A Solutions to the Theoretical Models

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

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

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

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

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

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

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

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

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

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

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

A.2 A New Keynesian model

Consider the following standard forward-looking New Keynesian model:

$$R_t = \phi_\pi \pi_{t+1|t} \tag{A.6}$$

$$\pi_t = \beta \pi_{t+1|t} + \kappa y_t \tag{A.7}$$

$$y_t = y_{t+1|t} - \sigma^{-1}[R_t - \pi_{t+1|t} - r_t^N] \tag{A.8}$$

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

$$r_t^N = \tilde{r}_t^N + v_t \tag{A.9}$$

$$\tilde{r}_t^N = \rho_N \tilde{r}_{t-1}^N + \epsilon_t^{NN} + \epsilon_{t-1}^{NE} \tag{A.10}$$

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

Although at time t agents learn about r_t^N, its two individual components, $\tilde{r}_t^N$ and v_t, are not observed. In each period, however, agents receive a signal, which is equal to the sum of the news shock and of a noise component as in (5).

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

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

equations (A.9), (A.10), and (5) can be cast in the state-space form (A.1)-(A.2). As before, the solution to the signal-extraction problem can be obtained by applying the Kalman filter recursions (A.3)-(A.4) to the state-space form (A.1)-(A.2), thus obtaining the solution

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

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

To obtain the model's solution, we first substitute (A.6) into (A.8) to obtain

$$y_t = y_{t+1|t} - \sigma^{-1}[(\phi_\pi - 1)\pi_{t+1|t} - r_t^N] = \tag{A.12}$$

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

From (A.9)-(A.10) we have that $r^N_{t+1|t} = \rho_N \tilde{r}^N_{t|t} + \epsilon^{NE}_{t|t}$, so that the previous equation becomes

$$y_t = -\sigma^{-1}[(\phi_\pi - 1)\pi_{t+1|t} - r^N_t] + y_{t+2|t} - \sigma^{-1}[(\phi_\pi - 1)\pi_{t+2|t} - (\rho_N \tilde{r}^N_{t|t} + \epsilon^{NE}_{t|t})] \quad \text{(A.13)}$$

From (A.7) we get $y_t = \kappa^{-1}[\pi_t - \beta\pi_{t+1|t}]$, and substituting this into the previous expression, we get the following expectational difference equation for inflation:

$$\pi_t - \pi_{t+1|t}[\beta - \kappa\sigma^{-1}(\phi_\pi - 1)] - \pi_{t+2|t}[1 - \kappa\sigma^{-1}(\phi_\pi - 1)] + \beta\pi_{t+3|t} =$$
$$= \kappa\sigma^{-1}[r^N_t + \rho_N \tilde{r}^N_{t|t} + \epsilon^{NE}_{t|t}] \quad \text{(A.14)}$$

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

$$\pi_t = \alpha_1 r^N_t + \alpha_1 \tilde{r}^N_{t|t} + \alpha_1 \epsilon^{NE}_{t|t} \quad \text{(A.15)}$$

the solution turns out to be equal to

$$\pi_t = \kappa\sigma^{-1} r^N_t + \kappa\sigma^{-1}\rho_N \frac{1+\Gamma}{1-\rho_N\Gamma}\tilde{r}^N_{t|t} + \kappa\sigma^{-1}\frac{1+\Gamma}{1-\rho_N\Gamma}\epsilon^{NE}_{t|t} \quad \text{(A.16)}$$

with the analogous solutions for R_t and y_t being

$$R_t = \phi_\pi\rho_N\kappa\sigma^{-1}\left(1 + \rho_N\frac{1+\Gamma}{1-\rho_N\Gamma}\right)\tilde{r}^N_{t|t} + \phi_\pi\kappa\sigma^{-1}\left(1 + \rho_N\frac{1+\Gamma}{1-\rho_N\Gamma}\right)\epsilon^{NE}_{t|t} \quad \text{(A.17)}$$

$$y_t = \sigma^{-1} r^N_t + \rho_N\sigma^{-1}\left[\frac{(1+\Gamma)(1\text{-}\beta\rho_N)}{1-\rho_N\Gamma}\text{-}\beta\right]\tilde{r}^N_{t|t} + \sigma^{-1}\left[\frac{(1+\Gamma)(1\text{-}\beta\rho_N)}{1-\rho_N\Gamma}\text{-}\beta\right]\epsilon^{NE}_{t|t} \quad \text{(A.18)}$$

where

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

This implies that

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

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

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

Just as with the previous model, ϵ^{NE}_t and u_t produce, on impact, the same IRFs for all of the model's endogenous variables, whereas, by assumption, they do not impact upon r^N_t (only the news shock impacts upon r^N_t with a one-period delay).

Figure [??] in the online appendix shows IRFs to news and noise shocks for the interest rate, inflation, and the output gap, conditional on a standard calibration of the model's structural parameters.[21] Consistent with the previous discussion, for each variable the impact at t=0 of news and noise shocks is identical. Further, the IRFs to news shocks lie above the corresponding IRFs to noise shocks, reflecting the fact that, just as in the model with dividends and stock prices, agents progressively learn whether a shock was news or noise. In the long run, the IRFs to news shocks progressively converge to their perfect-information counterpart. This can be seen by comparing the black lines and the blue lines in Figure [??], with the former showing the IRFs to news shocks, and the latter representing instead the same IRFs for the case of no noise shocks (i.e. based on the model calibrated as above, but with $\sigma_u^2 = 0$). An implication of this is that separation between the two sets of IRFs will be faster the smaller is the noise, whereas if the noise is substantial (i.e., σ_u^2 is comparatively large), it will take more time for the agents to learn the truth.

B Econometric Methods

B.1 General Framework

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

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

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

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

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

along with the restriction $1 \leq \text{rank}\tilde{\mathbf{\Theta}}_q \leq n - 1$. Observe that the VARMA process in (B.1) is *fundamental*, and as shown next, the structural VARMA(p, q) of interest

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

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

can be derived as a *basic non-fundamental* representation of this process (see Lippi and Reichlin (1994) for a precise definition and further discussion).

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

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

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

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

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

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

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

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

is observationally equivalent to (B.1).

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

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

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

except in the special case where the second column of $\mathbf{\Theta}_{q-1}$ is either zero or lies in the column space of $\mathbf{\Theta}_q$ (then, $\text{rank}(\mathbf{\Theta}_{q-1}\mathbf{C}_1 + \mathbf{\Theta}_q\mathbf{C}_0) = \text{rank}\mathbf{\Theta}_q$).

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

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

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

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

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

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

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

[23] As discussed in Section 4.3, we analyze all possible non-fundamental representations obtained by "flipping" the roots of the determinant of the fundamental $\tilde{\mathbf{\Theta}}(L)$ (Lippi and Reichlin (1994)).

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

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

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

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

   (b) the news shock is the one which, among all of the remaining shocks, explains the maximum fraction of the FEV of TFP at a long horizon (which we will take to be 20 years ahead).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Finally, let

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

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

B.2 Bayesian Algorithms

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

whereas $\tilde{\mathbf{B}}_j$ in the expanded form is identical to the corresponding $\tilde{\mathbf{B}}_j$ in the semi-structural form for $j = 0, \ldots, p^*$. Consequently, draws from (B.4) can be obtained by sampling directly from the expanded form (B.5) and then computing $\mathbf{M}_1, \ldots, \mathbf{M}_{p^*}$, $\mathbf{\Sigma}$ from each draw of $\mathbf{\Phi}_0, \ldots, \mathbf{\Phi}_{p^*}$, $\mathbf{\Omega}$, and $\mathbf{\Lambda}$ using the mapping in (B.6). The exact procedure based on *generalized eigenvalues* is provided in Section 2 of Chan and Eisenstat (2015) and Appendix D of Chan, Eisenstat, and Koop (2016). To economize on space, we do not reproduce it here, but only emphasize that it is a computationally simple procedure, even for large VARMA systems.

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

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

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

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

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

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

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

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

To complete the prior specification, we set

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

where $\mathcal{IG}(a,b)$ denotes the *inverse gamma distribution* with shape parameter a and rate parameter b. Note that these settings imply weakly informative priors on Ω_{ii} and improper priors on Λ_{ii}. In the paper, we report results holding fixed all of the above prior settings, but varying the dimension of the system n as well as the lag-lengths p and q.

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

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

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

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

1. Sample $\left(\boldsymbol{\gamma}_i, \mathbf{B}_{(i)}, \boldsymbol{\Phi}_{(i)} \,|\, \mathbf{f}, \Lambda_{ii}, \mathbf{y}_i\right)$ for each $i = 1, \ldots, n$, where $\mathbf{B}_{(i)}$ denotes the i-th row of $\mathbf{B} = (\mathbf{I}_n - \mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_p)$, $\boldsymbol{\Phi}_{(i)}$ the i-th row of $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_q)$, and $\boldsymbol{\gamma}_i$ is the set of all SSVS indicators pertaining to $\mathbf{B}_{(i)}, \boldsymbol{\Phi}_{(i)}$.

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

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

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

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

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

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

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

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

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

5. Fore each draw of $\mathbf{B}_1, \ldots, \mathbf{B}_p$, $\mathbf{A}_0, \ldots, \mathbf{A}_q$ compute $\mathbf{K}(L) = \mathbf{B}(L)^{-1}\mathbf{A}(L)$ to obtain draws from the posterior distribution of the impulse responses (and with a further trivial transformation, from the forecast error variance decompositions as well).

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

C.1 The first-order conditions

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

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

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

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

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

where λ_t and μ_t are two Lagrange multipliers.

C.2 The process for TFP

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

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

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

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

where, once again, ϵ_t^{NN} and ϵ_t^{NE} are a non-news and a news shock, respectively. We consider a 1-period anticipation horizon for the news shock. Although at time t agents observe a_t, its two individual components, $\tilde{a}_t$ and v_t, are never observed. In each period, however, agents receive a signal, which is equal to the sum of the news shock and of a noise component as in (5)—that is: $s_t = \epsilon_t^{NE} + u_t$—with u_t being once again $WN(0, \sigma_u^2)$.

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

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

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

and observation equation

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

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

C.4 Stationarizing the model's variables

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C.6 Model solution

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

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

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

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

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

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

where P solves the quadratic equation[28]

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

Q and V are given by

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

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

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

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

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

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

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

Let

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

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

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

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

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

as follows.

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

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

and B_1 is therefore given by

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

E.1 Details of the Monte Carlo experiment

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

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

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

E.2 Evidence

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

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

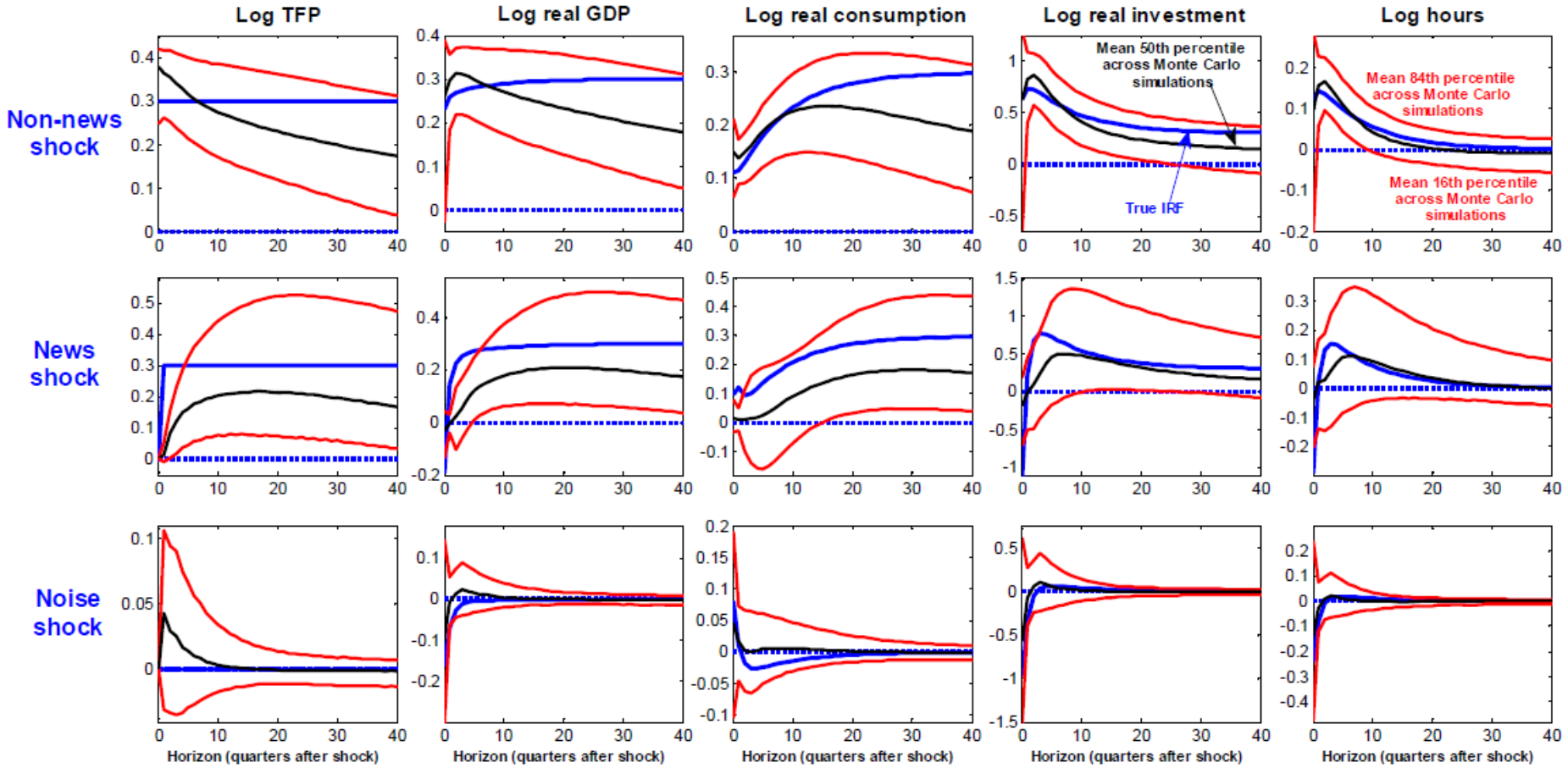

Figure E.1 Results from the Monte Carlo exercise: Means, across all of the Monte Carlo simulations, of the 50th, 16th, and 84th percentiles of the posterior distributions of the impulse-response functions

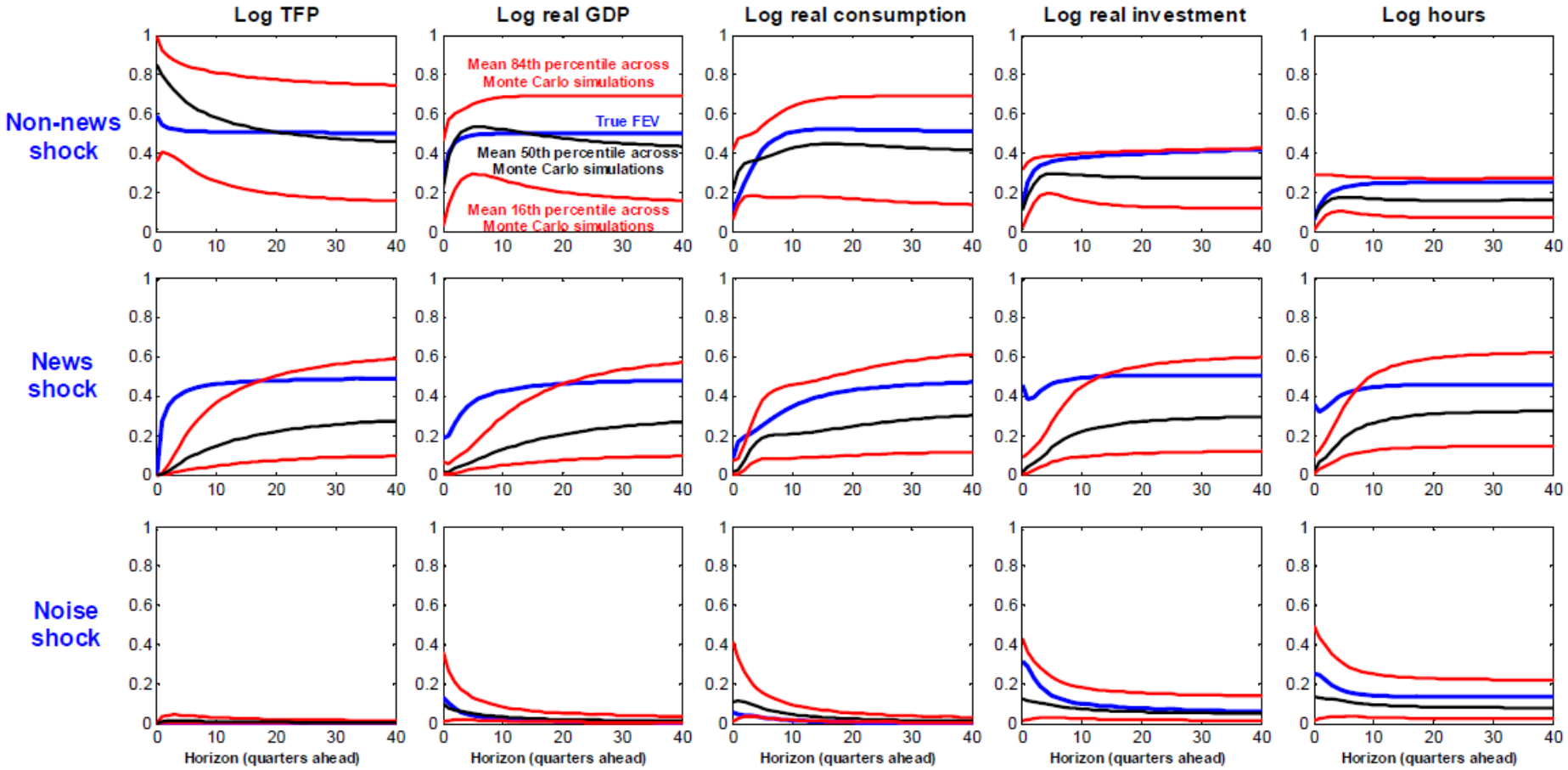

Figure E.2 Results from the Monte Carlo exercise: Means, across all of the Monte Carlo simulations, of the 50^{th}, 16^{th}, and 84^{th} percentiles of the posterior distributions of the fractions of forecast error variance

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

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

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

F The Data

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

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

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

G Model Comparison Exercise

G.1 Deviance Information Criterion

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

G.2 Estimated DIC values for alternative models

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

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

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

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

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

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

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

H Additional Empirical Results

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

I On the Identifiability of News and Noise Shocks

Chahrour and Jurado (2017) have recently argued that news and noise shocks are observationally equivalent, so that they cannot be separately identified. Their argument can be illustrated as follows, based on the example in Section 3.2, pages 13-14. Let

$$x_t = \lambda_{t-1} + \eta_t \tag{I.1}$$

$$s_t = \lambda_t + \xi_t \tag{I.2}$$

where x_t is the observed process, s_t is the signal, and η_t, λ_t, and ξ_t have the interpretation of non-news, news, and noise shocks, with $\eta_t \sim WN(0, \sigma_\eta^2)$, $\lambda_t \sim WN(0, \sigma_\lambda^2)$, and $\xi_t \sim WN(0, \sigma_\xi^2)$. Expression (I.2) implies that

$$\lambda_{t|t} = \underbrace{\frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma_\xi^2}}_{\kappa}[\lambda_t + \xi_t] = \kappa s_t \tag{I.3}$$

The representation for x_t and $x_{t|t-1}$ is given by

$$x_t = x_{t|t-1} + \epsilon_t \tag{I.4}$$

$$x_{t|t-1} = \lambda_{t-1|t-1} = \kappa[\lambda_{t-1} + \xi_{t-1}] \tag{I.5}$$

The process for x_t is uniquely characterized by two moments, $\text{Var}[x_t]$ and $\text{Cov}[x_t, x_{t|t-1}]$, with

$$\text{Var}[x_t] = \sigma_\lambda^2 + \sigma_\eta^2 \tag{I.6}$$

$$\text{Cov}[x_t, x_{t|t-1}] = \frac{\sigma_\lambda^4}{\sigma_\lambda^2 + \sigma_\xi^2} \tag{I.7}$$

Since (I.6)-(I.7) is a system of two equations in three unknowns, the three variances cannot be separately identified.

Suppose, however, that the econometrician can observe another series, y_t, which contains information about the signal s_t. The simplest example would be, e.g.,

$$y_t = s_t + v_t \tag{I.1}$$

where $v_t \sim WN(0, \sigma_v^2)$. For example, in Barsky and Sims' (2011) RBC model, all variables except neutral technology react to s_t, reflecting agents' signal-extraction process about $\tilde{a}_t$. Under these circumstances, the observed vector becomes $Y_t = [x_t \; y_t]'$, with two additional moments, $\text{Var}[y_t]$ and $\text{Cov}[x_t, y_{t-1}]$. The system of moment conditions now features four equations in four unknowns, and it therefore has a unique solution. In general, any model—e.g., Barsky and Sims' (2011) RBC model—features several additional observed variables beyond x_t, and, in general, all of them may depend on ξ_t, as well as on lags of all shocks. This, however, does not change the previous conclusion, as the additional parameters are all identified by the additional moment conditions.

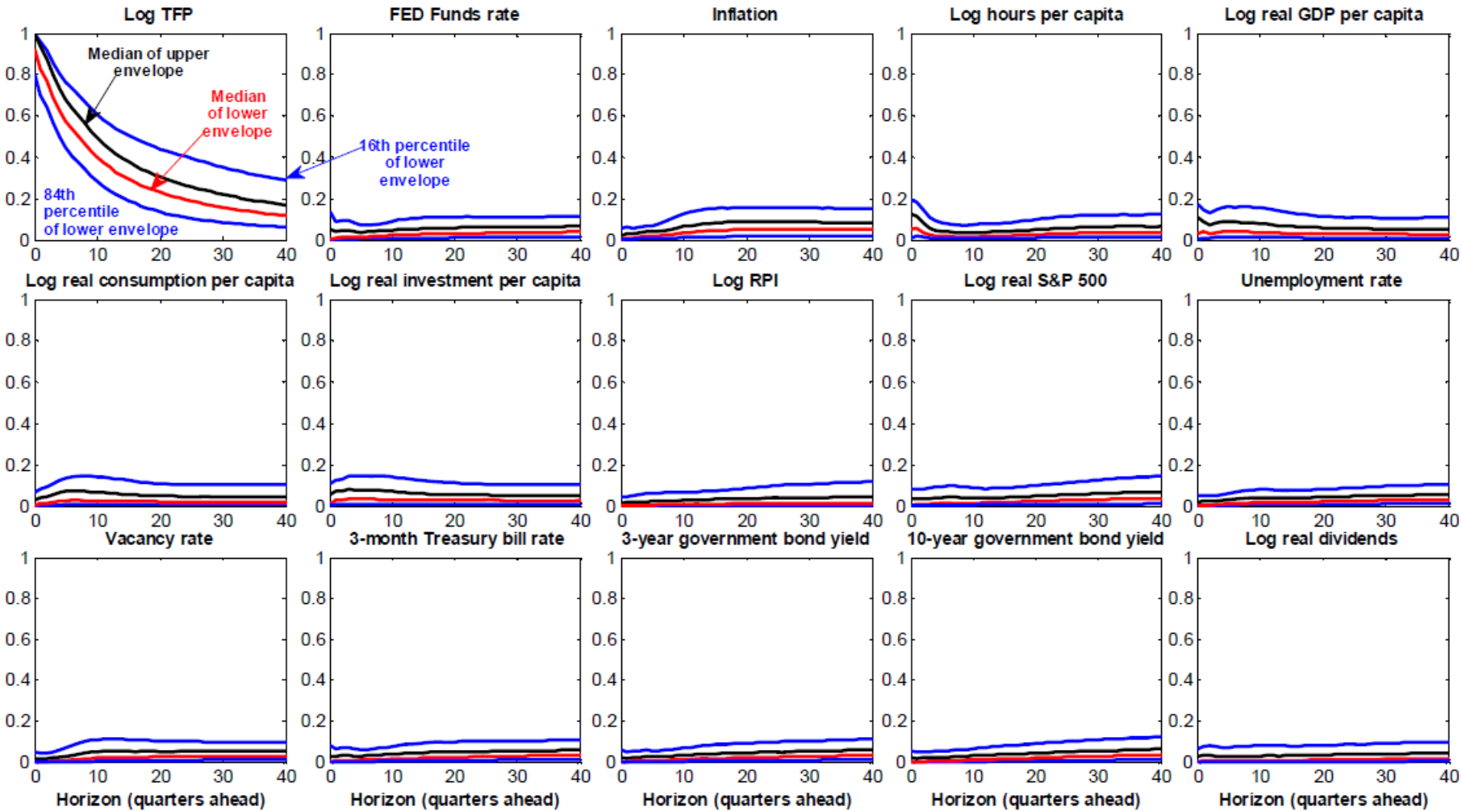

Figure H.1 Fractions of forecast error variance explained by non-news shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

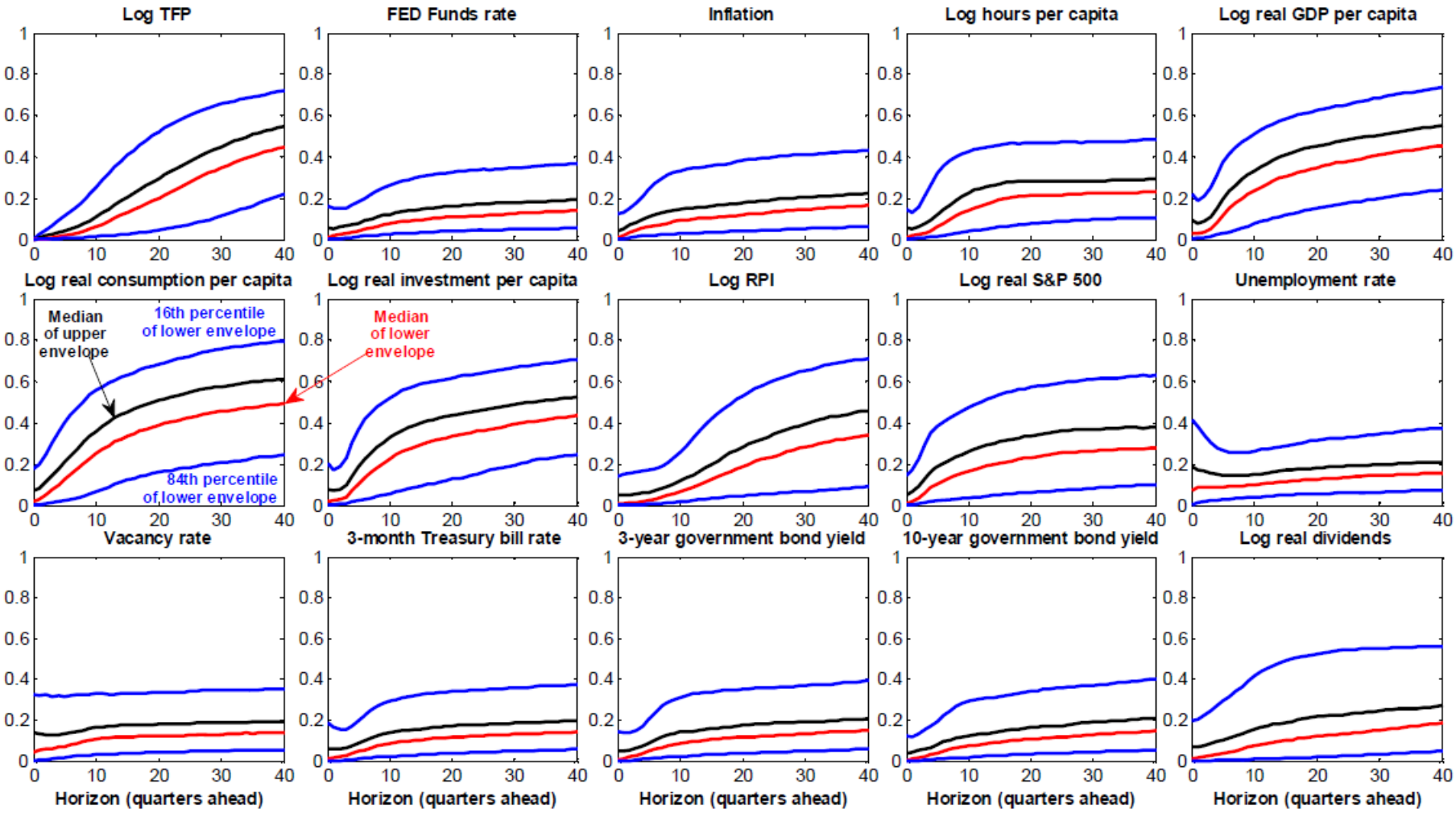

Figure H.2 Fractions of forecast error variance explained by news shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

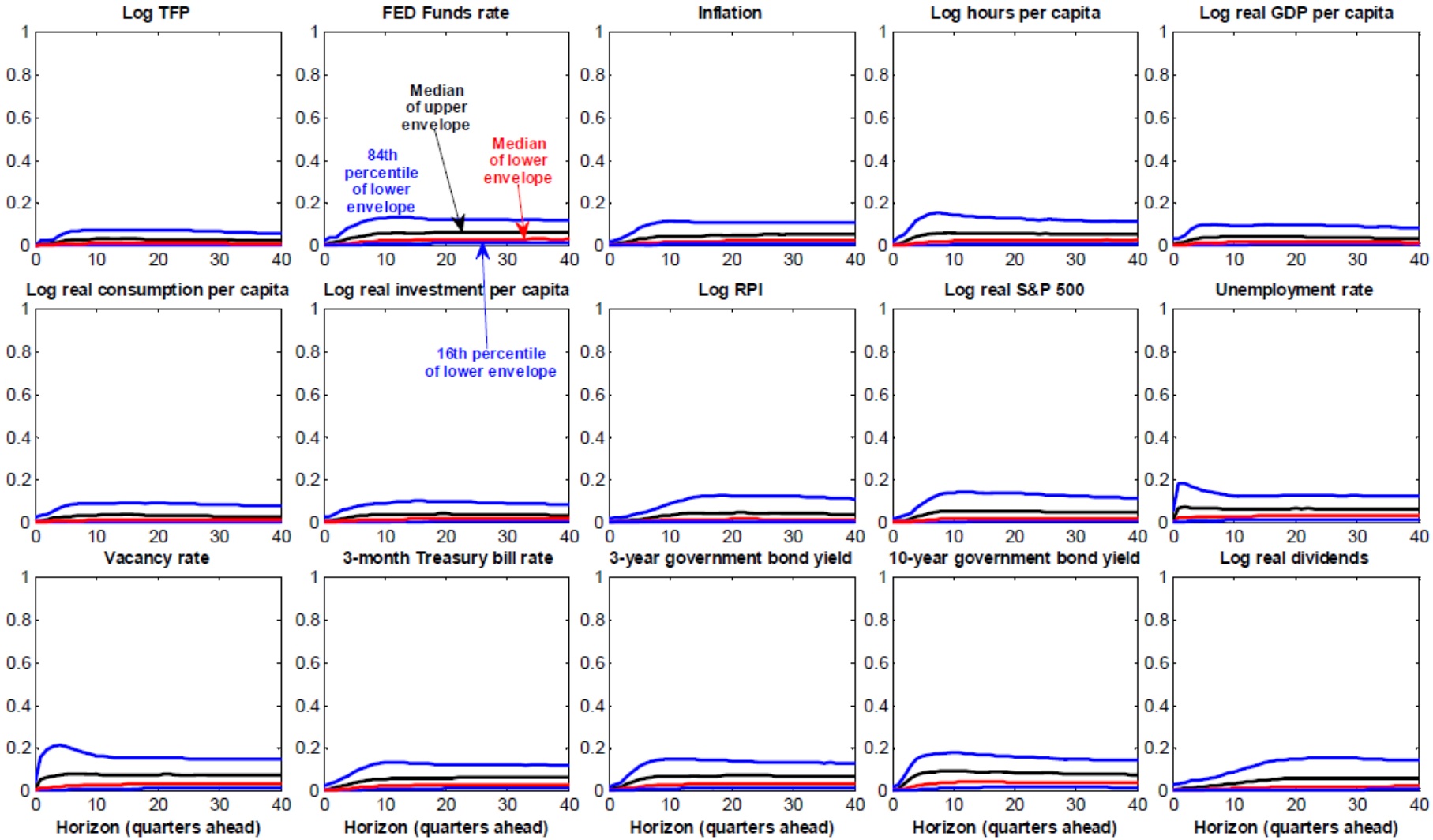

Figure H.3 Fractions of forecast error variance explained by noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

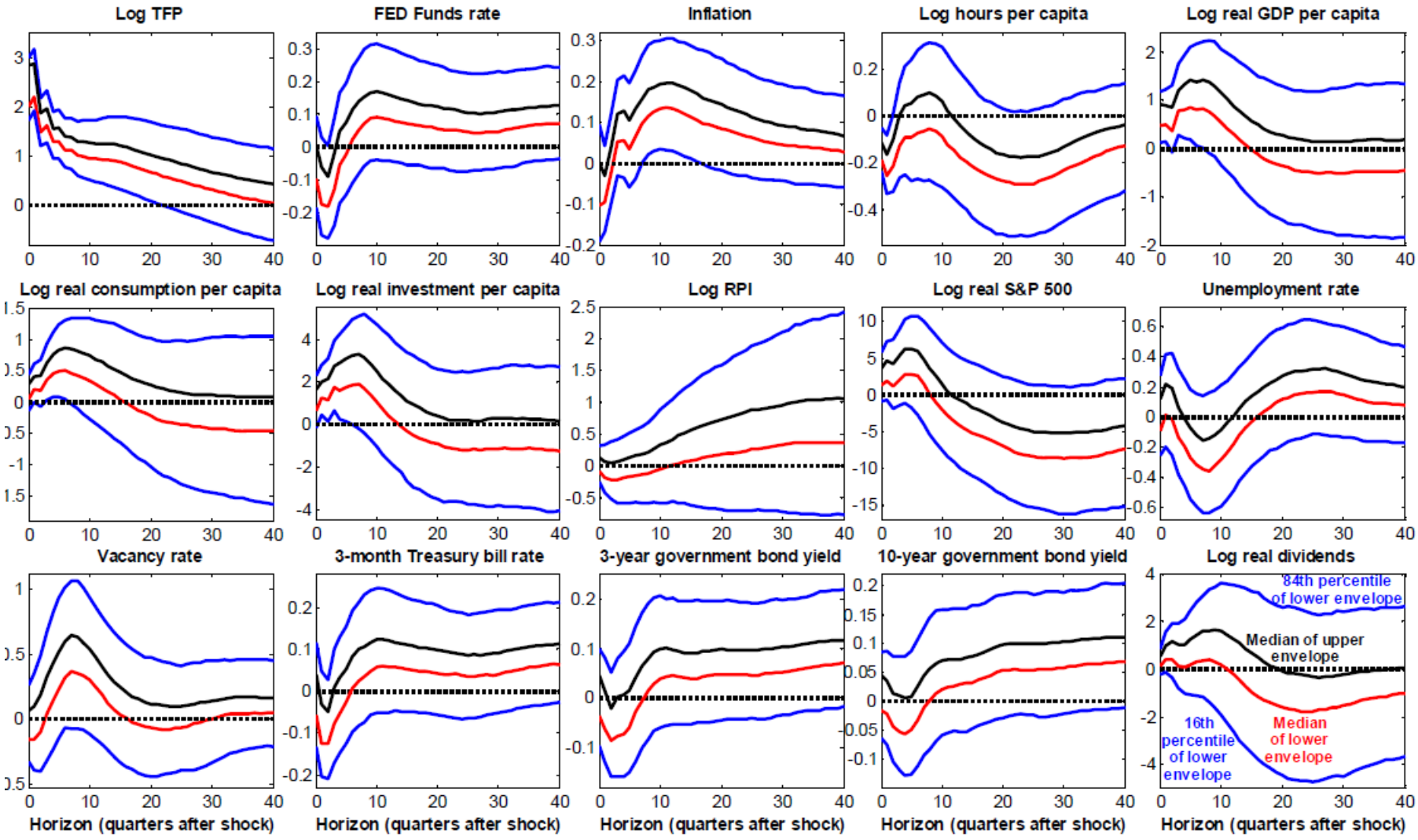

Figure H.4 Impulse-response functions non-news shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

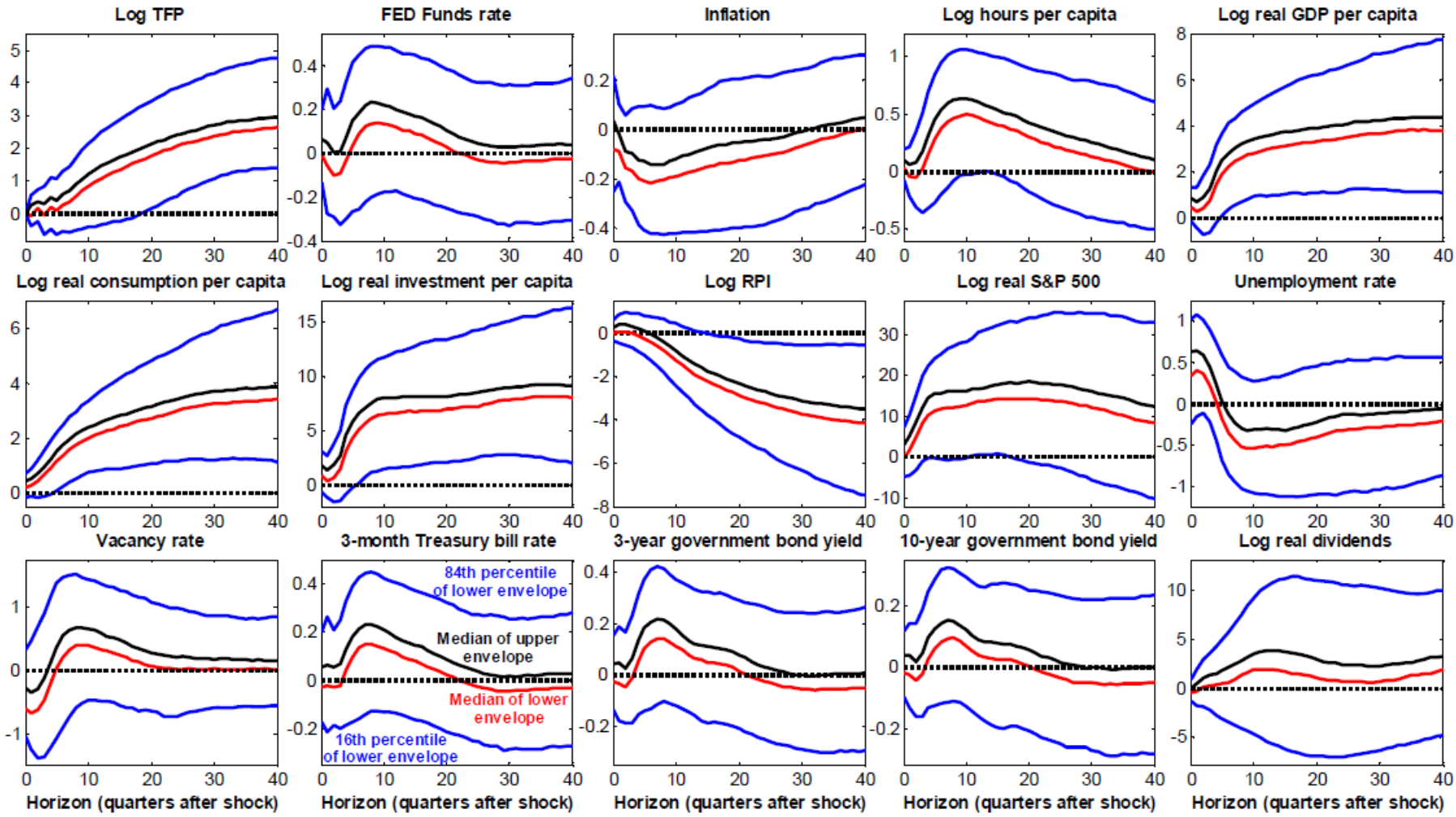

Figure H.5 Impulse-response functions news shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact

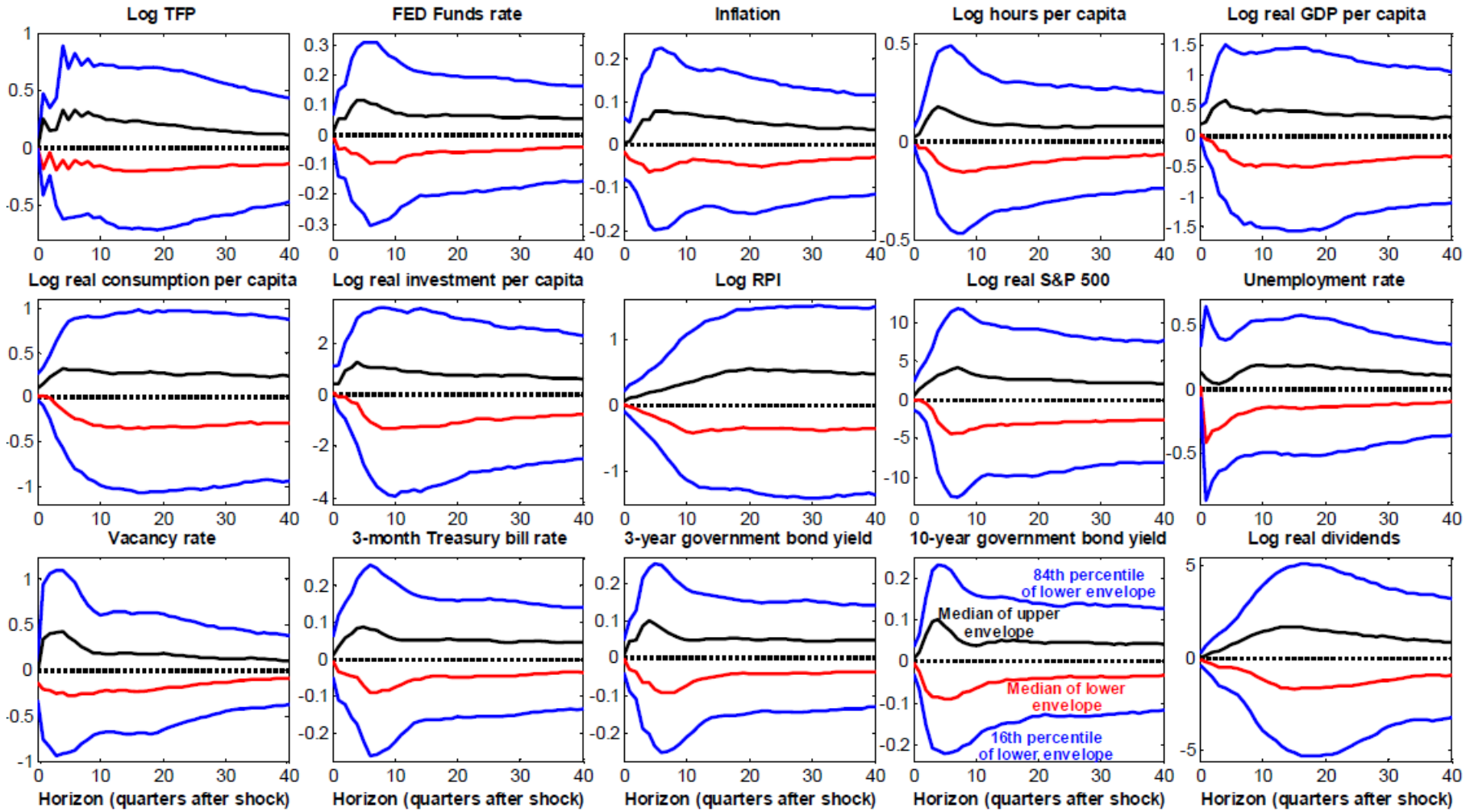

Figure H.6 Impulse-response functions noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) with 15 series, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks 2 quarters after impact