_{12}(L)}{\sqrt{1+c^2}} & \frac{K_{12}(L)}{\sqrt{1+c^2}} & \mathbf{0} \\ K_{21}(L) & K_{21}(L) \frac{cL^{-\tau}K_{12}(L)}{\sqrt{1+c^2}K_{11}(L)} & \frac{K_{22}(L)+cK_{23}(L)}{\sqrt{1+c^2}} & \mathbf{K}_{24}(L) \\ \vdots & \vdots & \vdots & \vdots \\ K_{n1}(L) & K_{n1}(L) \frac{cL^{-\tau}K_{12}(L)}{\sqrt{1+c^2}K_{11}(L)} & \frac{K_{n2}(L)+cK_{n3}(L)}{\sqrt{1+c^2}} & \mathbf{K}_{n4}(L) \end{pmatrix}, \quad (25)$$

observe that although it is singular (with normal rank  $n-1$ ), it can be decomposed as  $\Phi(L) = \Psi(L) \Upsilon(L)$ , where  $\Psi(L)$  is  $n \times (n-1)$  and  $\Upsilon(L)$  is  $(n-1) \times n$ . In particular,

$$\Psi(L) = \begin{pmatrix} \frac{1}{\hat{\chi}(L)} & \frac{K_{12}(L)}{\sqrt{1+c^2}} & \mathbf{0} \\ \frac{K_{21}(L)}{\hat{K}_{11}(L)} & \frac{K_{22}(L)+cK_{23}(L)}{\sqrt{1+c^2}} & \mathbf{K}_{24}(L) \\ \vdots & \vdots & \vdots \\ \frac{K_{n1}(L)}{\hat{K}_{11}(L)} & \frac{K_{n2}(L)+cK_{n3}(L)}{\sqrt{1+c^2}} & \mathbf{K}_{n4}(L) \end{pmatrix}, \quad (26)$$

$$\Upsilon(L) = \begin{pmatrix} \hat{K}_{11}(L) & \frac{cL^{-\tau}\hat{K}_{12}(L)}{\sqrt{1+c^2}} & \mathbf{0} \\ 0 & 0 & \mathbf{I}_{n-2} \end{pmatrix}, \quad (27)$$

Note that  $\Psi(L)$  is a full column rank rational transfer matrix, and  $\Upsilon(L)$  is a full row rank *polynomial* matrix (since  $\widehat{K}_{11}(L)$  and  $L^{-\tau}\widehat{K}_{12}(L)$  are scalar polynomials by Assumption 7.1). In addition, since  $\widehat{K}_{11}(L)$  and  $\widehat{K}_{12}(L)$  have no common roots,  $\Upsilon(L)$  clearly has no finite zeros, and it can be further verified that it has no infinite zeros either.

Specifically,  $\Upsilon(L)$  has a zero at infinity if and only if there exists a rational vector  $\mathbf{v}(L)$  such that for  $z \in \mathbb{C}$ ,

$$0 < \lim_{z \rightarrow \infty} \|\mathbf{v}(z)\| < \infty$$

and

$$\lim_{z \rightarrow \infty} \mathbf{v}(z)' \Upsilon(z) = 0.$$

Let  $\mathbf{v}(z) = (v_1(z), \mathbf{v}_2(z))'$ . It is immediately evident that  $\lim_{z \rightarrow \infty} \mathbf{v}_2(z) = 0$  so that we must have  $\lim_{z \rightarrow \infty} v_1(z) \neq 0$ . But since both  $\lim_{z \rightarrow \infty} \widehat{K}_{11}(z)$  and  $\lim_{z \rightarrow \infty} z^{-\tau} \widehat{K}_{12}(z)$  diverge, the conditions above cannot be satisfied.

Consequently,  $\Upsilon(L)$  has no zeros (finite or infinite) and no finite poles. It does, however, have a pole at infinity of multiplicity equal to  $\max\{\deg \widehat{K}_{11}(L), \deg L^{-\tau} \widehat{K}_{12}(L)\}$ . Moreover, we can construct the rational transfer matrix

$$\Xi(L) = \Psi(L) \begin{pmatrix} \psi(L) & \\ & \mathbf{I}_{n-2} \end{pmatrix},$$

where

$$\psi(L)\psi(L^{-1}) = \widehat{K}_{11}(L)\widehat{K}_{11}(L^{-1}) + \frac{c^2}{1+c^2} \widehat{K}_{12}(L)\widehat{K}_{12}(L^{-1})$$

and

$$\Xi(L)\Xi(L^{-1})' = \Phi(L)\Phi(L^{-1})' = \mathbf{K}(L)\mathbf{K}(L^{-1})'.$$

Observe that  $\Xi(L)$  is a  $n \times (n-1)$  rational transfer matrix of normal rank  $n-1$ .

Although  $\psi(L)$  in this formulation has  $\max\{\deg \widehat{K}_{11}(L), \deg L^{-\tau} \widehat{K}_{12}(L)\}$  roots, we can always choose  $\psi(L)$  to have only roots with modulus greater than or equal to unity (since  $\psi(L)\psi(L^{-1}) = \psi(L)\varpi(L)\varpi(L^{-1})\psi(L^{-1})$  for an arbitrary  $\varpi(L)\varpi(L^{-1}) = 1$ ). On the other hand,  $\Psi(L)$ , being a *tall* transfer matrix, is zero-free for almost all  $\mathbf{K}(L)$  (Anderson and Deistler, 2008). In the non-generic case, all the zeros (finite or infinite) along with the finite poles of  $\Psi(L)$  are equivalent to the zeros (finite or infinite) along with the finite poles of  $\Phi(L)$ .

The locations of the zeros of  $\Phi(L)$  and therefore  $\Psi(L)$ , if they exist, are determined by  $\omega$ . If  $\omega = (0, \dots, 0)$ , then  $\Psi(L)$  has no zeros of modulus less than unity, and neither does  $\Xi(L)$ . Accordingly, an identical procedure applied to  $\tilde{\Phi}(L)$  yields the rational transfer matrix  $\tilde{\Xi}(L)$  with no zeros of modulus less than unity, given  $\omega = (0, \dots, 0)$ . In this case, Baggio and Ferrante (2016, Theorem 2) confirms that  $\tilde{\Xi}(L) = \Xi(L)\mathbf{C}_0$ , where  $\mathbf{C}_0$  is a constant orthogonal matrix.<sup>10</sup>

<sup>10</sup>Note that Theorem 2 in Baggio and Ferrante (2016) is stated using the parameterization  $\Xi(z) = \Xi_0 + \Xi_1 z^{-1} + \dots + \Xi_l z^{-l} + \dots$  and the requirement that  $\Xi(z)$  has no zeros in  $\{z \in \mathbb{C} : |z| > 1\}$ ; here we work with  $\Xi(z) = \Xi_0 + \Xi_1 z + \dots + \Xi_l z^l + \dots$ , in which case the same requirement corresponds to  $\Xi(z)$  having no zeros in  $\{z \in \mathbb{C} : |z| < 1\}$ .

If  $\omega \neq (0, \dots, 0)$ , then  $\Xi(L)$  will have some zeros inside the unit circle. However, we can apply the arguments in Lippi and Reichlin (1994, Proof of Theorem 3) to obtain a rational transfer matrix  $\mathbf{N}(L)$  with no zeros of modulus less than unity and

$$\mathbf{N}(z) = \Xi(z)\mathbf{D}(\bar{z}^{-1})^*,$$

where  $\mathbf{D}(z)$  is a Blaschke matrix containing factors

$$\mathbf{R}(\alpha_h, z) = \begin{pmatrix} \frac{z - \alpha_h}{1 - \bar{\alpha}_h z} & \\ & \mathbf{I}_{n-2} \end{pmatrix}$$

for each zero  $\alpha_h$  of  $\Xi(z)$  satisfying  $|\alpha_h| < 1$ . Manipulating  $\tilde{\Xi}(L)$  in an identical way yields  $\tilde{\mathbf{N}}(L)$  with no zeros of modulus less than unity.

Baggio and Ferrante (2016, Theorem 2) may now be applied to obtain  $\tilde{\mathbf{N}}(L) = \mathbf{N}(L)\mathbf{C}_{00}$  for a constant orthogonal matrix  $\mathbf{C}_{00}$ , which also implies that  $\mathbf{N}(L)$  and  $\tilde{\mathbf{N}}(L)$  have identical zeros. Therefore,  $\Xi(L)$  and  $\tilde{\Xi}(L)$  corresponding to the same  $\omega$  must also have identical zeros. Applying again the arguments in Lippi and Reichlin (1994, Proof of Theorem 3), we obtain that  $\tilde{\Xi}(L) = \Xi(L)\mathbf{C}_0$  for some constant orthogonal matrix  $\mathbf{C}_0$ .

From  $\tilde{\Xi}(0) = \Xi(0)\mathbf{C}_0$  and given that  $\tilde{\Xi}_{1j}(0) = \Xi_{1j}(0) = 0$  for all  $j \geq 2$ , it follows that  $C_{0,1j} = C_{0,j1} = 0$  for all  $j \geq 2$  and  $|C_{0,11}| = 1$ . Accordingly, since for some  $z \neq 0$ ,  $\tilde{\Xi}_{12}(z) \neq 0$ ,  $\Xi_{12}(z) \neq 0$  and  $\tilde{\Xi}_{1j}(z) = \Xi_{1j}(z) = 0$  for all  $j \geq 3$ , we obtain  $C_{0,2j} = C_{0,j2} = 0$  and  $|C_{0,22}| = 1$ .

Therefore,

$$\frac{\tilde{\psi}(L)}{\tilde{\chi}(L)} = \pm \frac{\psi(L)}{\chi(L)}, \quad (28)$$

$$\frac{\tilde{K}_{12}(L)}{\sqrt{1 + \tilde{c}^2}} = \pm \frac{K_{12}(L)}{\sqrt{1 + c^2}}, \quad (29)$$

$$\frac{\tilde{\psi}(L)\tilde{K}_{i1}(L)}{\tilde{\hat{K}}_{11}(L)} = \pm \frac{\psi(L)K_{i1}(L)}{\hat{K}_{11}(L)}, \quad i = 2, \dots, n \quad (30)$$

$$\frac{K_{i2}(L) + cK_{i3}(L)}{\sqrt{1 + c^2}} = \pm \frac{\tilde{K}_{i2}(L) + \tilde{c}\tilde{K}_{i3}(L)}{\sqrt{1 + \tilde{c}^2}}, \quad i = 1, \dots, n. \quad (31)$$

Applying Assumption 7.2,

$$\frac{\sqrt{1 + \tilde{c}^2}}{\tilde{c}}\tilde{K}_{i3}(L) + \frac{\tilde{K}_{i1}(L)}{\tilde{\hat{K}}_{11}(L)}\frac{\tilde{K}_{12}(L)}{\sqrt{1 + \tilde{c}^2}} = \pm \left( \frac{\sqrt{1 + c^2}}{c}K_{i3}(L) + \frac{K_{i1}(L)}{\hat{K}_{11}(L)}\frac{K_{12}(L)}{\sqrt{1 + c^2}} \right),$$

and using the above equalities,

$$\begin{aligned} \frac{\tilde{K}_{i1}(L)}{\tilde{\hat{K}}_{11}(L)} &= \frac{\frac{\tilde{\psi}(L)\tilde{K}_{i1}(L)}{\tilde{\hat{K}}_{11}(L)}}{\frac{\tilde{\psi}(L)}{\tilde{\chi}(L)}} = \frac{\frac{\psi(L)K_{i1}(L)}{\hat{K}_{11}(L)}}{\frac{\psi(L)}{\chi(L)}} = \frac{K_{i1}(L)}{\hat{K}_{11}(L)}, \\ \frac{\tilde{K}_{i1}(L)}{\tilde{\hat{K}}_{11}(L)}\frac{\tilde{K}_{12}(L)}{\sqrt{1 + \tilde{c}^2}} &= \pm \frac{K_{i1}(L)}{\hat{K}_{11}(L)}\frac{K_{12}(L)}{\sqrt{1 + c^2}}, \end{aligned} \quad (32)$$

where the  $\pm$  in the last line is the sign of  $C_{0,22}$ , which is the same as the  $\pm$  in (31).

It follows that

$$\frac{\sqrt{1+\tilde{c}^2}}{\tilde{c}} \tilde{K}_{i3}(L) = \pm \frac{\sqrt{1+c^2}}{c} K_{i3}(L).$$

Then, applying  $\mathbf{g}$  to both sides of this equality and observing that

$$\frac{\tilde{K}_{\tau^*,12}}{\sqrt{1+\tilde{c}^2}} = \pm \frac{K_{\tau^*,12}}{\sqrt{1+c^2}}$$

yields  $\tilde{c} = c$ . Consequently, straightforward algebra leads to

$$\tilde{K}_{i1}(L) = \pm K_{i1}(L), \quad (33)$$

$$\tilde{K}_{i2}(L) = \pm K_{i2}(L), \quad (34)$$

$$\tilde{K}_{i3}(L) = \pm K_{i3}(L), \quad (35)$$

$$\tilde{\mathbf{K}}_{i4}(L) = \mathbf{K}_{i4}(L) \mathbf{C}_{0,33}, \quad (36)$$

for  $i = 1, \dots, n$  and where  $\mathbf{C}'_{0,33} \mathbf{C}_{0,33} = \mathbf{C}_{0,33} \mathbf{C}'_{0,33} = \mathbf{I}_{n-3}$ . Note that (33) follows from  $\tilde{K}_{11}(L) \tilde{K}_{11}(L^{-1}) = K_{11}(L) K_{11}(L^{-1})$  and the assumption that  $K_{11}(L)$  has no roots of modulus less than unity. The equality for  $i = 2, \dots, n$  then follows directly from (32). Finally, setting

$$\mathbf{\Gamma} = \begin{pmatrix} \pm 1 & & & \\ & \pm 1 & & \\ & & \pm 1 & \\ & & & \mathbf{C}_{0,33} \end{pmatrix}$$

yields  $\tilde{\mathbf{K}}(L) = \mathbf{K}(L) \mathbf{\Gamma}$  as desired.  $\square$

## B Econometric Methods

### B.1 Imposing Structural Identification Restrictions

We begin by detailing Step 5 of the algorithm outlined in Section 5, which entails implementing restrictions R2-R4. In fact, this 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  $j = 0, \dots, q$ . In our implementation, we set the horizon in R3 to be 20 quarters after impact.

Accordingly,  $\mathbf{\Gamma}_1$  is determined by setting the first column  $\mathbf{\Gamma}_{1,1} = \tilde{\mathbf{A}}'_{0,(1)} / \|\tilde{\mathbf{A}}_{0,(1)}\|^2$ , where  $\tilde{\mathbf{A}}_{0,(1)}$  denotes the first row of  $\tilde{\mathbf{A}}_0$ , and the remaining columns  $\mathbf{\Gamma}_{1,i}$  for  $i = 2, \dots, n$  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  $\mathbf{E}_2$ . The orthogonal matrix that identified news shocks according to restriction R3 is then given by

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

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 R2 and R3, 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

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

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

By construction, the  $n \times n - 1$  matrix  $(\check{\mathbf{A}}_{0,2}, \dots, \check{\mathbf{A}}_{0,n})$  has rank  $n - 2$  and, therefore, there exists a  $n - 1 \times 1$  vector  $\zeta_A = (\zeta_{A,1}, \zeta'_{A,2})'$ ,  $\|\zeta_A\| = 1$  such that<sup>11</sup>

$$(\check{\mathbf{A}}_{0,2}, \dots, \check{\mathbf{A}}_{0,n})\zeta_A = 0.$$

Accordingly, set

$$c = \frac{|\zeta_{A,1}|}{1 - \zeta_{A,1}^2} \tag{37}$$

$$\mathbf{E}_{3,1} = -\text{sgn}(\zeta_{A,1}) \frac{\zeta_{A,2}}{1 - \zeta_{A,1}^2}, \tag{38}$$

and the remaining columns  $\mathbf{E}_{3,2}, \dots, \mathbf{E}_{3,n-2}$  of  $\mathbf{E}_3$  to be the  $n - 3$  vectors orthogonal to  $\mathbf{E}_{3,1}$  (normalized such that  $\|\mathbf{E}_{3,i}\| = 1$ ). Subsequently, multiplying  $\mathbf{A}_j = \check{\mathbf{A}}_j\mathbf{\Gamma}_3$  for all  $j = 0, \dots, q$  yields the desired representation  $\mathbf{A}(L)$  where  $\mathbf{A}_{0,3} = c\mathbf{A}_{0,2}$  while preserving the restrictions R2 and R3.

As mentioned in Section 5, the algorithm involves finding an  $n \times r_*$  matrix  $\mathbf{\Delta}$  such that  $\mathbf{\Theta}_j\mathbf{\Theta}_0\mathbf{\Delta} = 0$  for all  $j = q - \tau + 1, \dots, q$ . When  $r_* > 1$ , this matrix is not unique as there exist infinitely many orthonormal matrices  $\mathbf{C}_\Delta$  such that  $\tilde{\mathbf{\Delta}} = \mathbf{\Delta}\mathbf{C}_\Delta$  is also a valid alternative. To obtain a unique  $\mathbf{\Delta}$  in the  $r_* > 1$  case, one may choose  $\mathbf{\Delta}$  to maximize the fraction of forecast error variance of TFP that is *jointly* explained by non-news and noise shocks at 20 quarters after impact.

Note that this objective is in line with the theoretical Assumption 3. Starting with an arbitrary  $\mathbf{\Delta}$  obtained in Step 3, the maximization may be performed numerically (over  $\mathbf{C}_\Delta$ ) by iterating Step 4 and Step 5, which themselves involve computationally simple operations. Of course, the complexity of the numerical optimisation approach

---

<sup>11</sup>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  $n$ .

will vary with  $r_*$  (for example, for  $r_* = 2$ ,  $\mathbf{C}_\Delta$  has only one free parameter, and hence, the optimisation is univariate). In all our empirical work with this algorithm, however, we never encountered a need to perform numerical optimisation to find a unique  $\Delta$  as every draw of  $\Theta_j$  from all our models satisfied  $\text{rank } \Theta_j = n - 1$ .

## B.2 Bayesian Algorithms

### B.2.1 Priors

A key advantage of the Bayesian approach in time-series modelling has proven to be the ability to incorporate information through prior probability distributions. 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 *stochastic search variable selection* (SSVS) prior introduced to the Bayesian VAR literature by George, Sun, and Ni (2008), and used in many VAR papers (e.g. Koop, 2013; Korobilis, 2013). The basic idea can be explained in terms of a generic VAR or MA coefficient, say  $\vartheta$ . 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:

$$(\vartheta | \nu) \sim (1 - \nu) \mathcal{N}(0, v_0^2) + \nu \mathcal{N}(0, v_1^2), \quad (39)$$

where  $\nu \in \{0, 1\}$  is an unknown parameter. If  $\nu = 0$  the prior for  $\vartheta$  is given by the first Normal distribution, and if it is  $\nu = 1$  its prior is given by the second. The prior is hierarchical since  $\nu$  is treated as an unknown parameter that is estimated in a data-based fashion. The first prior variance,  $v_0^2$ , is chosen to be ‘small’ (so that the coefficient is constrained to be virtually equal to zero) and the second prior variance,  $v_1^2$ , to be ‘large’ (implying a relatively non-informative 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  $v_0^2$  and  $v_1^2$ , 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 the following subsection.

### B.2.2 Bayesian VARMA

The ultimate goal of a Bayesian approach to estimating the effects of news and noise shocks is to obtain draws from the posterior distribution of representation 6. 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 (6). Such an approach has the advantage of allowing restrictions on impulse responses to be imposed directly in the sampling. However, it is computationally intensive and the requisite sampling algorithm deteriorates for systems larger than three variables.

Since in our applications we wish to estimate models potentially involving tens of variables, it is more appropriate to work with a finite order VARMA representation, such as the VARMA( $p, q$ ) specified in (15). As discussed in Section 5, however, even in

the reduced form (16) the parameters of  $\mathbf{B}(L)$  and  $\boldsymbol{\Theta}(L)$  will generally not be identified without further restrictions, and the same holds for  $\mathbf{B}(L)$  and  $\mathbf{A}(L)$  in the structural form. The simple reason for that is as follows: even though  $\mathbf{K}(L) = \mathbf{B}(L)^{-1}\mathbf{A}(L)$  is uniquely determined by identifying restrictions such as R2-R4 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 VARMA are further complicated by the fact that the fundamental and numerous non-fundamental representations are observationally equivalent, as discussed in sub-section 4.1.2. 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 is achieved by rewriting (16) in recursive form (17) and imposing two types of restrictions:

1. exclusion restrictions on  $\tilde{\mathbf{B}}_0, \dots, \tilde{\mathbf{B}}_{p^*}, \mathbf{M}_1, \dots, \mathbf{M}_{p^*}$  according to the *row degrees*  $p_1, \dots, p_n$  that define the lag structure of each equation in the system (with  $p^* = \max(p_1, \dots, p_n)$ );
2. non-linear restrictions on  $\mathbf{M}_1, \dots, \mathbf{M}_{p^*}$  to ensure all roots of  $\mathbf{M}(L)$  lie outside the unit circle.

However, estimating a VARMA in *echelon* 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  $p_1, \dots, p_n$ , which themselves need to be estimated in practice.<sup>12</sup>

Fortunately, point identification is not necessary in the Bayesian framework. To explain, 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  $\mathbf{B}_1, \dots, \mathbf{B}_p, \tilde{\boldsymbol{\Theta}}_1, \dots, \tilde{\boldsymbol{\Theta}}_q$ , and  $\boldsymbol{\Sigma}$ , we can readily obtain draws from

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

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).

The key insight in a Bayesian approach to analyzing VARMA models is that parameters  $\mathbf{B}_1, \dots, \mathbf{B}_p, \boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_q$ , and  $\boldsymbol{\Sigma}$  themselves are not of primary interest, but rather quantities such as forecasts and impulse 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

---

<sup>12</sup>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 Luetkepohl (2005) for a textbook treatment, including further details and explicit formulae.

van Dyk, 2005; Ghosh and Dunson, 2009; Koop, León-González, and Strachan, 2010, 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 (2017) and Chan, Eisenstat, and Koop (2016) develop MCMC algorithms on the *expanded* VARMA representation:

$$\tilde{\mathbf{B}}_0 \mathbf{y}_t = \tilde{\mathbf{B}}_1 \mathbf{y}_{t-1} + \cdots + \tilde{\mathbf{B}}_{p^*} \mathbf{y}_{t-p^*} + \mathbf{\Pi}_0 \mathbf{f}_t + \mathbf{\Pi}_1 \mathbf{f}_{t-1} + \cdots + \mathbf{\Pi}_{p^*} \mathbf{f}_{t-p^*} + \boldsymbol{\eta}_t, \quad (40)$$

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  $\mathbf{\Pi}_0$  is lower triangular with ones on the diagonal. Expanded form parameters are related to the VARMA parameters in (17) by the mapping:

$$\sum_{l=j}^{p^*} \boldsymbol{\Theta}_l \boldsymbol{\Sigma} \boldsymbol{\Theta}'_{l-j} = \sum_{l=j}^{p^*} \mathbf{\Pi}_l \mathbf{\Omega} \mathbf{\Pi}'_{l-j} + \mathbf{1}(j=0) \mathbf{\Lambda}, \quad \text{for all } j = 0, \dots, p^*, \quad (41)$$

whereas  $\tilde{\mathbf{B}}_j$  in the expanded form is identical to the corresponding  $\tilde{\mathbf{B}}_j$  in the semi-structural form for  $j = 0, \dots, p^*$ . Consequently, draws from (17) can be obtained by sampling directly from the expanded form (40) and then computing  $\mathbf{M}_1, \dots, \mathbf{M}_{p^*}$ ,  $\boldsymbol{\Sigma}$  from each draw of  $\mathbf{\Pi}_0, \dots, \mathbf{\Pi}_{p^*}$ ,  $\mathbf{\Omega}$ , and  $\mathbf{\Lambda}$  using the mapping in (41). The exact procedure based on *generalized Schur decompositions* / *generalized eigenvalues* is provided in Section 2.3 of Chan and Eisenstat (2017) and Appendix D of Chan, 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  $\mathbf{M}_1, \dots, \mathbf{M}_{p^*}$ ,  $\boldsymbol{\Sigma}$  from  $\mathbf{\Pi}_0, \dots, \mathbf{\Pi}_{p^*}$ ,  $\mathbf{\Omega}$ , and  $\mathbf{\Lambda}$ ). It is worth emphasizing that this also facilitates building set-identified IRFs by randomly selecting of the valid fundamental / non-fundamental representations to construct.

At the same time, the expanded form provides an extremely flexible approach to estimating VARMA systems. For example, Chan, Eisenstat, and Koop (2016) demonstrate how to construct a prior on the expanded form parameters—using SSVS methods (see Kuo and Mallick, 1997; George, Sun, and Ni, 2008)—such that the implied draws from the recursive form (17) 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 VARMA systems. 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 (40) with  $p = q = p^*$ , we impose parsimony by first setting (with probability one)

$$\Pi_{p^*-\tau+1,ni} = \cdots = \Pi_{p^*,ni} = 0, \quad i = 1, \dots, n,$$

which corresponds to restrictions (18) on  $\mathbf{M}_{p^*-\tau+1}, \dots, \mathbf{M}_{p^*}$ . This, in turn, leads to the restrictions

$$\Theta_{p^*-\tau+1,ni} = \cdots = \Theta_{p^*,ni} = 0, \quad i = 1, \dots, n,$$

such that  $\Theta_{p^*-\tau+1}, \dots, \Theta_{p^*}$  are singular matrices. Next, we specify SSVS priors on the individual free elements of  $\tilde{\mathbf{B}}_0, \dots, \tilde{\mathbf{B}}_{p^*}$  and  $\Pi_0, \dots, \Pi_{p^*}$  of the form:

$$\begin{aligned} (\tilde{B}_{j,ik} | \gamma_{j,ik}^B) &\sim \gamma_{j,ik}^B \mathcal{N}(0, 1) + (1 - \gamma_{j,ik}^B) \mathcal{N}(0, 0.01), \\ (\Pi_{j,ik} | \gamma_{j,ik}^\Pi) &\sim \gamma_{j,ik}^\Pi \mathcal{N}(0, 1) + (1 - \gamma_{j,ik}^\Pi) \mathcal{N}(0, 0.01), \end{aligned}$$

subject to the restriction that  $\det \mathbf{B}(L)$  has no roots inside the unit circle, and with

$$\Pr(\gamma_{j,ik}^B = 1) = \Pr(\gamma_{j,ik}^\Pi = 1) = \begin{cases} 0.5 & \text{if } n < 6, \\ 0.4 & \text{if } n = 6, \\ 0.2 & \text{if } n = 10, \\ \frac{1.5}{n} & \text{if } n > 10. \end{cases}$$

Note that we specify  $\Pr(\gamma_{j,ik}^B = 1)$  and  $\Pr(\gamma_{j,ik}^\Pi = 1)$  as decreasing functions of  $n$  in line with the theory developed in Mol, Giannone, and Reichlin (2008) and empirical evidence reported in Korobilis (2013).

Through extensive experimentation with the resulting algorithm, we find these settings to produce satisfactory. 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

$$\begin{aligned} \Omega_{ii} &\sim \mathcal{IG}(5, 1), \\ \Lambda_{ii} &\sim \mathcal{IG}(10^{-12}, 0.1) \mathbf{1}(\Lambda_{ii} \leq 50), \end{aligned}$$

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  $\Omega_{ii}$  and extremely heavy-tailed but proper priors on  $\Lambda_{ii}$ . In the paper, we report results holding fixed all of the above prior settings, but varying the dimension of the system  $n$ .

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

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

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  $(\gamma_i, \tilde{\mathbf{B}}_{(i)}, \mathbf{\Pi}_{(i)} | \mathbf{f}, \Lambda_{ii}, \mathbf{y}_i)$  for each  $i = 1, \dots, n$ , where  $\tilde{\mathbf{B}}_{(i)}$  denotes the  $i$ -th row of  $\tilde{\mathbf{B}} = (\mathbf{I}_n - \tilde{\mathbf{B}}_0, \tilde{\mathbf{B}}_1, \dots, \tilde{\mathbf{B}}_{p^*})$ ,  $\mathbf{\Pi}_{(i)}$  the  $i$ -th row of  $\mathbf{\Pi} = (\mathbf{\Pi}_0, \dots, \mathbf{\Pi}_{p^*})$ , and  $\gamma_i$  is the set of all SSVS indicators pertaining to  $\tilde{\mathbf{B}}_{(i)}, \mathbf{\Pi}_{(i)}$ .
2. Sample  $(\Lambda_{ii} | \tilde{\mathbf{B}}_{(i)}, \mathbf{\Pi}_{(i)}, \gamma_i, \mathbf{f}, \mathbf{y}_i)$  for each  $i = 1, \dots, n$ .
3. Sample  $(\Omega_{ii} | \mathbf{f}_i)$  for each  $i = 1, \dots, n$ .
4. Sample  $(\mathbf{f} | \tilde{\mathbf{B}}, \mathbf{\Pi}, \Omega, \Lambda, \gamma, \mathbf{y})$ .

Details and extensive discussion of each sampling step above are provided in Appendix B of Chan, Eisenstat, and Koop (2016). If set identified IRFs are desired, then we also add

5. Sample  $(\varrho | \mathbf{\Pi}, \Omega, \Lambda)$  using rejection: obtain a uniform draw from 1 to  $2^{np^*}$  and accept it *if and only if* the roots configuration  $\omega_\varrho$  leads to a valid representation; continue until a draw of  $\varrho$  is accepted.<sup>13</sup>

Note that this extra step is computationally fast because the *generalized Schur decompositions / generalized eigenvalues* algorithm discussed above yields all  $nk^*$  roots. Given  $\omega_\varrho$ , the combination of roots is checked that every complex root is paired with its conjugate.<sup>14</sup>

Following the MCMC, each draw of the expanded form parameters is transform to a draw of  $\mathbf{B}_1, \dots, \mathbf{B}_p, \tilde{\mathbf{\Theta}}_1, \dots, \tilde{\mathbf{\Theta}}_q$  and  $\mathbf{\Sigma}$ , which constitutes Step 1 of the algorithm outlined in Section 5. Finally, Step 2-Step 5 applied to each draw of  $\tilde{\mathbf{\Theta}}_1, \dots, \tilde{\mathbf{\Theta}}_q$  yield a draw of  $\mathbf{A}_0, \dots, \mathbf{A}_q$  and  $\mathbf{K}(L) = \mathbf{B}(L)^{-1} \mathbf{A}(L)$  provides a draw of the IRFs.

### B.2.3 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 Titterton, 2006). Given a likelihood function  $f(\mathbf{y} | \boldsymbol{\theta})$ , the DIC is defined as:

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

where

$$\overline{D(\boldsymbol{\theta})} = -2\mathbb{E}_{\boldsymbol{\theta}}[\log f(\mathbf{y} | \boldsymbol{\theta}) | \mathbf{y}]$$

<sup>13</sup>In practice, we set  $\omega_\varrho$  to be the  $(nk^*)$ -bit binary representation of the decimal number  $2^{np^*} - 1$ , which is straightforward to compute—e.g., using Matlab’s `dec2bin` command.

<sup>14</sup>In a non-fundamental representation, if a complex root is “flipped” then its conjugate must also be flipped for the representation to be valid.

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

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

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

Following Chan, Eisenstat, and Koop (2016), we use the likelihood implied by the system in (16), or equivalently

$$\mathbf{y}_t = \sum_{j=1}^p \mathbf{B}_j \mathbf{y}_{t-j} + \sum_{j=1}^q \boldsymbol{\Theta}_j \boldsymbol{\epsilon}_{t-j} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad (42)$$

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

To derive this density, we stack (42) over  $t$  and obtain:

$$\mathbf{y} = \mathbf{a} + \boldsymbol{\Theta} \boldsymbol{\epsilon}, \quad (43)$$

where  $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}'_1, \dots, \boldsymbol{\epsilon}'_T)' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes \boldsymbol{\Sigma})$ ,  $\mathbf{b} = ((\sum_{j=1}^p \mathbf{B}_j \mathbf{y}_{1-j})', \dots, (\sum_{j=1}^p \mathbf{B}_j \mathbf{y}_{T-j})')'$  and  $\boldsymbol{\Theta}$  is a  $Tn \times Tn$  lower triangular matrix with the identity matrix  $\mathbf{I}_n$  on the main diagonal block,  $\boldsymbol{\Theta}_1$  on first lower diagonal block,  $\boldsymbol{\Theta}_2$  on second lower diagonal block, and so forth. Hence, we have

$$(\mathbf{y} | \mathbf{B}_1, \dots, \mathbf{B}_p, \boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_q, \boldsymbol{\Sigma}) \sim \mathcal{N}(\mathbf{b}, \boldsymbol{\Theta}(\mathbf{I}_T \otimes \boldsymbol{\Sigma})\boldsymbol{\Theta}').$$

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

## C News, Noise and Mixed Representations

Consider (as in Section 3) a fundamental process defined by

$$a_t = \epsilon_t^{\text{NN}} + \epsilon_{t-1}^{\text{NE}} + \theta \epsilon_{t-2}^{\text{NE}}, \quad (44)$$

along with the signal process defined by

$$s_t = \epsilon_t^{\text{NE}} + u_t. \quad (45)$$

We show that the resulting mixed representation is unique for systems that contain observations of the fundamental  $a_t$  and future expectations  $\hat{a}_{j,t} \equiv \mathbb{E}_t(a_{t+j})$ ,  $j \geq 1$ .

Indeed, with  $\theta \neq 0$ , agents whose time- $t$  information set consists of the history  $\{a_{t-\tau}, s_{t-\tau}\}_{\tau=0}^{\infty}$  formulate expectations at time- $t$  that are given by

$$\hat{a}_{1,t} \equiv \mathbb{E}_t(a_{t+1}) = \kappa_0 s_t + \theta \kappa_1 a_t + \theta \kappa_2 s_{t-1}, \quad (46)$$

$$\hat{a}_{2,t} \equiv \mathbb{E}_t(a_{t+2}) = \theta \kappa_0 s_t, \quad (47)$$

$$\hat{a}_{\tau,t} \equiv \mathbb{E}_t(a_{t+\tau}) = 0, \quad \tau \geq 3, \quad (48)$$

where

$$\kappa_0 = \frac{\sigma_{\text{NE}}^2}{\sigma_{\text{NE}}^2 + \sigma_u^2}, \quad \kappa_1 = \frac{\kappa_0 \sigma_u^2}{\theta^2 \sigma_{\text{NE}}^2 + (2 - \kappa_0) \sigma_{\text{NN}}^2}, \quad \kappa_2 = \frac{\kappa_0 (\theta^2 \sigma_{\text{NE}}^2 + \sigma_{\text{NN}}^2)}{\theta^2 \sigma_{\text{NE}}^2 + (2 - \kappa_0) \sigma_{\text{NN}}^2}.$$

The fundamentals and expectations have the following important properties:

$$E(\hat{a}_{2,t} a_t) = 0, \quad E(\hat{a}_{1,t} a_{t-1}) = \theta \kappa_1 \sigma_{\text{NE}}^2. \quad (49)$$

Now, consider a potential news representation for  $a_t$ , which by definition has the form:

$$a_t = \eta_{0,t} + \eta_{1,t-1} + \eta_{2,t-2} + \eta_{3,t-3} + \cdots,$$

where each  $\{\eta_{j,t}; t \in \mathbb{Z}\}$  is a stationary, Gaussian, mean-zero process with unconditional variance  $\sigma_{\eta_j}^2 \geq 0$ . Moreover,  $E(\eta_{j,t} \eta_{k,t-\tau}) = 0$  for all  $j \neq k$  and all  $\tau \in \mathbb{Z}$ . However, each  $\eta_{j,t}$  is potentially correlated with  $\eta_{j,t-\tau}$  for  $\tau \neq 0$ . In the news representation, agents' information consists of the space spanned by  $\{\eta_{j,t-\tau}\}_{j,\tau=0}^\infty$ . This implies the expectation

$$\begin{aligned} E(\hat{a}_{2,t} a_t) &= E(E_t(\eta_{0,t+2}) \eta_{0,t}) + E(E_t(\eta_{1,t+1}) \eta_{1,t-1}) + \sum_{j=2}^{\infty} E(\eta_{j,t+2-j} \eta_{j,t-j}) \\ &= E(\eta_{0,t+2} \eta_{0,t}) + E(\eta_{1,t+1} \eta_{1,t-1}) + \sum_{j=2}^{\infty} E(\eta_{j,t+2-j} \eta_{j,t-j}) \end{aligned} \quad (50)$$

$$= E(\eta_{0,t+1} \eta_{0,t-1}) + E(\eta_{1,t} \eta_{1,t-2}) + \sum_{j=2}^{\infty} E(\eta_{j,t+1-j} \eta_{j,t-1-j}) \quad (51)$$

$$\begin{aligned} &= E(E_t(\eta_{0,t+1}) \eta_{0,t-1}) + E(\eta_{1,t} \eta_{1,t-2}) + \sum_{j=2}^{\infty} E(\eta_{j,t+1-j} \eta_{j,t-1-j}) \\ &= E(\hat{a}_{1,t} a_{t-1}), \end{aligned} \quad (52)$$

where (50) and (52) follow from iterated expectations along with the fact that  $\eta_{j,t-\tau}$  belongs the agents' time- $t$  information set for all  $\tau \geq 0$ , and (51) follows from  $\eta_{j,t}$  being stationary and Gaussian.

Since,  $E(\hat{a}_{2,t} a_t)$  and  $E(\hat{a}_{1,t} a_{t-1})$  must satisfy (49) in the mixed representation, observational equivalence only holds if either  $\theta = 0$  or  $\sigma_u^2 = 0$  (which holds if and only if  $\kappa_1 = 0$ ). Of course if  $\sigma_u^2 = 0$  then noise is not present in the model and (44)-(45) reduces to a news representation. When  $\theta \neq 0$ , on the other hand, there exists a “pure news” representation that is observationally equivalent to a “pure noise” representation, but *neither of these* will be observationally equivalent to the mixed representation defined by (44) and (45). Hence,  $\theta \neq 0$  is the key feature that breaks the observational equivalence between this mixed representation and any potential news / noise pair.

## D The Data

We use John Fernald's purified TFP series available from the San Francisco Fed's website. 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 GDP (GDPC96) is from the U.S. Department of Commerce: Bureau of Economic Analysis. The seasonally adjusted series for consumption of non-durables and services, real chain-weighted investment, 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.

The seasonally adjusted civilian unemployment rate (UNRATE) is from the U.S. Department of Labor: Bureau of Labor Statistics. The vacancy rate is computed from the help wanted index. The industrial production index (INDPRO), real M2 (M2REALx), USD/GBP exchange rate (EXUSUKx), and the University of Michigan consumer sentiment index (UMCSENTx) are obtained directly from the FRED-QD database, dated 19 August 2018.

All the preceding 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 all the government bond yield (TB3MS, GS1, GS3, GS5, GS10) 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. 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 non-institutional population (CNP16OV), which is used in transforming some of the variables, is from the U.S. Department of Labor: Bureau of Labor Statistics.

## E Additional Empirical Results

In this section, we present additional empirical results for models with  $n = 3$ ,  $n = 10$  and  $n = 15$  variables. The FEVDs are in Figures 7-12 while the IRFs are in Figures 13-18. The results for  $n = 10$  and  $n = 15$  models highlight the large degree of similarities to the FEVDs and IRFs reported in Section 6 for the  $n = 6$  and  $n = 20$  models. Hence, inference about the effects of news and noise shocks, and particularly our conclusion that noise plays an overall minor role in driving business cycles is fairly robust to the inclusion of more than six variables in the model. However, including more variables is preferred by the data (see our DIC results) and also sharpens the conclusion that news shocks are the most significant drivers of economic activity.

In contrast, the posterior distributions for IRFs, and to a larger extent, the FEVDs in the  $n = 3$  case are rather erratic. Indeed, many of the posteriors exhibit multimodality and otherwise non-standard shapes. This suggests a large degree of uncertainty in identifying the effects of non-news, news, and noise shocks with only TFP, consumption, and stocks as the observed variables. Hence, reliable inference regarding these effects requires taking into account additional information, such as the variables included in our larger models.

It is also worth pointing out that with only three variables, we estimate that noise

shocks explain a large fraction of the forecast error variance of stocks even in the  $\mathcal{R}_2$  representation. This is in contrast to all  $n \geq 6$  models. On the other hand, the  $n = 3$  specification agrees with all the other models in estimating no effects of noise shocks on consumption.

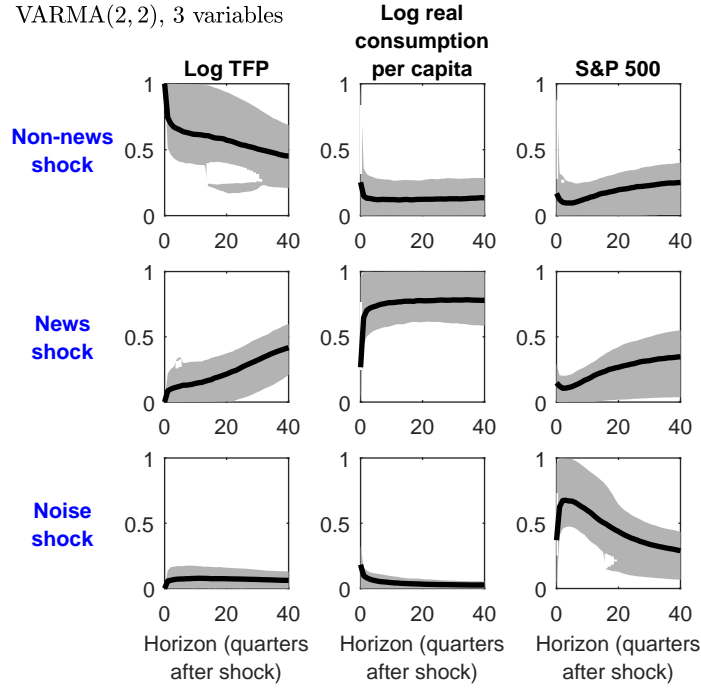
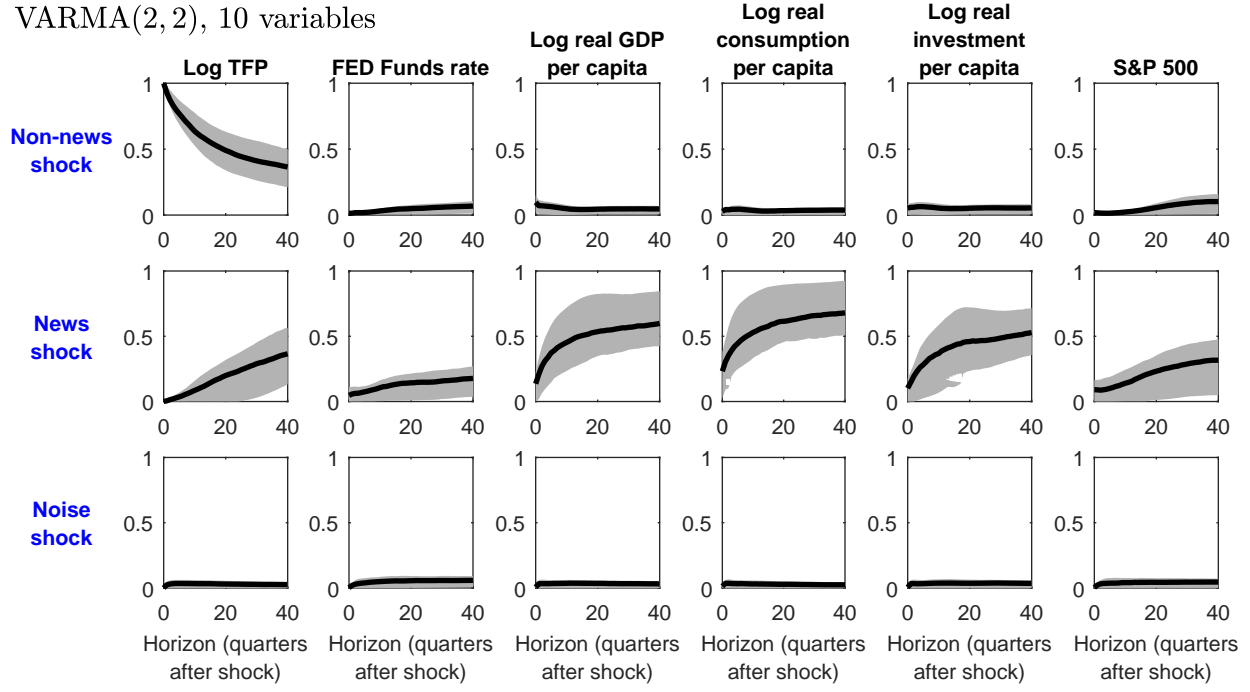


Figure 7: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_2$  (all non-zero, finite roots outside the unit circle). Top panel:  $n = 10$ , bottom panel:  $n = 15$ .

VARMA(2, 2), 10 variables



VARMA(2, 2), 15 variables

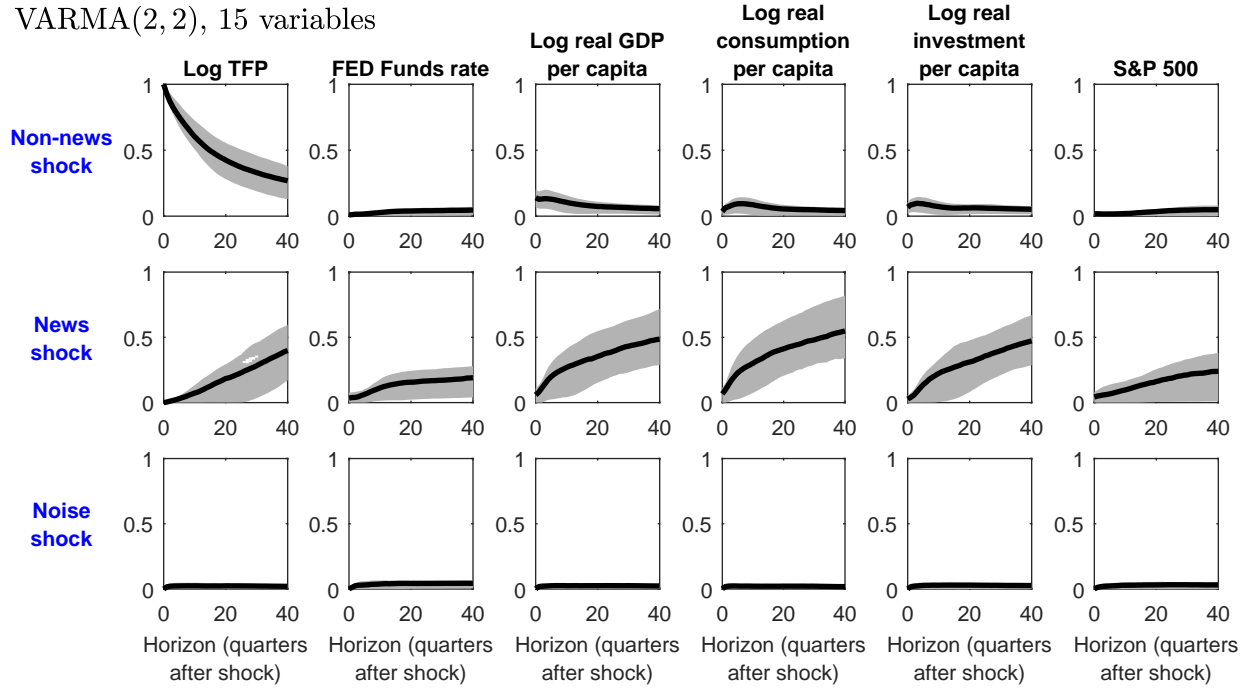


Figure 8: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_2$  (all non-zero, finite roots outside the unit circle).

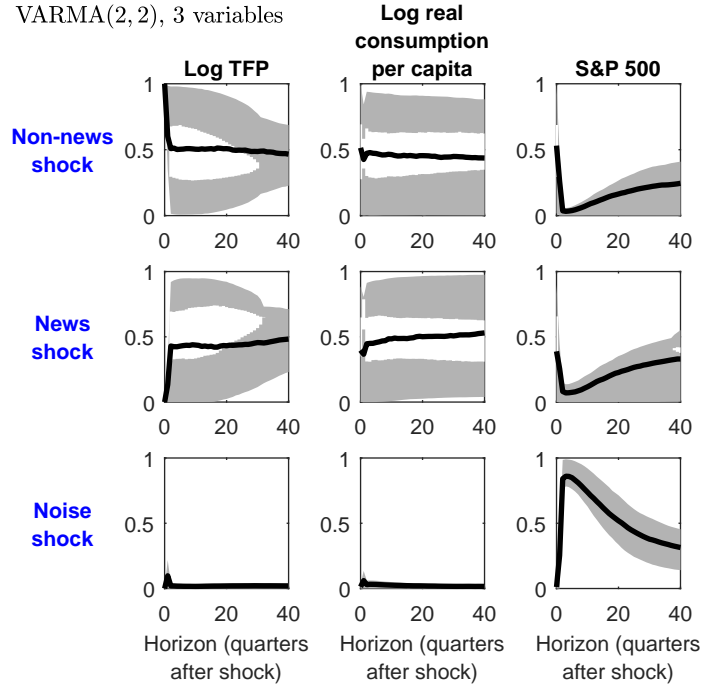
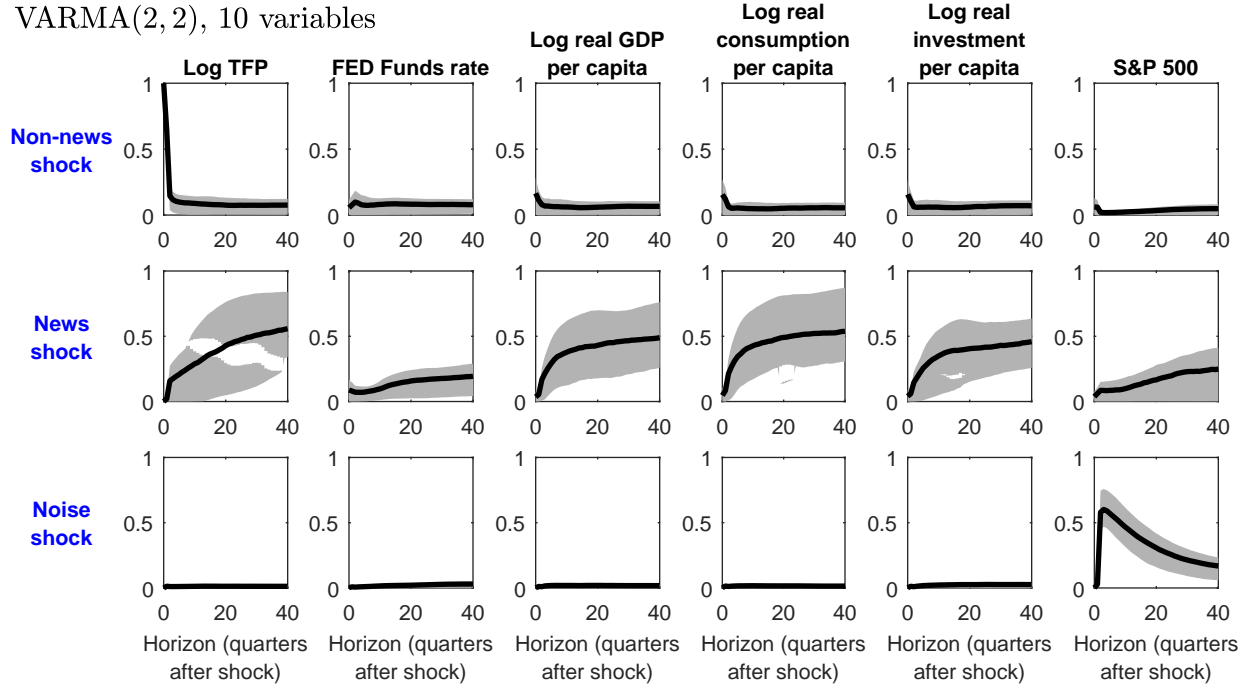


Figure 9: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_\infty$  (all finite roots inside the unit circle).

VARMA(2, 2), 10 variables



VARMA(2, 2), 15 variables

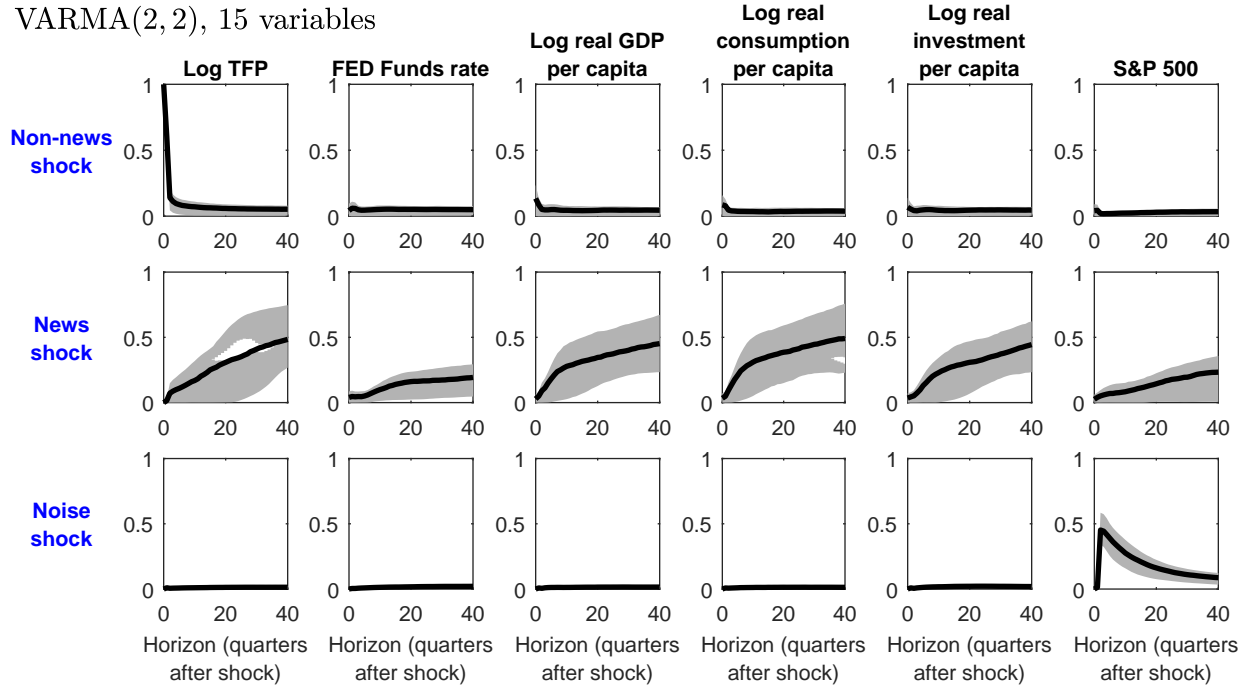


Figure 10: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_\infty$  (all finite roots inside the unit circle). Top panel:  $n = 10$ , bottom panel:  $n = 15$ .

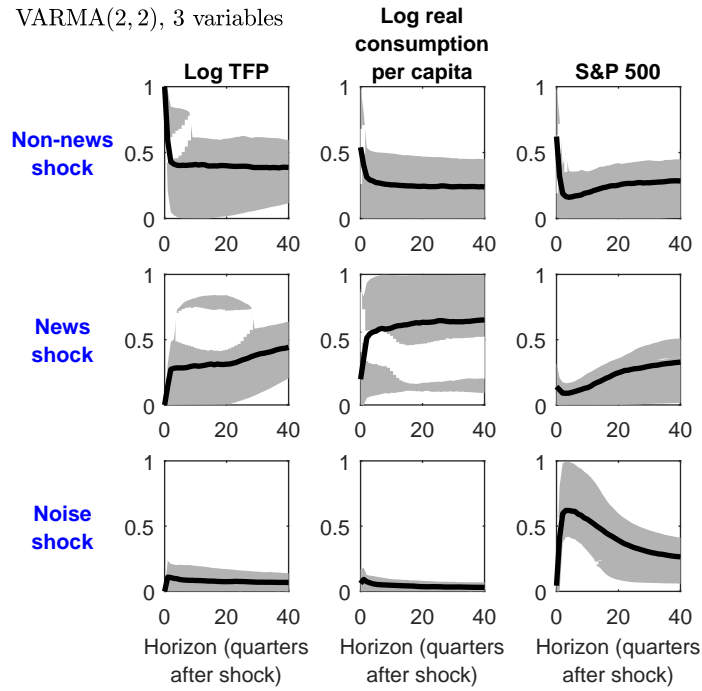
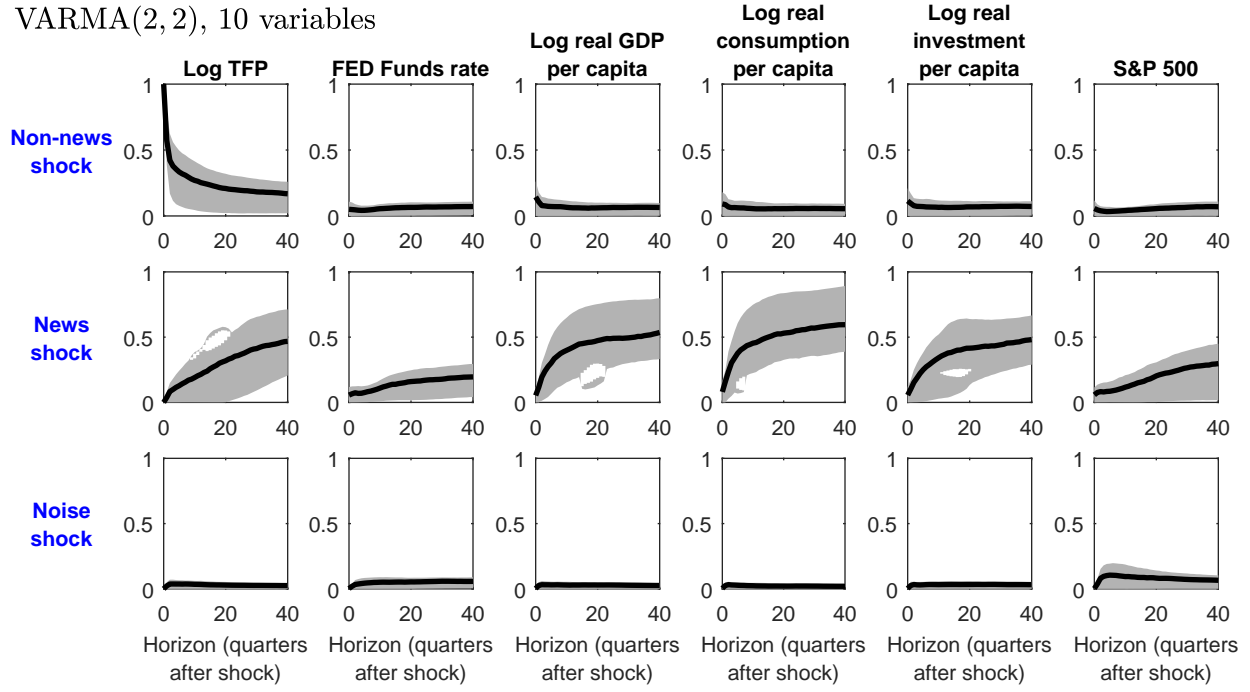


Figure 11: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on set identification over all possible representations.

VARMA(2, 2), 10 variables



VARMA(2, 2), 15 variables

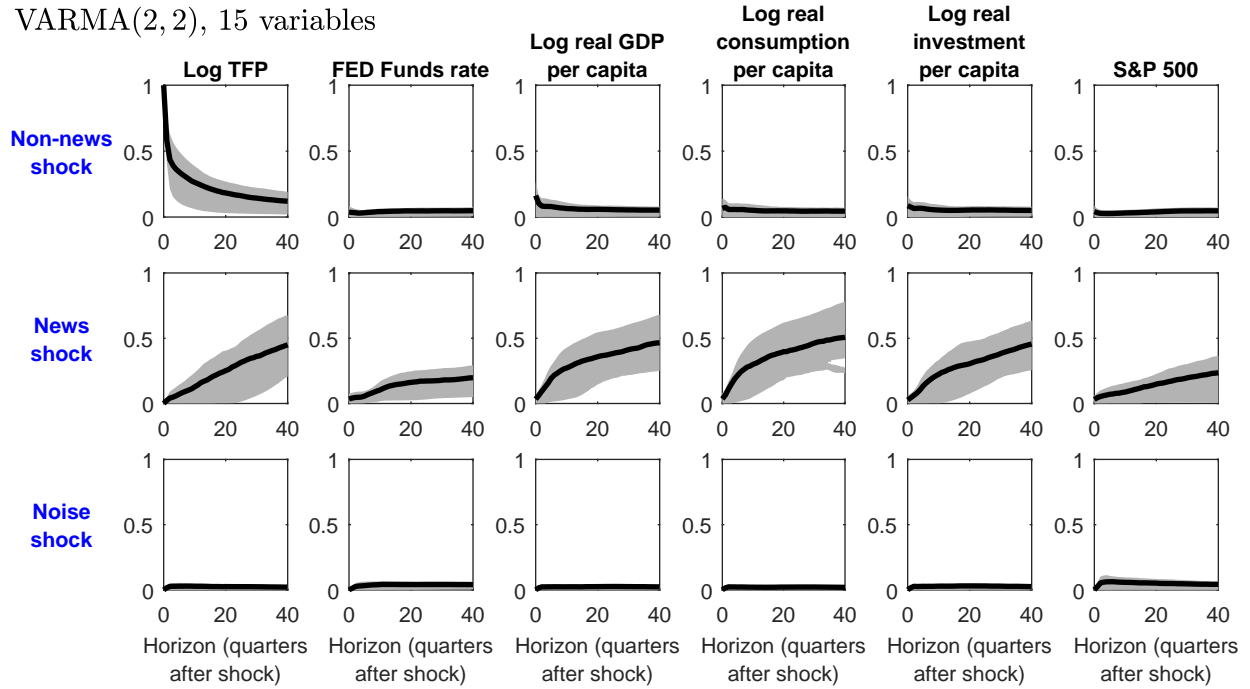


Figure 12: Fractions of forecast error variance explained by non-news, news, and noise shocks (posterior median and 68% highest density region), based on set identification over all possible representations. Top panel:  $n = 10$ , bottom panel:  $n = 15$ .

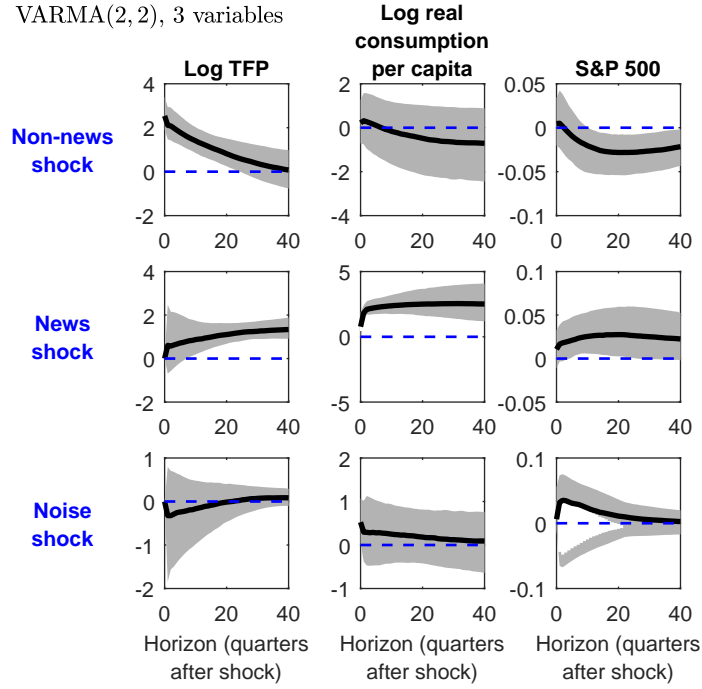
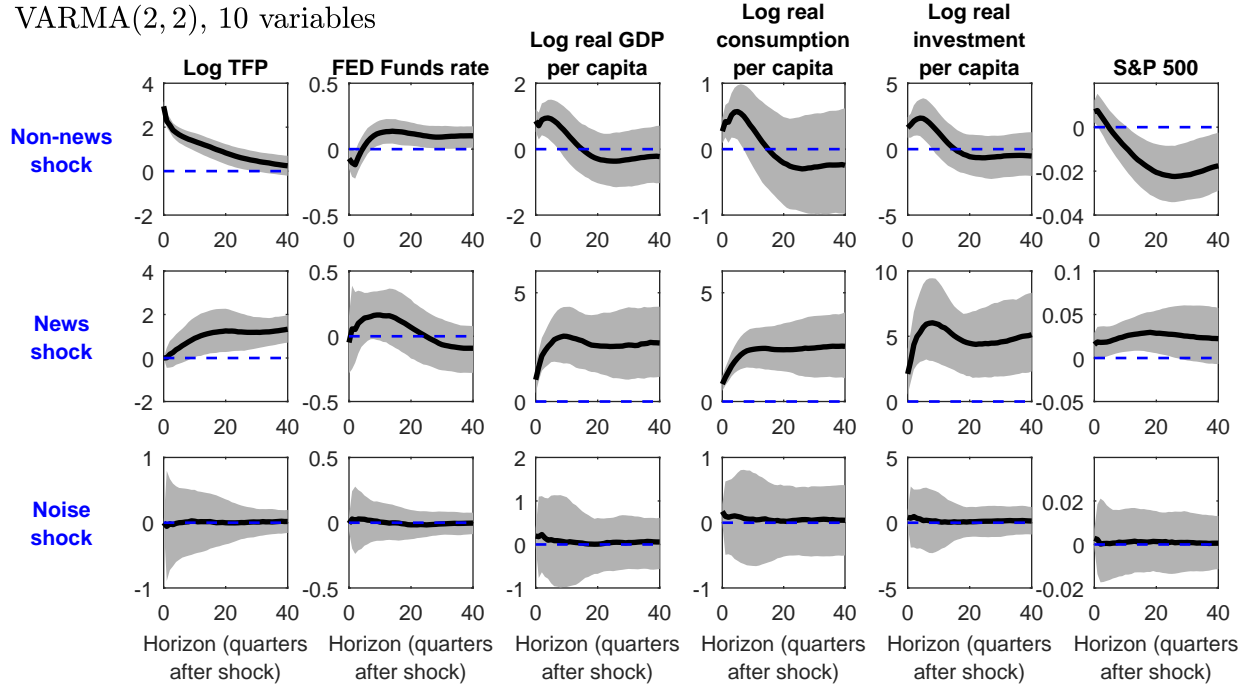


Figure 13: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_2$  (all non-zero, finite roots outside the unit circle).

VARMA(2, 2), 10 variables



VARMA(2, 2), 15 variables

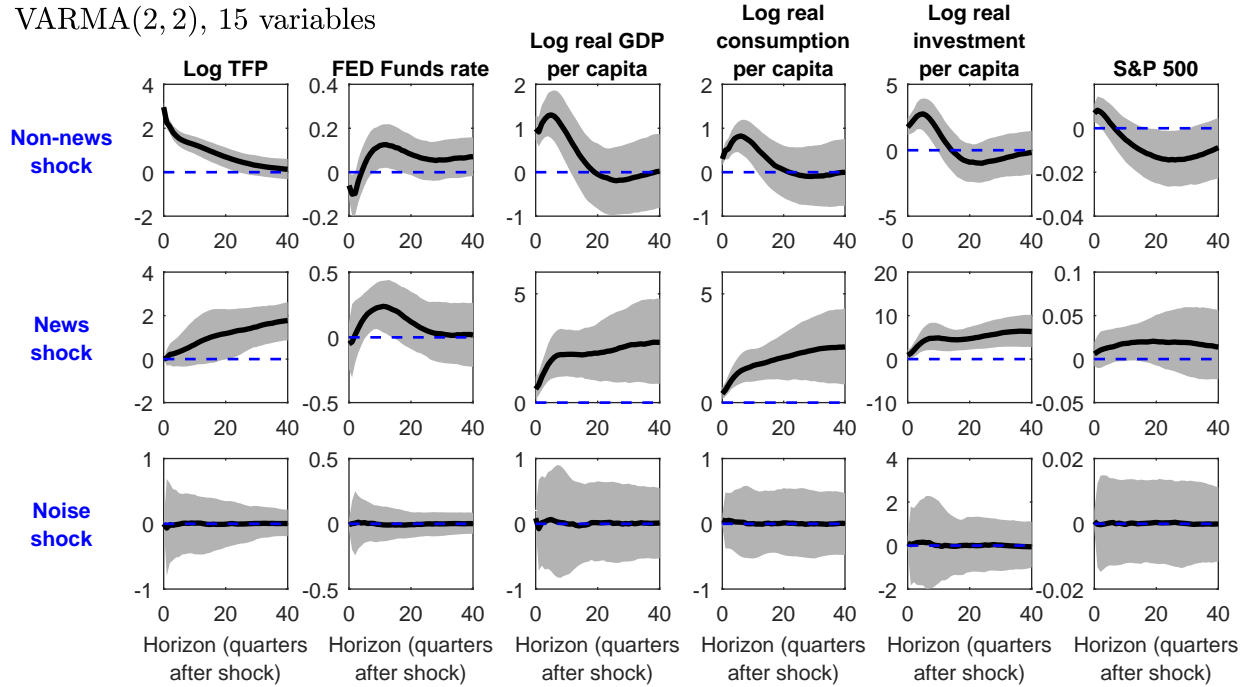


Figure 14: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_2$  (all non-zero, finite roots outside the unit circle). Top panel:  $n = 10$ , bottom panel:  $n = 15$ .

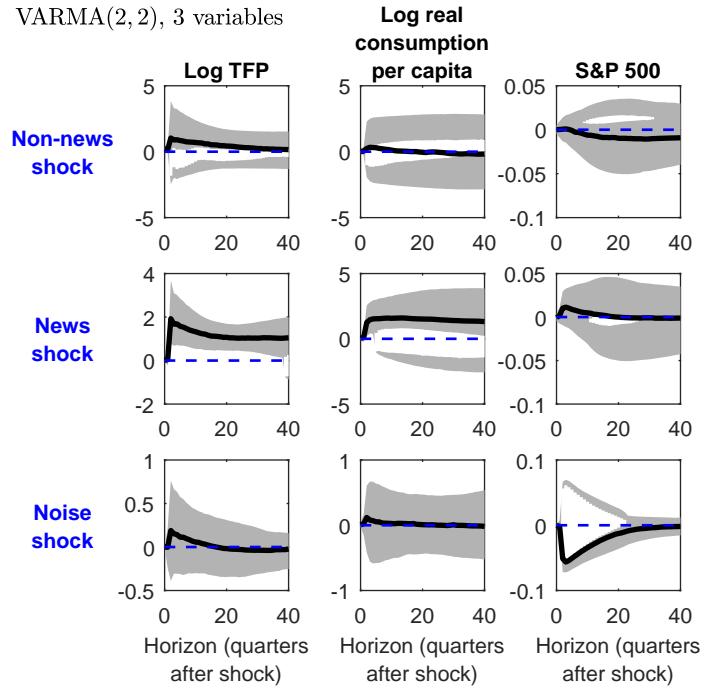
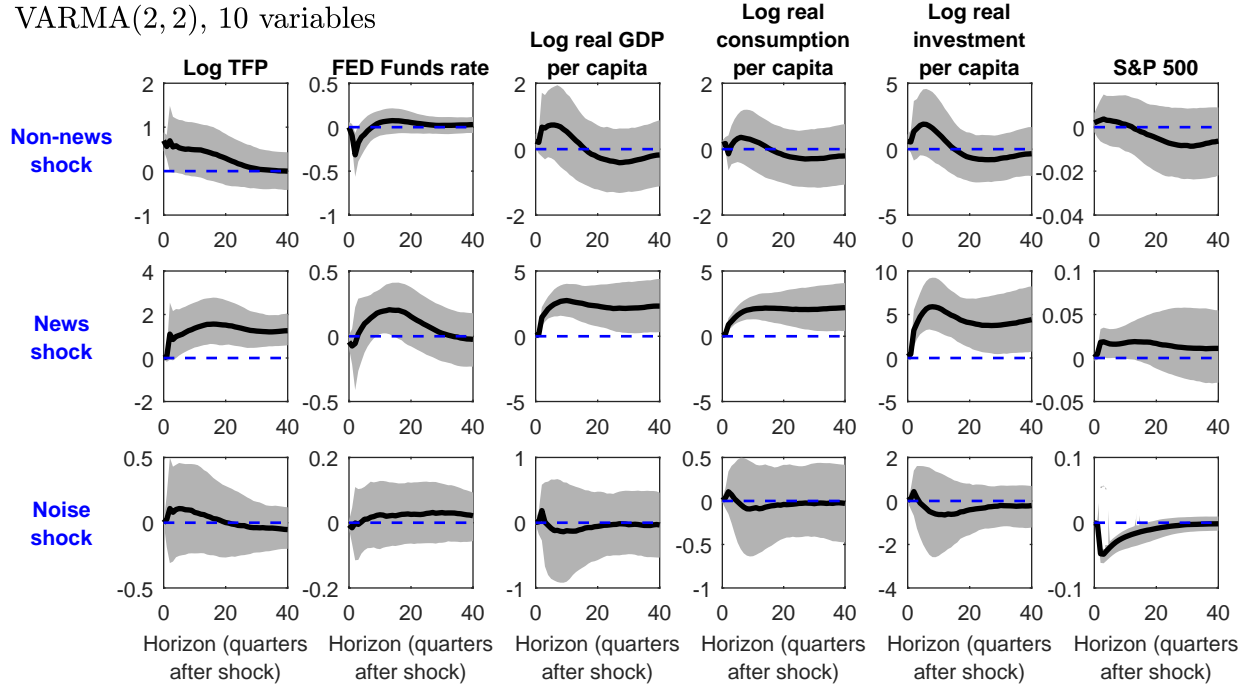


Figure 15: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_\infty$  (all finite roots inside the unit circle).

VARMA(2, 2), 10 variables



VARMA(2, 2), 15 variables

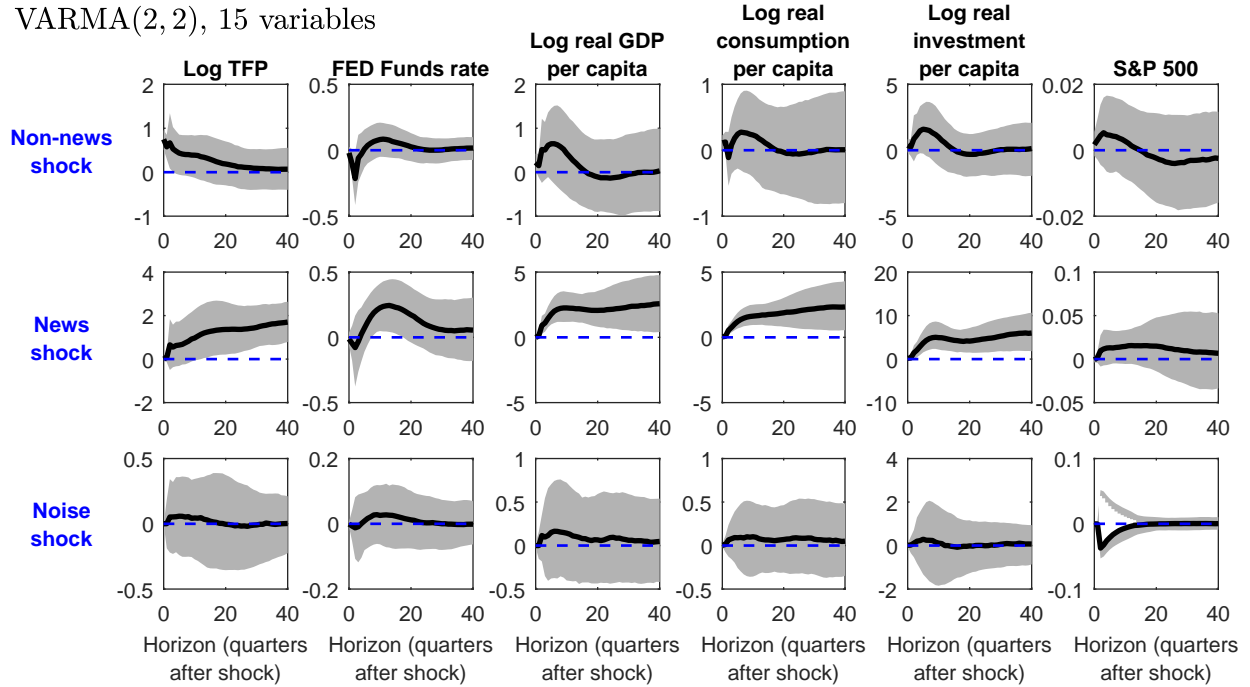


Figure 16: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on unique representation  $\mathcal{R}_\infty$  (all finite roots inside the unit circle). Top panel:  $n = 10$ , bottom panel:  $n = 15$ .

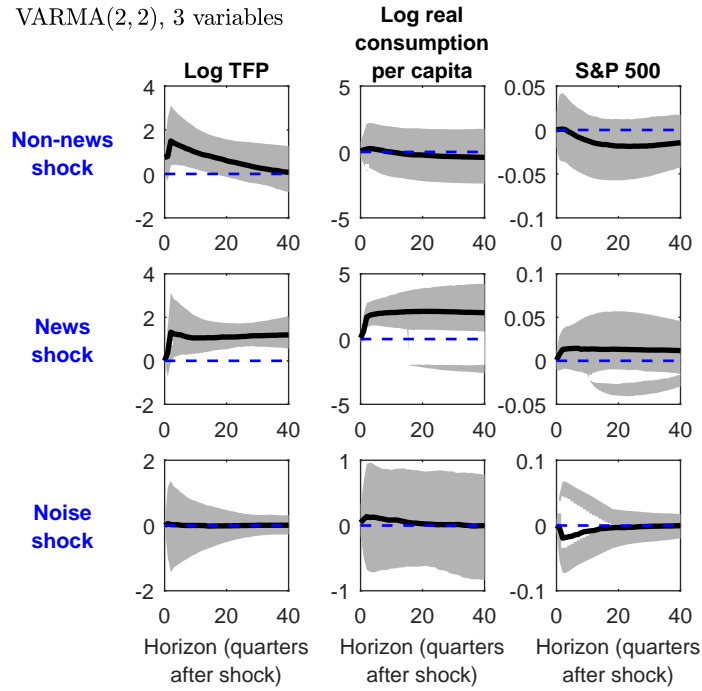
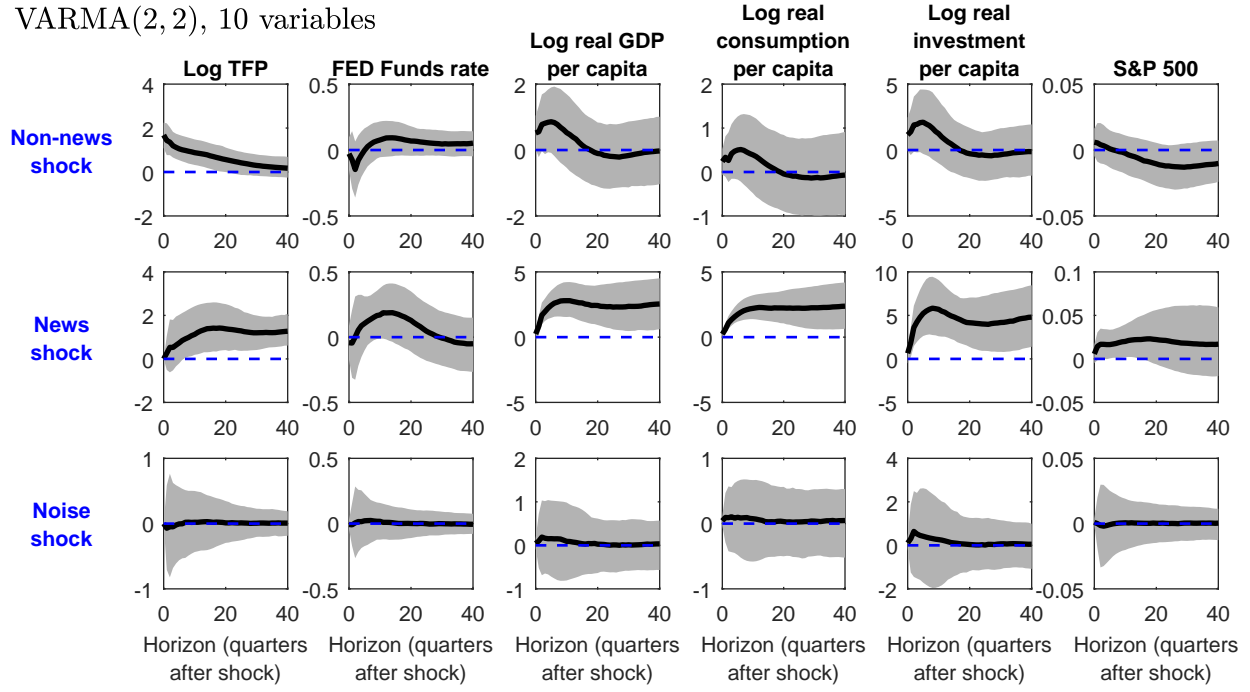


Figure 17: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on set identification over all possible representations.

VARMA(2, 2), 10 variables



VARMA(2, 2), 15 variables

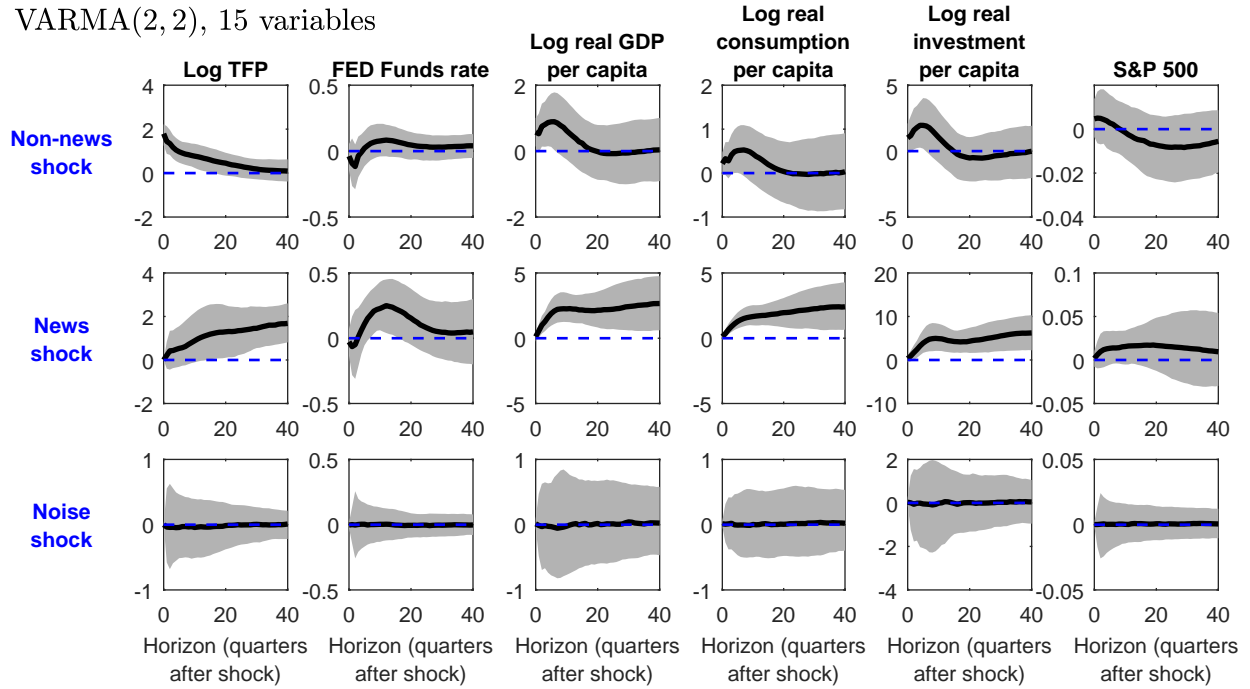


Figure 18: Impulse response functions to non-news, news, and noise shocks (posterior median and 68% highest density region), based on set identification over all possible representations. Top panel:  $n = 10$ , bottom panel:  $n = 15$ .