

# Online Appendix to: Identifying Noise Shocks

Luca Benati  
University of Bern\*

Joshua Chan  
Purdue University†

Eric Eisenstat  
University of Queensland‡

Gary Koop  
University of Strathclyde§

## 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

$$F \equiv \begin{bmatrix} 1 & 0 & 1 \\ 0 & \rho_T & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad H \equiv \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}'$$

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

equations (2)-(5) in Section 2.1 of the paper 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  $\tilde{K} \equiv (I_3 - KH')F$ ,

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

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

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

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

where  $I_3$  is the  $3 \times 3$  identity matrix, the solution to the the signal-extraction problem is therefore given by

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

## A.2 A New Keynesian model

Consider the following standard forward-looking New Keynesian model:

$$R_t = \phi_\pi \pi_{t+1|t} \quad (\text{A.6})$$

$$\pi_t = \beta \pi_{t+1|t} + \kappa y_t \quad (\text{A.7})$$

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

where  $R_t$ ,  $\pi_t$ , 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:

$$r_t^N = \tilde{r}_t^N + v_t \quad (\text{A.9})$$

$$\tilde{r}_t^N = \rho_N \tilde{r}_{t-1}^N + \epsilon_t^{NN} + \epsilon_t^{NE} \quad (\text{A.10})$$

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  $\epsilon_t^{NN}$ ,  $\epsilon_t^{NE}$ , 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 equation (5) in Section 2.1 of the paper.

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 equation (5) in Section 2.1 of the paper 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

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

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  $2 \times 2$  identity matrix.

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

$$\begin{aligned} y_t &= y_{t+1|t} - \sigma^{-1} [(\phi_\pi - 1)\pi_{t+1|t} - r_t^N] = \\ &= -\sigma^{-1} [(\phi_\pi - 1)\pi_{t+1|t} - r_t^N] + y_{t+2|t} - \sigma^{-1} [(\phi_\pi - 1)\pi_{t+2|t} - r_{t+1|t}^N] \end{aligned} \quad (\text{A.12})$$

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

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

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:

$$\begin{aligned} \pi_t - \pi_{t+1|t}[\beta - \kappa\sigma^{-1}(\phi_\pi - 1)] - \pi_{t+2|t}[1 - \kappa\sigma^{-1}(\phi_\pi - 1)] + \beta\pi_{t+3|t} &= \\ &= \kappa\sigma^{-1}[r_t^N + \rho_N \tilde{r}_{t|t}^N + \epsilon_{t|t}^{NE}] \end{aligned} \quad (\text{A.14})$$

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

$$\pi_t = \alpha_1 r_t^N + \alpha_1 \tilde{r}_{t|t}^N + \alpha_1 \epsilon_{t|t}^{NE} \quad (\text{A.15})$$

the solution turns out to be equal to

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

with the analogous solutions for  $R_t$  and  $y_t$  being

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

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

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

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

Just as with the previous model,  $\epsilon_t^{NE}$  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  $r_t^N$  (only the news shock impacts upon  $r_t^N$  with a one-period delay).

Figure I.1 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.<sup>1</sup> 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 I.1, 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.,  $\sigma_u^2$  is comparatively large), it will take more time for the agents to learn the truth.

<sup>1</sup>Specifically, we set  $\beta=0.99$ ,  $\kappa=0.05$ ,  $\sigma=1$ ,  $\phi_\pi=1.5$ , and  $\rho_N=0.95$ .

## B Econometric Methods

The econometric methodology employed in this paper is outlined in sub-section 5.2 of the main text. A key feature of this methodology is transforming an estimated fundamental reduced-form VARMA to the non-fundamental structural representation, which is used to compute impulse response functions and forecast error variance decompositions. These are summarized by Steps 3 and 4 in the procedure outlined at the end of sub-section 5.2. The details of these two steps are as follows.

### B.1 Transforming to a Non-fundamental Representation

Here, we provide details of Step 3, which transforms the *fundamental* representation that is estimated directly, with one characteristic root at infinity by construction, to a *non-fundamental* representation, with one root at zero. Accordingly, the approach in Step 3 involves traversing the following sub-steps:

- 3.1. Compute an arbitrary decomposition (e.g. Cholesky)  $\Theta_0\Theta'_0 = \Sigma$ .
- 3.2. Let  $\delta$  be the  $n \times 1$  vector such that  $\Theta_q\Theta_0\delta = 0$  and  $\delta'\delta = 1$ . Note that  $\delta$  is unique (up to sign). Set  $\Gamma_0 = (\delta, \Delta_\perp)$ , such that  $\delta'\Delta_\perp = 0$  and  $\Gamma'_0\Gamma_0 = \Gamma\Gamma'_0 = \mathbf{I}_n$  (i.e. a constant orthogonal matrix).
- 3.3. Compute  $\tilde{\Theta}_q = \Theta_q\Gamma_0$ , which results in the first column of  $\tilde{\Theta}_q$  being zero. 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

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

Note that because  $\delta$  in 3.2 is unique up to sign, the resulting polynomial matrix  $\tilde{A}(L) = \tilde{A}_0 + \tilde{A}_1L + \dots + \tilde{A}_qL^q$  is such that  $A(L) = \tilde{A}(L)\Gamma$ , where  $\Gamma$  is a constant orthogonal matrix. It therefore only remains to construct the appropriate orthogonal rotations matrix based on the identifying restrictions proposed in sub-section 3.3 of the main text.

### B.2 Imposing Structural Identification Restrictions

In the final post-processing Step 4 of the procedure outlined in sub-section 5.2 of the main text, we obtain the structural  $A_0, \dots, A_q$  that satisfy the identifying restrictions by applying a series of (constant) orthogonal rotations to  $\tilde{A}_0, \dots, \tilde{A}_q$  (obtained after executing 3.1-3.3 in the preceding sub-section) as in typical VAR settings. To implement our identification scheme R1-R3 (outlined in sub-section 3.3 of the main text), we assume  $n \geq 4$ ,  $\epsilon_{1,t}$  is non-news,  $\epsilon_{2,t}$  is news,  $\epsilon_{3,t}$  is noise, and  $\epsilon_{4,t}$  is the transitory TFP shock. We then obtain the requisite orthogonal rotations matrices as follows.

Ideally, the orthonormal matrices should be constructed using the algorithm described in sub-section 4.2, which involves numerical optimisation. While we use such an approach in executing the population exercise of Section 4, we find the numerical optimisation to be computationally too demanding to be practical in our Monte Carlo study and empirical applications involving substantially more variables. Instead, we employ an approximate algorithm that involves five types of analytically tractable orthogonal rotations,  $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$ , and  $\Gamma_5$ , such that their product  $\Gamma = \Gamma_1\Gamma_2\Gamma_3\Gamma_4\Gamma_5$  yields the comprehensive set of orthogonal rotations that transform  $\tilde{A}(L)$  into the structural representation of interest  $A(L)$ , where the two VMA polynomials are related by  $A_j = \tilde{A}_j\Gamma$  for  $j = 0, \dots, q$ . This approximate algorithm is computationally efficient and as detailed in Appendix E performs extremely well in recovering true IRFs/FEVs; its step-by-step implementation proceeds as follows.

We start with  $\Gamma_1$ , which is determined by setting the first column

$$\Gamma_{1,1} = \tilde{A}'_{0,(1)} / \|\tilde{A}_{0,(1)}\|^2,$$

where  $\tilde{A}_{0,(1)}$  denotes the first row of  $\tilde{A}_0$ , and the remaining columns  $\Gamma_{1,i}$  for  $i = 2, \dots, n$  equal to the  $n - 1$  vectors that are orthogonal to  $\tilde{A}'_{0,(1)}$  (normalized such that  $\|\Gamma_{1,i}\| = 1$ ).

Next, let  $\tilde{K}(L) = B(L)^{-1}\tilde{A}(L)\Gamma_1$  be the impulse responses obtained after applying the first set of orthogonal rotations  $\Gamma_1$ , and define  $\tilde{K}_{j,1[2:n]}$  for  $j \geq 0$  as the  $1 \times (n - 1)$  row vector constructed from the first row and columns 2 to  $n$  of  $\tilde{K}_j$ . Compute the eigenvalue decomposition of  $\sum_{j=1}^{20} \tilde{K}'_{j,1[2:n]}\tilde{K}_{j,1[2:n]}$ , with eigenvalues sorted in *descending* order, and store the eigenvectors in  $E_2$ . Using this, construct the orthogonal matrix

$$\Gamma_2 = \begin{pmatrix} 1 & 0 \\ 0 & E_2 \end{pmatrix}.$$

Observe that rotations  $\Gamma_1$  and  $\Gamma_2$  result in  $\check{A}(L) = \tilde{A}(L)\Gamma_1\Gamma_2$  that is *unique up to signs*, and at the same time, a news shock that satisfies restriction R2. 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.

We then proceed to identify the noise shock that satisfies restriction R3 by constructing an orthogonal matrix  $\Gamma_3$  such that the third column of  $\check{A}_0\Gamma_3$  is proportional to the second. Specifically, let

$$\Gamma_3 = \begin{pmatrix} \mathbf{I}_2 & 0 \\ 0 & E_3 \end{pmatrix},$$

where the first column  $E_{3,1}$  of the  $n - 2 \times n - 2$  orthogonal matrix  $E_3$  must satisfy  $(\check{A}_{0,3}, \dots, \check{A}_{0,n})E_{3,1} = c\check{A}_{0,2}$ .

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

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

(i.e.  $\zeta_A$  is the orthonormal basis for the null space of  $(\check{A}_{0,2}, \dots, \check{A}_{0,n})$ , with  $\zeta_{A,1}$  a scalar and  $\zeta_{A,2}$  a  $(n - 2) \times 1$  vector). Accordingly, set

$$c = \frac{|\zeta_{A,1}|}{1 - \zeta_{A,1}^2} \quad (1)$$

$$E_{3,1} = -\text{sign}(\zeta_{A,1}) \frac{\zeta_{A,2}}{1 - \zeta_{A,1}^2}, \quad (2)$$

and the remaining columns  $E_{3,2}, \dots, E_{3,n-2}$  of  $E_3$  to be the  $n - 3$  vectors orthogonal to  $E_{3,1}$  (normalized such that  $\|E_{3,i}\| = 1$ ).

Finally, construct  $\Gamma_4$  and  $\Gamma_5$  to identify the non-news and transitory TFP shocks that satisfies restriction R1. To this end, let  $\check{K}(L) = B(L)^{-1}\check{A}(L)\Gamma_3$  be the impulse responses obtained after applying the three orthogonal rotations described above, and define  $\check{K}_{j,1[1,4:n]}$  for  $j \geq 0$  as the  $1 \times (n - 2)$  row vector constructed from the first row and columns 1, 4,  $\dots, n$  of  $\check{K}_j$ . Compute the eigenvalue decomposition of  $\sum_{j=1}^{20} \check{K}'_{j,1[1,4:n]}\check{K}_{j,1[1,4:n]}$ , with eigenvalues sorted in *descending* order, and store the eigenvectors in the  $(n - 2) \times (n - 2)$  matrix  $E_4$ .

Now, set

$$\Gamma_4 = \begin{pmatrix} E_{4,11} & 0 & E_{4,1[4:n-2]} \\ 0 & \mathbf{I}_2 & 0 \\ E_{4,[4:n-2]1} & 0 & E_{4,[4:n-2][4:n-2]} \end{pmatrix},$$

where  $E_{4,11}$  is the (1, 1) element of  $E_4$ ,  $E_{4,[4:n-2]1}$  is the  $(n - 4) \times 1$  vector constructed from the first column and last  $n - 4$  rows of  $E_4$ ,  $E_{4,1[4:n-2]}$  is the  $1 \times (n - 4)$  vector constructed from the first row

<sup>2</sup>Recall that rank  $\tilde{A}_0 = n - 1$ , transformation by  $\Gamma_1$  preserves the linear independence of the first column, and transformation by  $\Gamma_2$  only alters columns 2 to  $n$ .

and last  $n - 4$  columns of  $E_4$ , and  $E_{4,[4:n-2][4:n-2]}$  is the  $(n - 4) \times (n - 4)$  sub-matrix constructed from the last  $n - 4$  rows and  $n - 4$  columns of  $E_4$ .

The polynomial matrix  $\hat{A}(L) = \tilde{A}(L)\Gamma_1\Gamma_2\Gamma_3\Gamma_4$  yields a representation in which non-news, news and noise satisfy restrictions R1-R3. Thus, all that remains is to identify the transitory TFP shock as the only other shock that impacts TFP at  $t = 0$  besides non-news, if it is also of interest. To this end, construct the  $(n - 3) \times (n - 3)$  matrix  $E_5$  by setting the first column  $E_{5,1} = \hat{A}'_{0,1[4:n]} / \|\hat{A}_{0,1[4:n]}\|^2$ , where  $\hat{A}_{0,1[4:n]}$  denotes the  $1 \times (n - 3)$  row vector constructed from the first row and last  $n - 3$  columns of  $\hat{A}_0$ . Set the remaining columns  $E_{5,i}$  for  $i = 2, \dots, n - 3$  equal to the  $n - 4$  vectors that are orthogonal to  $A'_{0,1[4:n]}$  (normalized such that  $\|E_{5,i}\| = 1$ ). Defining

$$\Gamma_5 = \begin{pmatrix} \mathbf{I}_3 & 0 \\ 0 & E_5 \end{pmatrix},$$

we obtain the desired structural representation as  $A(L) = \hat{A}(L)\Gamma_5$ . Once again, we normalize the signs of the non-news shock and transitory TFP shocks by requiring that the maximum impulse response (over the horizon  $0 : 20$ ) of TFP to both shocks is positive.

## C Barsky and Sims' (2011) RBC Model Augmented with Noise Shocks about 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

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

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

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

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

where  $\lambda_t$  and  $\mu_t$  are two Lagrange multipliers.

### C.2 The process for TFP

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

$$a_t = \tilde{a}_t + v_t \quad (\text{C.5})$$

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

$$\tilde{a}_t = \tilde{a}_{t-1} + \epsilon_t^{NN} + \epsilon_t^{NE} \quad (\text{C.6})$$

where, once again,  $\epsilon_t^{NN}$  and  $\epsilon_t^{NE}$  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 equation (5) in Section 2.1 of the paper—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 described by equation (5) in Section 2.1 of the paper, (C.5), and (C.6) can be put into state-space form, with state equation

$$\begin{bmatrix} \Delta\tilde{a}_t \\ \epsilon_t^{NE} \\ v_t \\ v_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_A \begin{bmatrix} \Delta\tilde{a}_{t-1} \\ \epsilon_{t-1}^{NE} \\ v_{t-1} \\ v_{t-2} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_B \begin{bmatrix} \epsilon_t^{NN} \\ \epsilon_t^{NE} \\ v_t \\ u_t \end{bmatrix} \quad (\text{C.7})$$

and observation equation

$$\begin{bmatrix} \Delta a_t \\ s_t \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_C \begin{bmatrix} \Delta\tilde{a}_t \\ \epsilon_t^{NE} \\ v_t \\ v_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_D \begin{bmatrix} \epsilon_t^{NN} \\ \epsilon_t^{NE} \\ v_t \\ u_t \end{bmatrix} \quad (\text{C.8})$$

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).<sup>3</sup> 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  $K_t^* \equiv K_t/\Gamma_{t-1}$ ;<sup>4</sup> and we stationarize the two Lagrange multipliers,  $\lambda_t$  and  $\mu_t$ , 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  $\lambda_t^*$ ,  $\hat{y}_t^*$  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:

$$\begin{aligned} & \hat{\mu}_t^* + \hat{c}_t^*[S_C\Delta_G + (1 - S_C)(1 - \Delta_G)] + (1 - S_C)\Delta_G\hat{c}_{t+1|t}^* + \\ & + S_C(1 - \Delta_G)\hat{c}_{t-1}^* + S_C\frac{1 - \Delta_G}{1 - \theta}\Delta a_t - \frac{(1 - S_C)\Delta_G}{1 - \theta}\Delta a_{t+1|t} - \epsilon_t^c = 0 \end{aligned} \quad (\text{C.9})$$

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

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

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

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

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

<sup>3</sup>We wish to thank Eric Sims for providing extensive details about the solution to their original model.

<sup>4</sup>The capital stock is divided by  $\Gamma_{t-1}$ , rather than by  $\Gamma_t$ , in order to make sure that the stationarized capital stock,  $K_t^*$ , is still predetermined at time  $t$ .

$$\hat{\mu}_t^* - \hat{\lambda}_t^* + \gamma \hat{\nu}_t^* + \frac{\gamma}{1-\theta} \Delta a_t - \gamma \hat{\nu}_{t-1}^* - \epsilon_t^i = 0 \quad (\text{C.15})$$

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  $\epsilon_t^c$  and  $\epsilon_t^i$  being white noise shocks with variances  $\sigma_c^2$  and  $\sigma_i^2$ , 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:  $S_C = (1 - b\bar{g}_A^{-1/(1-\theta)})^{-1} / [(1 - b\bar{g}_A^{-1/(1-\theta)})^{-1} - b\beta(\bar{g}_A^{1/(1-\theta)} - b)^{-1}]$ ;  $\Delta_G = \bar{g}_A^{1/(1-\theta)} / (\bar{g}_A^{1/(1-\theta)} - b)$ ;  $S_K = \theta \bar{\mu}_{ss}^* \rho_{YK} / [\theta \bar{\mu}_{ss}^* \rho_{YK} + (1-\delta) \bar{\lambda}_{ss}^* \bar{g}_A^{-1/(1-\theta)}]$ ;  $\alpha_C = \bar{C} / \bar{Y}$ ;  $\alpha_I = \bar{I} / \bar{Y}$ ;  $\alpha_K = (1-\delta) / [(1-\delta) + \delta \bar{g}_A^{1/(1-\theta)}]$ ;  $\bar{\mu}_{ss}^* = [(1 - b\bar{g}_A^{-1/(1-\theta)})^{-1} - b\beta(\bar{g}_A^{1/(1-\theta)} - b)^{-1}] / \bar{C}$ ;  $\bar{\lambda}_{ss}^* = \beta \theta \bar{\mu}_{ss}^* \rho_{YK} / [1 - \beta(1-\delta) \bar{g}_A^{-1/(1-\theta)}]$ , 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  $\rho_{YK} = \bar{Y} / \bar{K}$  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  $\beta=0.99$ ,  $\delta=0.05$ ,  $\theta=1/3$ ,  $\gamma=0.05$ ,  $\bar{g}=0.2$ ,  $\bar{g}_A=1.02^{1/4}$ ,  $\bar{c}=2/3$ ,  $\alpha_C=2/3$ ,  $\alpha_I=0.2$ . 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  $\sigma_{NN}=0.3$ ,  $\sigma_{NE}=0.3$ ,  $\sigma_u=0.25$ ,  $\sigma_v=0.25$ ,  $\sigma_c=0.25$ ,  $\sigma_n=0.25$ ,  $\sigma_g=0.25$ ,  $\sigma_i=0.25$ . As for  $\rho_{YK}$ , 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  $\rho_{YK}=1/1.7=0.5882$ . Finally, we set  $\bar{g}_I = \bar{g}_A^{1/(1-\theta)}$ : the rationale for doing this is simply that, in the steady-state,  $I_t/I_{t-1} = \bar{g}_I$ , and since the steady-state gross rate of growth of investment is equal to  $\bar{g}_A^{1/(1-\theta)}$ , it ought to be the case that  $\bar{g}_I = \bar{g}_A^{1/(1-\theta)}$ .

## 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

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

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

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

where  $P$  solves the quadratic equation<sup>5</sup>

$$FP^2 + GP + H = 0, \quad (\text{C.18})$$

$Q$  and  $V$  are given by

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

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

<sup>5</sup>See Uhlig (1999) for the solution to the quadratic equation.

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

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

Finally,  $\xi_{t|t}$  is the agents' estimate of the vector  $\xi_t$  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

In this appendix we describe how we compute the truncated theoretical SVARMA representations of Barsky and Sims' (2011) RBC model augmented with noise shocks we discuss in Section 4.1.2.

Let

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

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

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

as follows.

By post-multiplying (D.2) by  $Y'_{t-i}$ , for  $i = 1, 2, 3, \dots, p$ , we obtain a system of  $p$  equations,

$$\begin{aligned} Y_t Y'_{t-i} &= B_1 Y_{t-1} Y'_{t-i} + B_2 Y_{t-2} Y'_{t-i} + \dots + B_p Y_{t-p} Y'_{t-i} + \\ &+ A_0 \epsilon_t Y'_{t-i} + A_1 \epsilon_{t-1} Y'_{t-i} + \dots + A_q \epsilon_{t-q} Y'_{t-i}, \quad i = 1, 2, 3, \dots, p. \end{aligned} \quad (\text{D.3})$$

By the same token, by post-multiplying (D.2) by  $\epsilon'_{t-j}$ , for  $j = 1, 2, 3, \dots, q$ , we obtain a system of  $q$  equations,

$$\begin{aligned} Y_t \epsilon'_{t-j} &= B_1 Y_{t-1} \epsilon'_{t-j} + B_2 Y_{t-2} \epsilon'_{t-j} + \dots + B_p Y_{t-p} \epsilon'_{t-j} + \\ &+ A_0 \epsilon_t \epsilon'_{t-j} + A_1 \epsilon_{t-1} \epsilon'_{t-j} + \dots + A_q \epsilon_{t-q} \epsilon'_{t-j}, \quad j = 1, 2, 3, \dots, q. \end{aligned} \quad (\text{D.4})$$

By taking unconditional expectations of (D.3) and (D.4) we obtain the system of  $p+q$  equations

$$\begin{aligned} E[Y_t Y'_{t-i}] &= B_1 E[Y_{t-1} Y'_{t-i}] + B_2 E[Y_{t-2} Y'_{t-i}] + \dots + B_p E[Y_{t-p} Y'_{t-i}] + \\ &+ A_0 E[\epsilon_t Y'_{t-i}] + A_1 E[\epsilon_{t-1} Y'_{t-i}] + \dots + A_q E[\epsilon_{t-q} Y'_{t-i}], \quad i = 1, 2, 3, \dots, p. \end{aligned} \quad (\text{D.5})$$

$$\begin{aligned} E[Y_t \epsilon'_{t-j}] &= B_1 E[Y_{t-1} \epsilon'_{t-j}] + B_2 E[Y_{t-2} \epsilon'_{t-j}] + \dots + B_p E[Y_{t-p} \epsilon'_{t-j}] + \\ &+ A_0 E[\epsilon_t \epsilon'_{t-j}] + A_1 E[\epsilon_{t-1} \epsilon'_{t-j}] + \dots + A_q E[\epsilon_{t-q} \epsilon'_{t-j}], \quad j = 1, 2, 3, \dots, q. \end{aligned} \quad (\text{D.6})$$

which ought to be jointly solved in order to recover the matrices  $B_1, B_2, \dots, B_p, A_1, A_2, \dots, A_q$ . As for  $A_0$ , on the other hand, we have that  $A_0 = \tilde{A}_0$ . Equations (D.5)-(D.6) can be put into the matrix form

$$h = H\beta \quad (\text{D.7})$$

where

$$h = \begin{bmatrix} E[Y_{t-1}Y_t'] \\ \dots \\ E[Y_{t-p}Y_t'] \\ E[\epsilon_{t-1}Y_t'] \\ \dots \\ E[\epsilon_{t-q}Y_t'] \end{bmatrix} \quad H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \quad \beta = \begin{bmatrix} B_1' \\ \dots \\ B_p' \\ A_1' \\ \dots \\ A_q' \end{bmatrix}$$

and

$$H_{11} = \begin{bmatrix} E[Y_{t-1}Y_{t-1}'] & E[Y_{t-1}Y_{t-2}'] & \dots & E[Y_{t-1}Y_{t-p}'] \\ E[Y_{t-2}Y_{t-1}'] & E[Y_{t-2}Y_{t-2}'] & \dots & E[Y_{t-2}Y_{t-p}'] \\ \dots & \dots & \dots & \dots \\ E[Y_{t-p}Y_{t-1}'] & E[Y_{t-p}Y_{t-2}'] & \dots & E[Y_{t-p}Y_{t-p}'] \end{bmatrix}$$

$$H_{12} = \begin{bmatrix} E[Y_{t-1}\epsilon'_{t-1}] & E[Y_{t-1}\epsilon'_{t-2}] & \dots & E[Y_{t-1}\epsilon'_{t-q}] \\ 0_{N \times N} & E[Y_{t-2}\epsilon'_{t-2}] & \dots & E[Y_{t-2}\epsilon'_{t-q}] \\ \dots & \dots & \dots & \dots \\ 0_{N \times N} & 0_{N \times N} & \dots & E[Y_{t-p}\epsilon'_{t-q}] \end{bmatrix}, \quad H_{21} = H_{12}'$$

$$H_{22} = \begin{bmatrix} \Omega & 0_{N \times N} & \dots & 0_{N \times N} \\ 0_{N \times N} & \Omega & \dots & 0_{N \times N} \\ \dots & \dots & \dots & \dots \\ 0_{N \times N} & 0_{N \times N} & \dots & \Omega \end{bmatrix}$$

The expectations terms in the expressions for  $h$  and  $H$  are equal to

$$E[Y_t Y_{t-i}'] = \sum_{k=i}^{\infty} \tilde{A}_k \Omega \tilde{A}'_{k-i} \quad \text{and} \quad E[Y_{t-i} Y_t'] = [E(Y_t Y_{t-i}')]', \quad i = 0, 1, 2, \dots, p \quad (\text{D.8})$$

$$E[Y_t \epsilon'_{t-j}] = \tilde{A}_j \Omega \quad \text{and} \quad E[\epsilon_{t-j} Y_t'] = [E(Y_t \epsilon'_{t-j})]', \quad j = 0, 1, 2, \dots, q \quad (\text{D.9})$$

Based on this, the solution to (D.7) can be immediately be computed as  $\beta = H^{-1}h$ , from which we can recover the matrices  $B_1, B_2, \dots, B_p, A_1, A_2, \dots, A_q$ . One practical issue in the computation of the solution is the following. Since in Barsky and Sims' RBC model  $Y_t$  is I(1), all of the matrices  $E[Y_t Y_{t-i}']$ ,  $i = 0, 1, 2, \dots, p$ , explode to infinity. (The elements of the matrices  $E[Y_t \epsilon'_{t-j}] = \tilde{A}_j \Omega$ , on the other hand, are always finite.) In order to solve equation (D.7), therefore, instead of computing  $E[Y_t Y_{t-i}']$  based on the infinite sum in (D.8), we compute it based on

$$E[Y_t Y_{t-i}'] = \sum_{k=i}^K \tilde{A}_k \Omega \tilde{A}'_{k-i}, \quad i = 0, 1, 2, \dots, p \quad (\text{D.8})$$

with  $K$  'large'. We compute the solution to equation (D.7) for different values of  $K$ . If  $K$  is sufficiently large (i.e., greater than about 100), further increasing it produces results which are numerically indistinguishable.

## 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?

Based on the DGP described in Section 3 and Appendix C, we generate 1,000 artificial samples of length  $T = 1,000$ , in order to focus on the ability of the proposed methodology to recover the DGP’s main features asymptotically (in the present case, in ‘very large samples’). Based on each sample we then estimate a VARMA(1,1) based on the methodology described in Section 4, imposing *exactly* the same restrictions we impose when we work with the actual data.

Figures E.1 and E.2 show the means—across all of the 1,000 Monte Carlo simulations—of the 50th, 16th, 84th, 5th, and 95th percentiles of the posterior distributions of the IRFs and the fractions of FEV, respectively, based on *point identification*. Figures E.3 and E.4 show the means—again, across all of the 1,000 Monte Carlo simulations—of the medians of the posterior distributions of the upper and lower envelopes of the IRFs and the fractions of FEV, respectively (the upper and lower envelopes have been computed, for each draw, across the set of all possible representations obtained by ‘flipping the roots’—see the discussion in Sections 4.2.2 and 5.1); the 84th percentile of the posterior distribution of the upper envelope; and the 16th percentile of the posterior distribution of the lower envelope, all based on *set identification*.

Results based on *point identification* require little discussion, as they are uniformly excellent. In particular, the fractions of FEV in Figure E.2 are all captured with great precision. As for the IRFs, although they are typically captured precisely, the following should be noted. First, in a few cases (see, in particular, the responses of TFP to news and non-news shocks) the IRFs exhibit some slight tendency towards mean-reversion. Since we are here using samples of length 1,000, this might be seen, at first sight, as problematic. In fact, it is not, and it simply originates from the fact that, as mentioned in the text, in estimation we are imposing stationarity (i.e., we are rejecting all of the draws for which the largest eigenvalue of the VAR portion of the VARMA is greater than or equal to one). This means that, in principle, we could trivially get rid of this simply by not imposing the stationarity constraint. Second, in two cases—consumption, and especially GDP—the IRFs to noise shocks exhibit some imprecision. The obvious explanation is that, since in the DGP these shocks explain very little of anything, precisely estimating their impact is comparatively more difficult than for news and non-news shocks.

Turning to the results based on *set identification*, the main feature emerging from Figures E.3–E.4 is that whereas the *upper envelopes* of the posterior distributions of the IRFs and fractions of FEV systematically capture correctly the true objects, the lower envelopes do not. The obvious explanation for this is that whereas a VARMA entails many observationally equivalent representations, only one of them will be the ‘correct’ one. Within the present context, the evidence in Figures E.3–E.4—especially when compared to the corresponding evidence in Figures E.1–E.2—naturally suggests that the assumption that all non-zero roots are outside the unit circle is indeed justified, in the sense of being consistent with the DGP. On the other hand, the other representations—which we obtain by flipping the non-zero roots to lie inside the unit circle—systematically fail to capture the true DGP.

To focus ideas, consider the the MA(1) process:  $y_t = a\epsilon_t + \epsilon_{t-1}$ , with  $\epsilon_t \sim N(0, \sigma^2)$ . As it is well known, an observationally equivalent representation for  $y_t$  is given by  $y_t = (1/a)\epsilon_t + \epsilon_{t-1}$ , with  $\epsilon_t \sim N(0, a^2\sigma^2)$ . Because of this identification problem, when we estimate the MA(1) we need to impose a restriction on the parameter—either  $|a| < 1$ , which corresponds to the MA root being outside the unit circle, or  $|a| > 1$ , which corresponds to the MA root being inside the unit circle. Now, suppose we do a Monte Carlo experiment where the DGP is  $y_t = 0.5\epsilon_t + \epsilon_{t-1}$ , with  $\epsilon_t \sim N(0, 1)$ . If we estimate the MA(1) under the restriction  $|a| < 1$ , for large enough  $T$  we will recover  $a=0.5$ ,  $\sigma^2=1$ . However, if we estimate the process using the alternative restriction  $|a| > 1$ , we will obtain, for large  $T$ ,  $a=2$ ,  $\sigma^2 = 0.25$ . Crucially, in the second representation the variance is smaller than the true variance. This is exactly what is happening with the results from the Monte Carlo experiment: One of the representations—let’s call it ‘almost fundamental’, to mean that all non-zero roots are outside the unit circle—is correct, whereas the others with the roots flipped inside the unit circle are observationally equivalent, but not correct with respect to the true DGP.

Overall, our own assessment is that the performance of the proposed estimation and identification

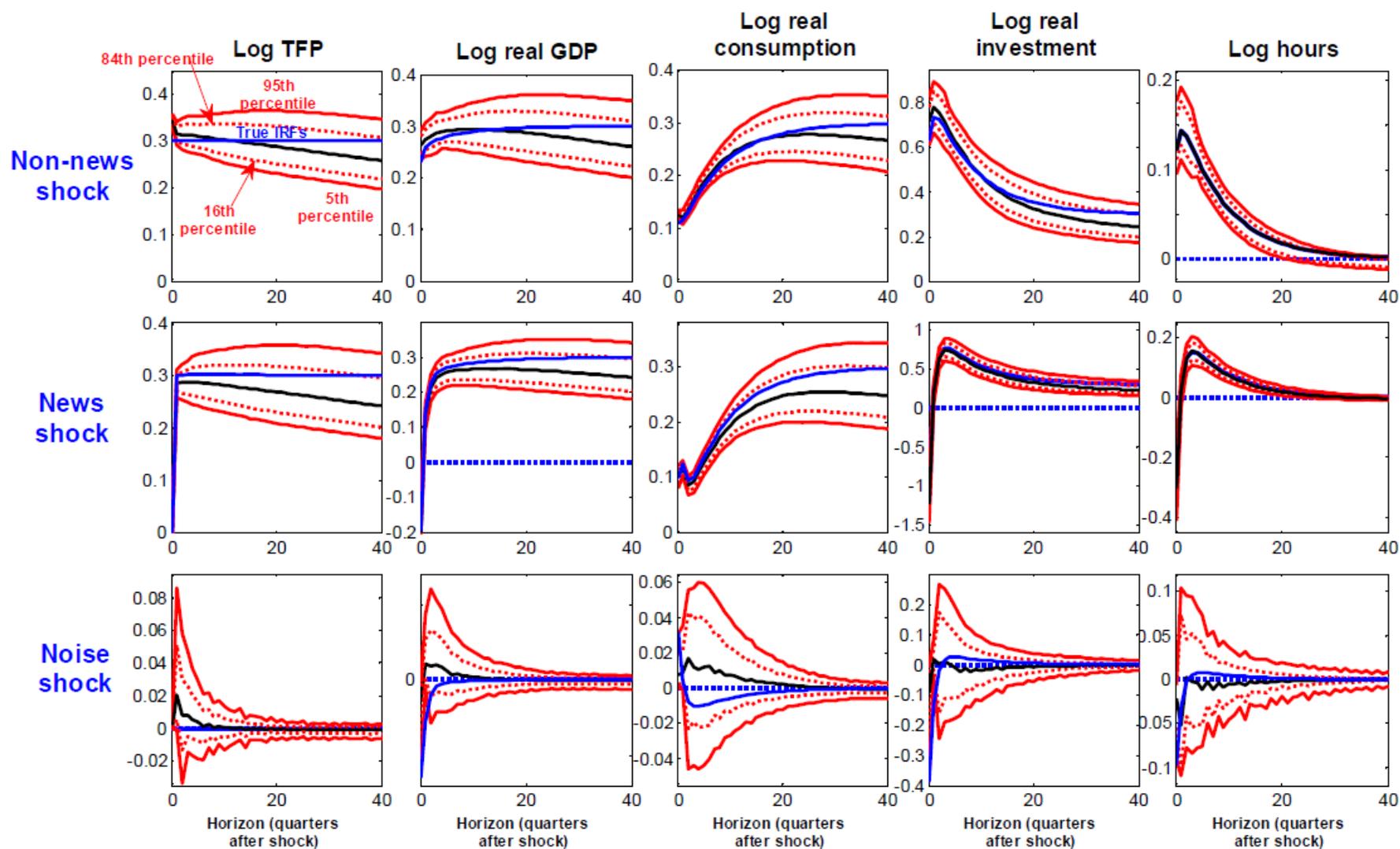


Figure E.1 Results from the Monte Carlo exercise based on point-identification: Means, across all of the Monte Carlo simulations, of the 50th, 16th, 84th, 5th, and 95th percentiles of the posterior distributions of the impulse-response functions

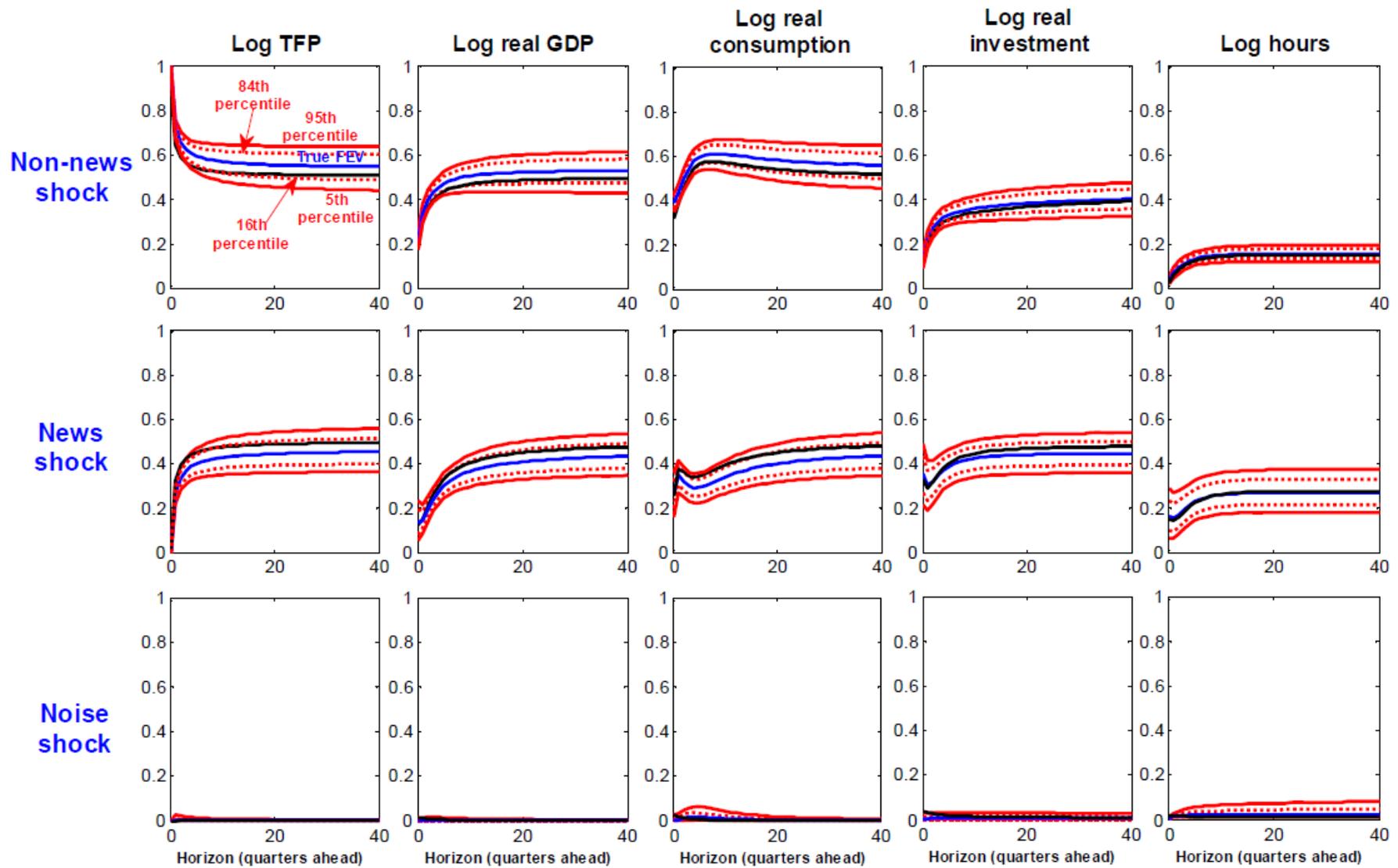


Figure E.2 Results from the Monte Carlo exercise based on point-identification: Means, across all of the Monte Carlo simulations, of the 50th, 16th, 84th, 5th, and 95th percentiles of the posterior distributions of the fractions of forecast error variance

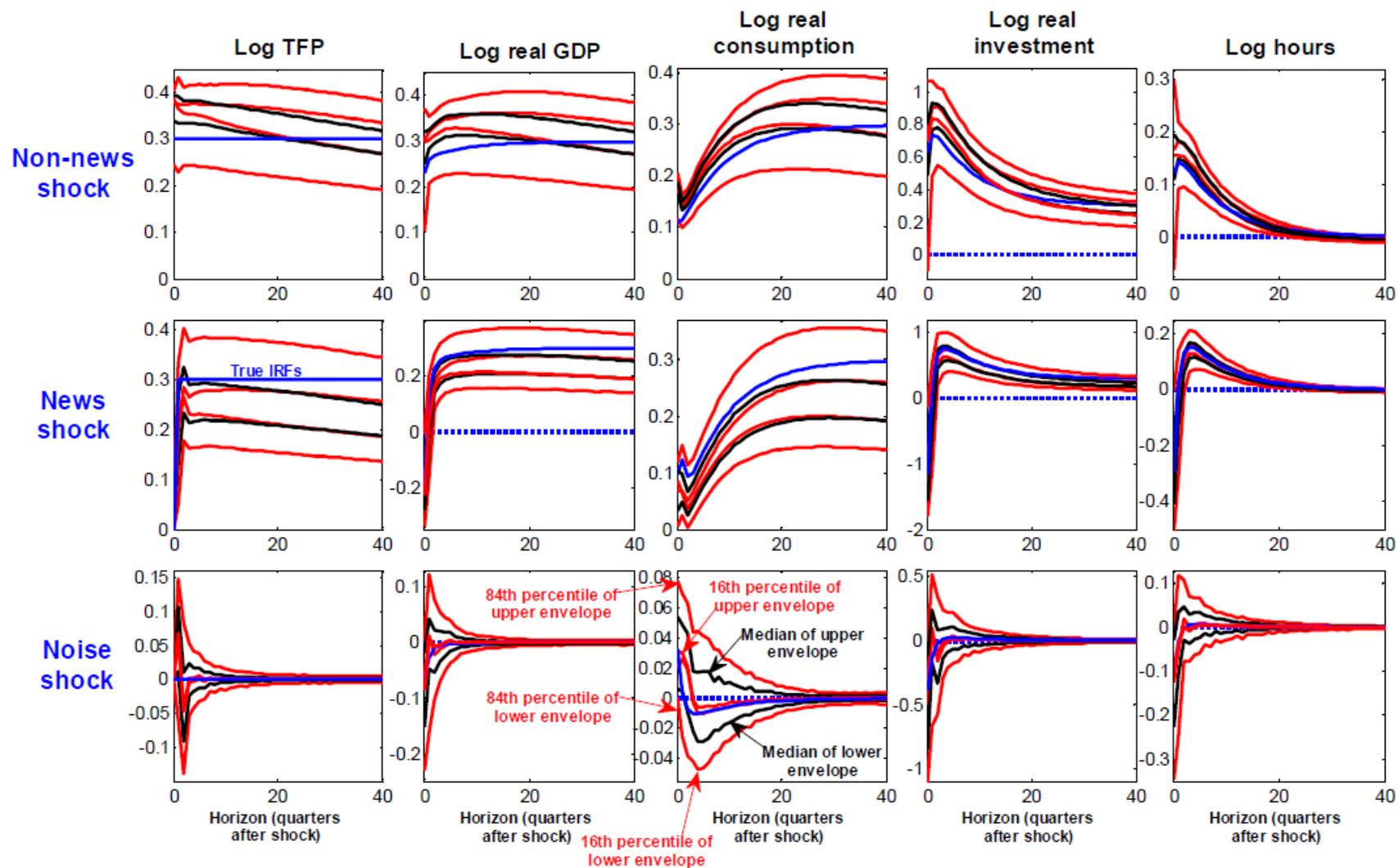


Figure E.3 Results from the Monte Carlo exercise based on set-identification: Means, across all of the Monte Carlo simulations, of the median, the 16th, and 84th percentiles of the posterior distributions of the upper and lower envelopes of the impulse-response functions

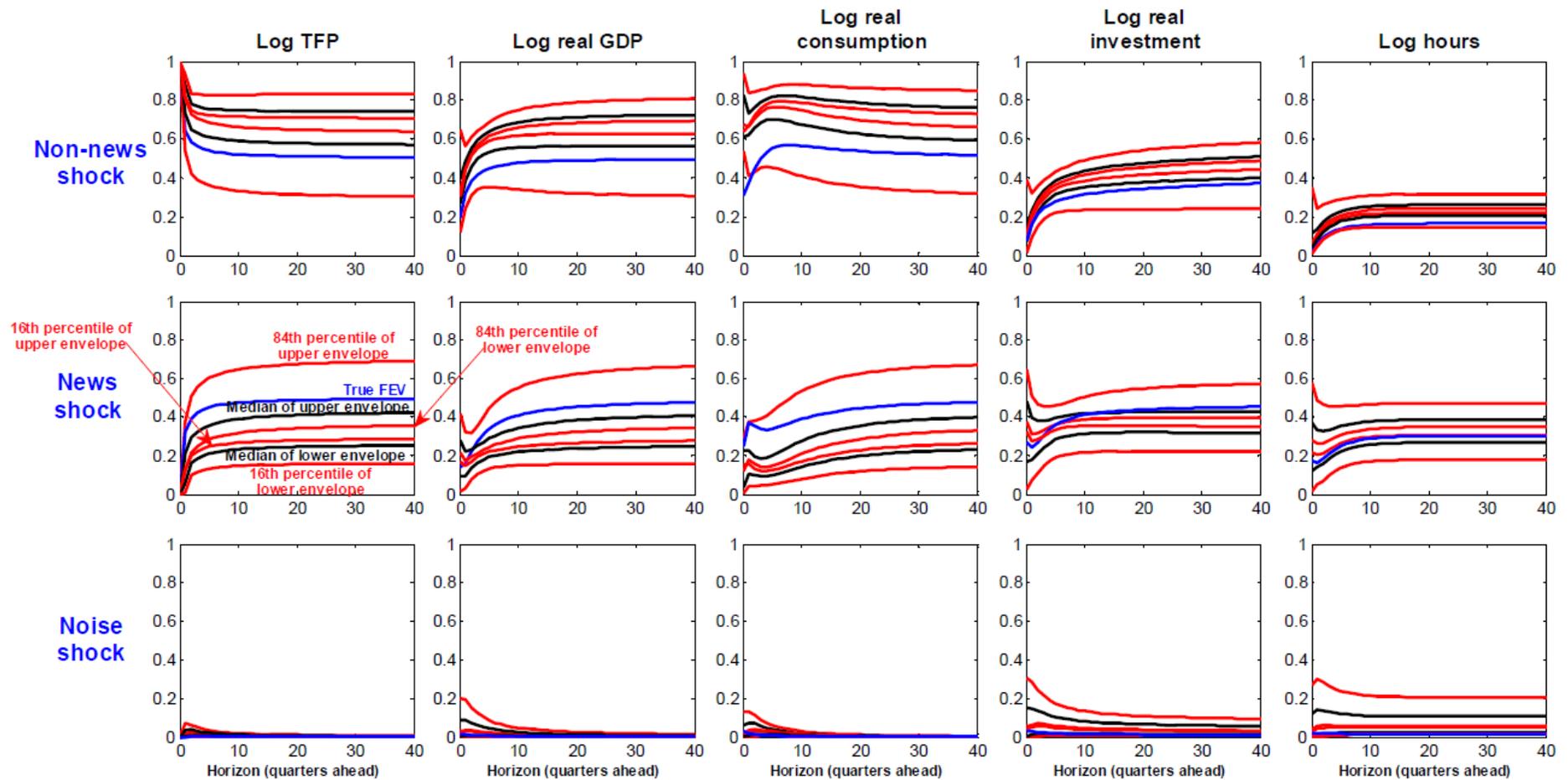


Figure E.4 Results from the Monte Carlo exercise based on set-identification: Means, across all of the Monte Carlo simulations, of the median, the 16th, and 84th percentiles of the posterior distributions of the upper and lower envelopes of the fractions of forecast error variance

methodology against this specific DGP is excellent, which justifies applying it to the data.

## 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.<sup>6</sup> 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) 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 et al. (2002). For latent variable models there are a few distinct variants depending on the exact notion of the likelihood Celeux et al. (2006). Given a likelihood function  $f(y|\theta)$ , the DIC is defined as:

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

where

$$\overline{D(\theta)} = -2\text{E}_\theta(\ln f(y|\theta) | y)$$

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

Following Chan et al. (2016), we use the likelihood implied by the system  $B(L)y_t = \Theta(L)\varepsilon_t$ ,  $\varepsilon_t \sim N(0, \Sigma)$ , or equivalently

$$y_t = \sum_{j=1}^p B_j y_{t-j} + \sum_{j=1}^q \Theta_j \varepsilon_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma), \quad (3)$$

---

<sup>6</sup>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  $r_t = (1 + R_t/100)^{1/4} - 1$ .

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

To derive this density, we stack (3) over  $t$  and obtain:

$$y = b + \Theta\epsilon, \tag{4}$$

where  $\epsilon = (\epsilon'_1, \dots, \epsilon'_T)' \sim N(0, I_T \otimes \Sigma)$ ,  $b = ((\sum_{j=1}^p B_j y_{1-j})', \dots, (\sum_{j=1}^p B_j y_{T-j})')'$  and  $\Theta$  is a  $Tn \times Tn$  lower triangular matrix with the identity matrix  $I_n$  on the main diagonal block,  $\Theta_1$  on first lower diagonal block,  $\Theta_2$  on second lower diagonal block, and so forth. Hence, we have

$$(y | B_1, \dots, B_p, \Theta_1, \dots, \Theta_q, \Sigma) \sim N(b, \Theta(I_T \otimes \Sigma)\Theta').$$

Since the covariance matrix  $\Theta(I_T \otimes \Sigma)\Theta'$  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  $n=8$  dominates 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  $n=8$ .

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

<sup>a</sup> 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, this online appendix contains (below) several additional results. In particular, for our main application with TFP, we present results based on systems featuring either 6 or 8 series; based on VARMA( $p, q$ )'s with  $p = 4$  and  $q = 1, 2, 3$ ; based on either point or set identification; and, in a few cases, based on imposing the restriction that the absolute magnitudes of the IRFs to news shocks are greater than the corresponding magnitudes of the IRFs to noise shocks two periods after impact. We also show results for systems featuring either 10 or 15 series; based on  $p = 4$  and  $q = 1$ ; based on either point or set identification; and without imposing the restriction on the absolute magnitudes of the IRFs to news and noise shocks.

The main points we want to stress here are that

- (1) our main finding—noise shocks play a uniformly negligible role in driving macroeconomic fluctuations—is remarkably robust across all specifications; and
- (2) results pertaining to both the IRFs and the fractions of FEVs are very close across alternative specifications.

## I News, Noise and Mixed Representations

In this section, we illustrate several properties of structural representations involving news and noise shocks. The goal is to highlight conditions under which such representations are observationally equivalent and when they are unique. This has been the subject of close scrutiny in recent literature, with important implications for the type of time series methods that are suitable in empirical work (e.g., Blanchard et al. (2013); Forni et al. (2017); Chahrour and Jurado (2018)).

As a simple example, consider a *news* representation defined by the fundamental process

$$a_t = \epsilon_{t-1}^{NE} + \epsilon_t^{NN}, \quad \begin{pmatrix} \epsilon_t^{NN} \\ \epsilon_t^{NE} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{NN}^2 & 0 \\ 0 & \sigma_{NE}^2 \end{pmatrix} \right),$$

along with the agents' time- $t$  information set consisting of the history  $\{a_{t-\tau}, s_{t-\tau}\}_{\tau=0}^{\infty}$ , with  $s_t \equiv \epsilon_t^{NE}$ .<sup>7</sup> In this case, agents formulate rational expectations about future fundamentals as  $\hat{a}_{1,t} \equiv E_t(a_{t+1}) = s_t = \epsilon_t^{NE}$ .

Alternatively, Chahrour and Jurado (2018) define a *noise* representation by the signal process

$$s_t = a_{t+1} + u_t, \quad \begin{pmatrix} a_t \\ u_t \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right),$$

along with the agents' time- $t$  information set consisting of the history  $\{a_{t-\tau}, s_{t-\tau}\}_{\tau=0}^{\infty}$ . In this case, agents formulate rational expectations about future fundamentals as  $\hat{a}_{1,t} \equiv E_t(a_{t+1}) = \kappa s_t = \kappa a_{t+1} + \kappa u_t$ , where  $\kappa = \sigma_a^2 / (\sigma_a^2 + \sigma_u^2)$ .

Chahrour and Jurado (2018) show (in Proposition 1) that if an econometrician observes *only* the two variables  $(a_t, \hat{a}_{1,t})$ , then *news* and *noise* representations of this bivariate system are *observationally equivalent* if and only if

$$\sigma_a^2 = \sigma_{NN}^2 + \sigma_{NE}^2, \quad \frac{\sigma_u^2}{\sigma_a^2} = \frac{\sigma_{NN}^2}{\sigma_{NE}^2}. \quad (5)$$

It follows that to estimate the effects of noise shocks, one may proceed by first estimating  $\sigma_{NN}^2$  and  $\sigma_{NE}^2$  in the news representations, then obtaining  $\sigma_a^2$  and  $\sigma_u^2$  in the noise representation through (5). Moreover, if the econometrician observes additional variables that are linear combinations of  $a_{t-\tau}$  and  $\hat{a}_{1,t-\tau}$ ,  $\tau \geq 0$ , then the observational equivalence continues to hold, so that the approach of estimating

<sup>7</sup>Note that this information set is *equivalent* to  $\{\epsilon_{t-\tau}^{NN}, \epsilon_{t-\tau}^{NE}\}_{\tau=0}^{\infty}$  as stated in Chahrour and Jurado (2018).

the news representation, then transforming to the noise representation, may be generalized under these assumptions.

However, Forni et al. (2017) put forth the idea that certain observable variables available to the econometrician effectively “reveal” the signal  $s_t$  that is observed by the agents. Indeed, this concept plays a central role in their strategy aimed at identifying news and noise shocks from macroeconomic data. A simple example of such a variable may be  $y_t = s_t + \xi_t$ , with  $\xi_t \sim N(0, \sigma_\xi^2)$  being measurement error, uncorrelated with any structural shocks in the model.

Consequently, if the econometrician observes the trivariate system  $(a_t, \hat{a}_{1,t}, y_t)'$  then the two representations are given by:

$$\begin{aligned} \text{news :} \quad & \begin{pmatrix} a_t \\ \hat{a}_{1,t} \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & L & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_t^{NN} \\ \epsilon_t^{NE} \\ \xi_t \end{pmatrix}, \quad \begin{pmatrix} \epsilon_t^{NN} \\ \epsilon_t^{NE} \\ \xi_t \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{NN}^2 & 0 & 0 \\ 0 & \sigma_{NE}^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{pmatrix} \right), \\ \text{noise :} \quad & \begin{pmatrix} a_t \\ \hat{a}_{1,t} \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \kappa L^{-1} & \kappa & 0 \\ L^{-1} & 1 & 1 \end{pmatrix} \begin{pmatrix} a_t \\ u_t \\ \xi_t \end{pmatrix}, \quad \begin{pmatrix} a_t \\ u_t \\ \xi_t \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{pmatrix} \right), \end{aligned}$$

where  $L^{-1}$  is the *forward* operator (i.e.,  $x_{t+1} = L^{-1}x_t$ ).

However, it is easy to verify that the spectral density generated by the news representation is *not equal* to the spectral density under the noise representation unless  $\sigma_u^2 = 0$ . To see this, observe that under the news representation,  $\text{cov}(a_t, \hat{a}_{1,t-1}) = \text{cov}(a_t, y_{t-1}) = \sigma_{NE}^2$ . Under the noise representation, on the other hand,  $\text{cov}(a_t, \hat{a}_{1,t-1}) = \kappa\sigma_a^2$  and  $\text{cov}(a_t, y_{t-1}) = \sigma_a^2$ . For observational equivalence to hold, we must have  $\text{cov}(a_t, \hat{a}_{1,t-1}) = \text{cov}(a_t, y_{t-1})$ , or equivalently  $\kappa\sigma_a^2 = \sigma_a^2$ . Assuming  $\sigma_a^2 \neq 0$ , this condition is only satisfied when  $\sigma_u^2 = 0$ ; that is, when noise shocks are not present. Hence, one can no longer obtain an estimate of the noise representation by transforming an estimated news representation—the noise representation must be estimated directly in this case.

In the present paper, we are not interested in estimating either a “pure news” representation or a “pure noise” representation. Instead, we focus on a representation where macroeconomic data is driven by four shocks: a permanent surprise shocks, a transitory surprise shock, a news shock and a noise shock. Ignoring for ease of exposition the transitory shock (and because it is sufficient to illustrate the point), such a *mixed* representation may be formulated as:

$$a_t = \epsilon_{t-1}^{NE} + \epsilon_t^{NN}, \quad (6)$$

where agents observe in each period  $a_t$  along with the signal

$$s_t = \epsilon_t^{NE} + u_t, \quad (7)$$

and with

$$\begin{pmatrix} \epsilon_t^{NN} \\ \epsilon_t^{NE} \\ u_t \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{NN}^2 & 0 & 0 \\ 0 & \sigma_{NE}^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix} \right).$$

It turns out that for a bivariate system  $(a_t, \hat{a}_{1,t})$ , the mixed representation in (6)-(7) is observationally equivalent to a *pair* of news and noise representations (Chahrour and Jurado, 2018, Proposition 2). An important insight in this result, however, is that it depends crucially on the way the fundamental process is defined in (6).

In particular, the mixed representation will be unique for a system involving variables that reveal the signal (as illustrated above), but would also be unique for systems *not including* signal reveal variables if a more general fundamental process is specified. To clarify the intuition, consider a slightly more elaborate fundamental process defined by

$$a_t = \epsilon_{t-1}^{NE} + \theta\epsilon_{t-2}^{NE} + \epsilon_t^{NN}. \quad (8)$$

Clearly, when  $\theta = 0$  this process reverts to the one given by (6). However, with  $\theta \neq 0$ , the resulting mixed representation is unique, even for systems that contain only observations of the fundamental  $a_t$  and future expectations  $\hat{a}_{j,t}$ ,  $j \geq 1$ , but not any particular variables that reveal the signal.

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

$$\hat{a}_{1,t} \equiv E_t(a_{t+1}) = \kappa_0 s_t + \theta \kappa_1 a_t + \theta \kappa_2 s_{t-1}, \quad (9)$$

$$\hat{a}_{2,t} \equiv E_t(a_{t+1}) = \theta \kappa_0 s_t, \quad (10)$$

$$\hat{a}_{\tau,t} \equiv E_t(a_{t+\tau}) = 0, \quad \tau \geq 3, \quad (11)$$

where

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

The fundamentals and expectations have the following important properties:

$$E(\hat{a}_{2,t} a_t) = 0, \quad E(\hat{a}_{1,t} a_{t-1}) = \theta \kappa_1 \sigma_{NE}^2. \quad (12)$$

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

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

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

$$\begin{aligned} E(\hat{a}_{2,t} a_t) &= E(E_t(\eta_{0,t+2}) \eta_{0,t}) + E(E_t(\eta_{1,t+1}) \eta_{1,t-1}) + \sum_{j=2}^{\infty} E(\eta_{j,t+2-j} \eta_{j,t-j}) \\ &= E(\eta_{0,t+2} \eta_{0,t}) + E(\eta_{1,t+1} \eta_{1,t-1}) + \sum_{j=2}^{\infty} E(\eta_{j,t+2-j} \eta_{j,t-j}) \end{aligned} \quad (13)$$

$$= E(\eta_{0,t+1} \eta_{0,t-1}) + E(\eta_{1,t} \eta_{1,t-2}) + \sum_{j=2}^{\infty} E(\eta_{j,t+1-j} \eta_{j,t-1-j}) \quad (14)$$

$$= E(E_t(\eta_{0,t+1}) \eta_{0,t-1}) + E(\eta_{1,t} \eta_{1,t-2}) + \sum_{j=2}^{\infty} E(\eta_{j,t+1-j} \eta_{j,t-1-j}) \quad (15)$$

$$= E(\hat{a}_{1,t} a_{t-1}),$$

where (13) and (15) 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 (14) 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 (12) 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 (8)-(7) 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 (7) and (8).

Hence,  $\theta \neq 0$  is the key feature that breaks the observational equivalence between this mixed representation and any potential news / noise pair. It follows that mixed representations of the type analysed in this paper are unique, unless very restrictive assumptions are imposed on the fundamental process. In the general case, therefore, mixed representations must be estimated directly, and the structural VARMA methodology employed in this paper once again emerges as the most suitable approach.

## References

- Robert Barsky and Eric Sims. News shocks and business cycles. *Journal of Monetary Economics*, 58 (3):273–289, 2011.
- Robert Barsky, Susanto Basu, and K. Lee. Whither news shocks? in *Jonathan Parker and Michael Woodford, editors, NBER Macroeconomics Annuals*, 29:225–264, 2014.
- Olivier Blanchard, Jean-Paul L’Huillier, and Guido Lorenzoni. News, noise, and fluctuations: An empirical exploration. *American Economic Review*, 103(7):3045–70, 2013.
- G. Celeux, F. Forbes, C.P. Robert, and D.M. Titterton. Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4):651–674, 2006.
- Ryan Chahrour and Kyle Jurado. News or noise? The missing link. *American Economic Review*, 108 (7):1702–1736, 2018.
- Joshua Chan and Angelia Grant. Fast computation of the deviance information criterion for latent variable models. *Computational Statistics and Data Analysis*, 100:847–859, 2016.
- Joshua Chan, Eric Eisenstat, and Gary Koop. Large Bayesian VARMA’s. *Journal of Econometrics*, 192 (2):374–390, 2016.
- Joshua Chan, Eric Eisenstat, and Gary Koop. Quantifying the effects of noise shocks: A structural VARMA approach. *Working Paper*, 2019. manuscript available at <http://joshuachan.org/papers/QuantifyingNoiseShocks.pdf>.
- V.V. Chari, P. Kehoe, and E. McGrattan. Are structural VARs with long-run restrictions useful in developing business cycle theory? *Journal of Monetary Economics*, 55(8):1337–1352, 2008.
- Carlo Dadda and Antonello Scorcu. On the time stability of the output-capital ratio. *Economic Modelling*, 20(6):1175–1189, 2003.
- J. Fernandez-Villaverde, J. Rubio-Ramirez, T.J. Sargent, and M. Watson. A’s, B’s, C’s (and D’s) for understanding VARs. *American Economic Review*, 97:1021–1026, 2007.
- Mario Forni, Luca Gambetti, Marco Lippi, and Luca Sala. Noisy news in business cycles. *American Economic Journal: Macroeconomics*, forthcoming, 2017.
- E. George, D. Sun, and S. Ni. Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142:553–580, 2008.
- J.D. Hamilton. *Time Series Analysis*. Princeton, NJ, Princeton University Press, 1994.
- Lars Peter Hansen, William T. Roberds, and Thomas J. Sargent. Time series implications of present value budget balance and of martingale models of consumption and taxes. in *Lars Peter Hansen and Thomas J. Sargent, editors, Rational Expectations Econometrics*, Boulder: Westview Press, pages 121–161, 1991.
- D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. vanderLinde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64(4):583–639, 2002.
- Harald Uhlig. Comment on: The robustness of identified VAR conclusions about money. *Carnegie-Rochester Conference Series on Public Policy*, 49:245–263, 1998.
- Harald Uhlig. A toolkit for analysing nonlinear dynamic stochastic models easily. in *Ramon Marimon and Andrew Scott, eds, Computational Methods for the Study of Dynamic Economies*, Oxford University Press, pages 30–61, 1999.

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

# Figures for Online Appendix

Additional results for Barsky and Sims' (2011) model  
augmented with noise shocks about TFP

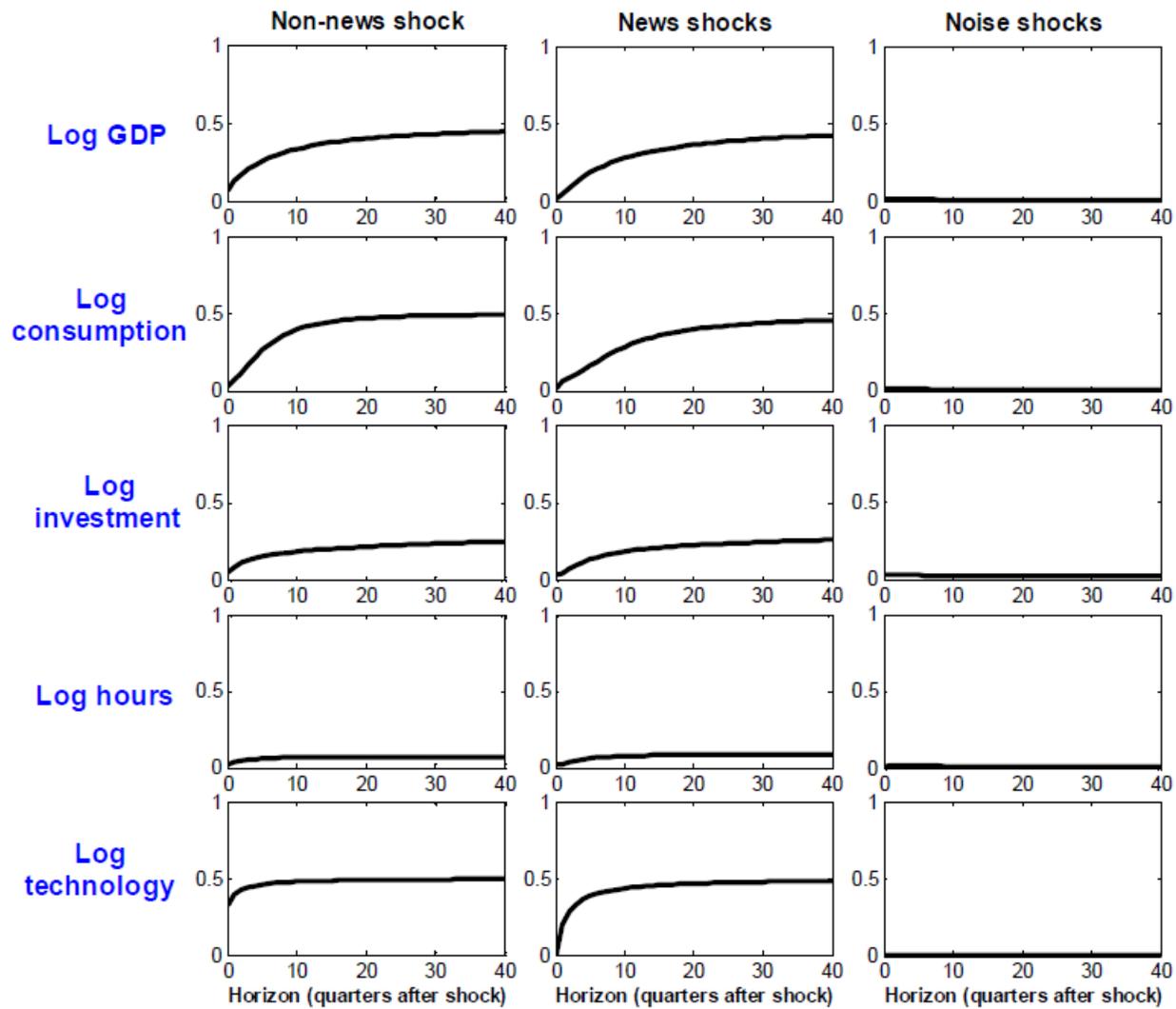


Figure I.1 Fractions of forecast error variance explained by non-news, news and noise shocks for Barsky and Sims' (2011) model augmented with noise shocks about TFP

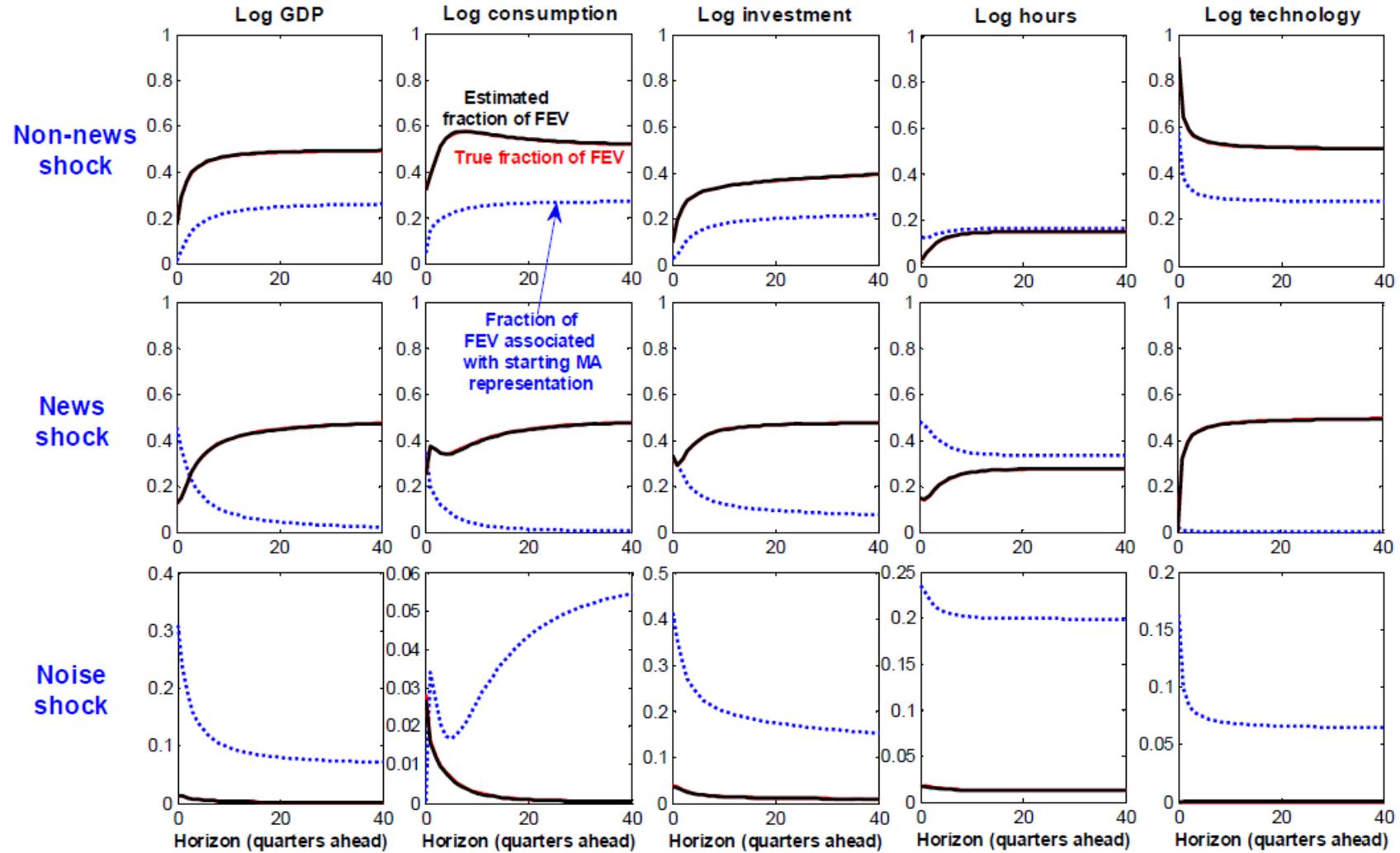


Figure I.2 Assessing the identifying restrictions in population: Theoretical and estimated fractions of FEV based on Barsky and Sims' (2011)RBC model augmented with noise shocks

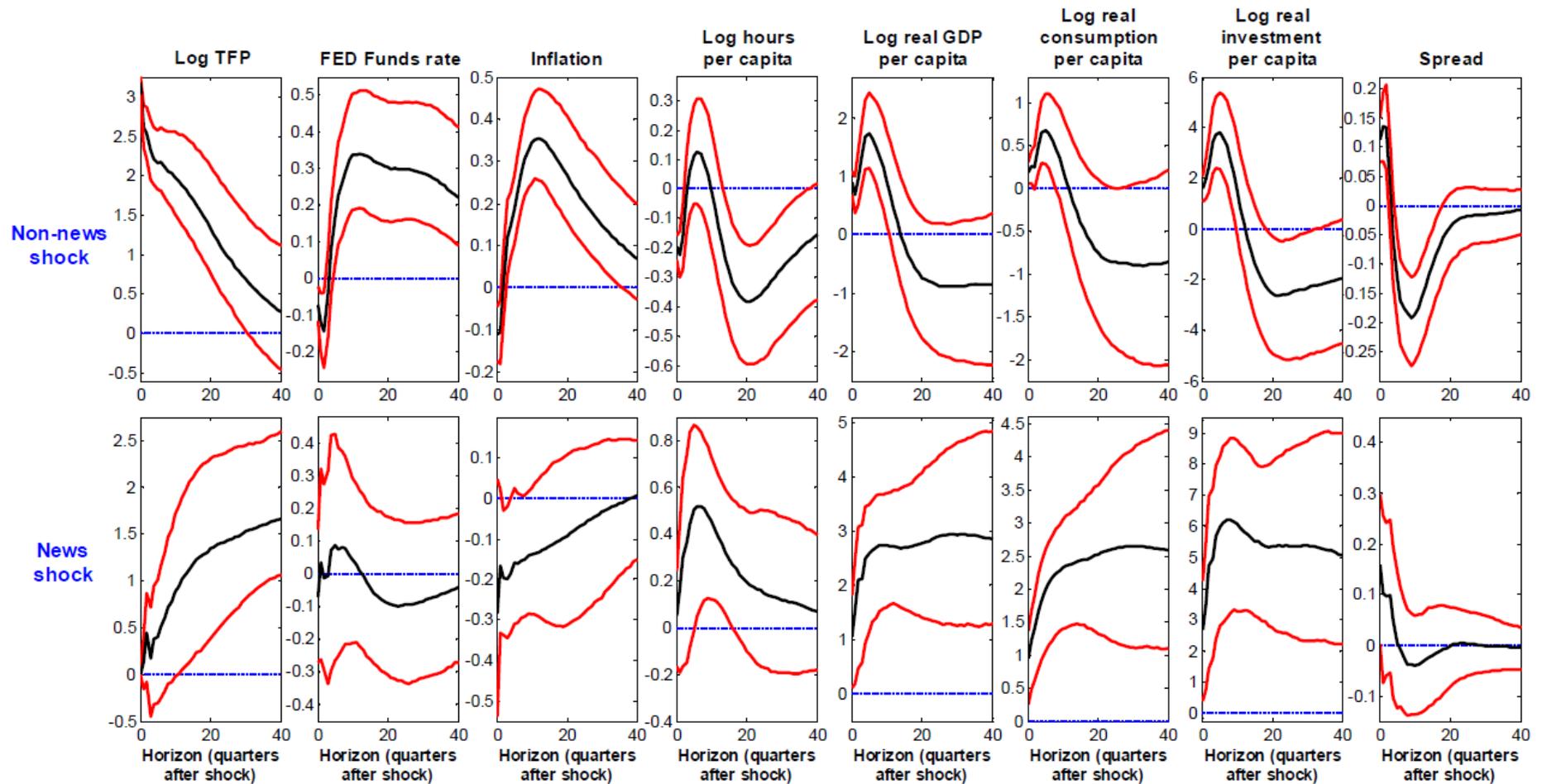


Figure I.3 Comparison with Barsky and Sims (2011) for the application with TFP: Medians, and 16-84 and 5-95 percentiles of the posterior distributions of the impulse-response functions to the news and non-news shocks

Impulse-response functions to news and noise  
shocks in the New Keynesian model

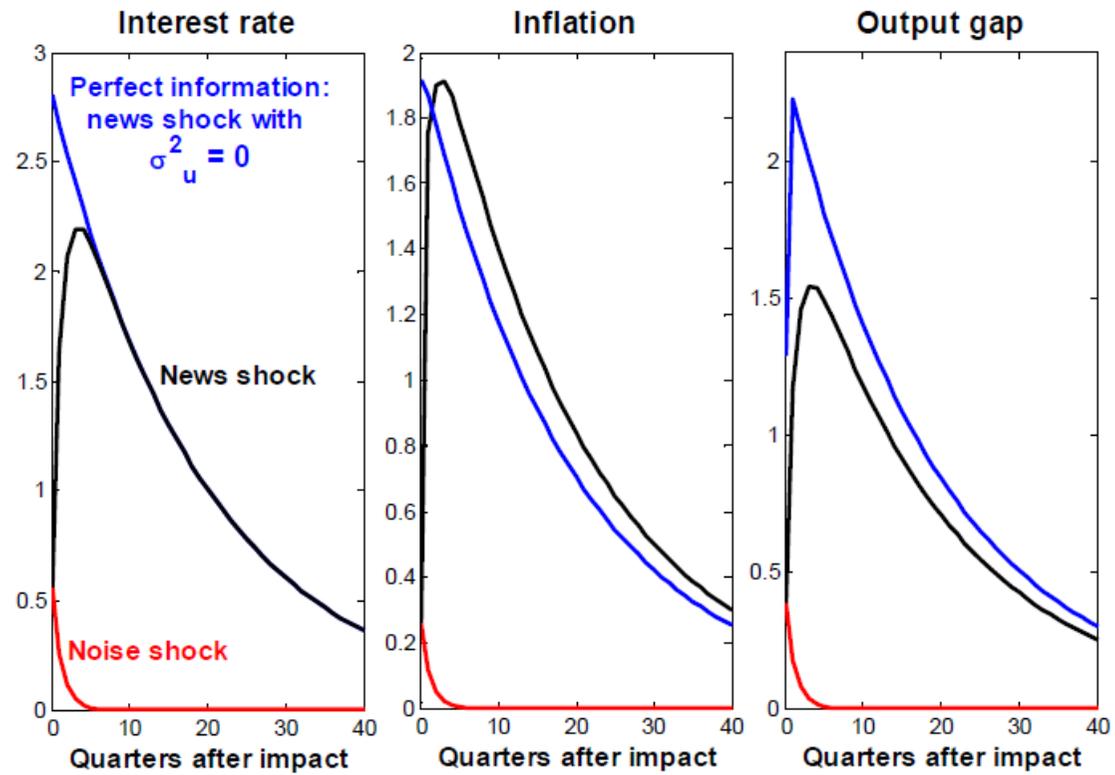


Figure II.1 Impulse-response functions to news and noise shocks within the standard New Keynesian model

Results based on VARMA with 6 series and point identification, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

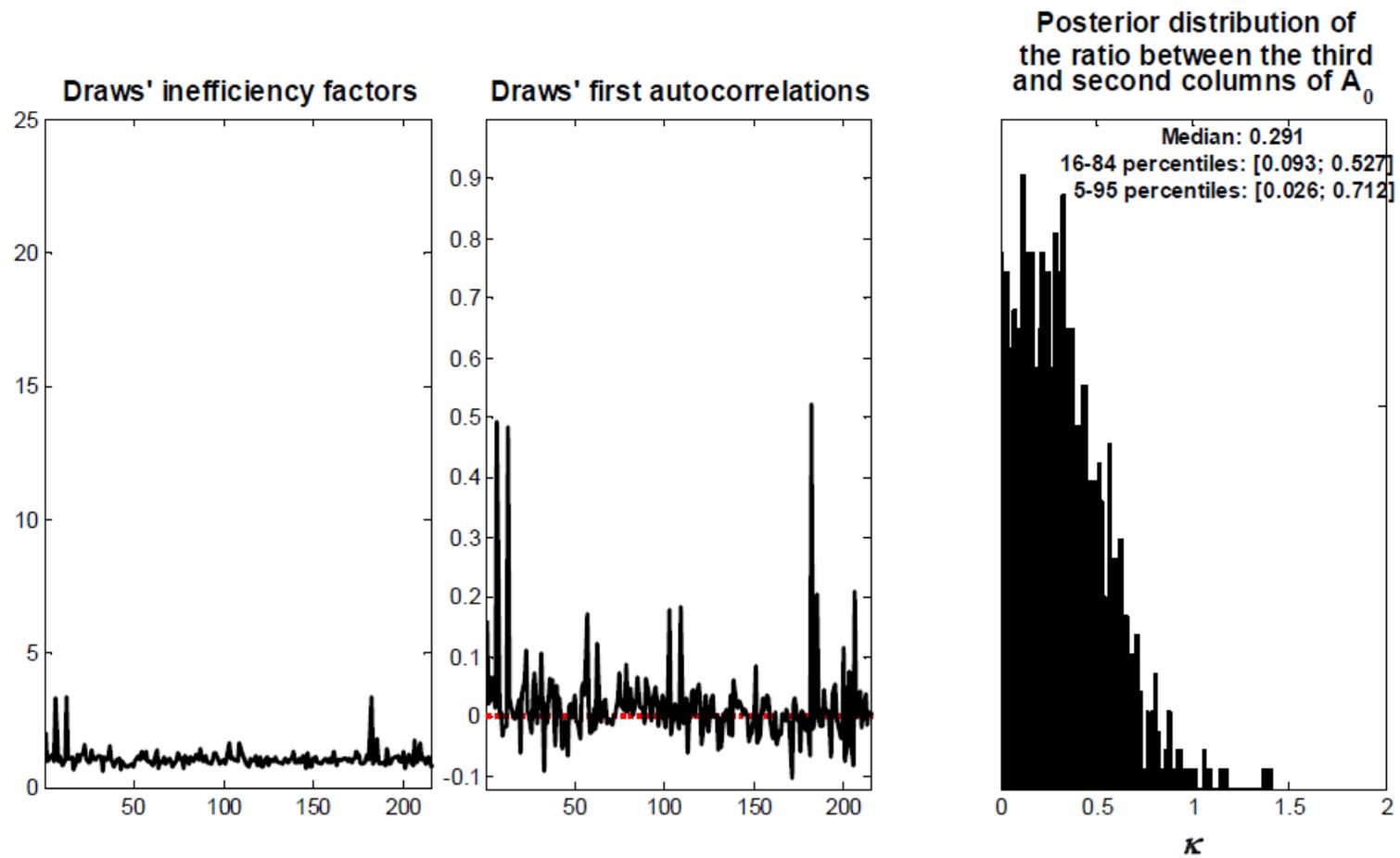


Figure III.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

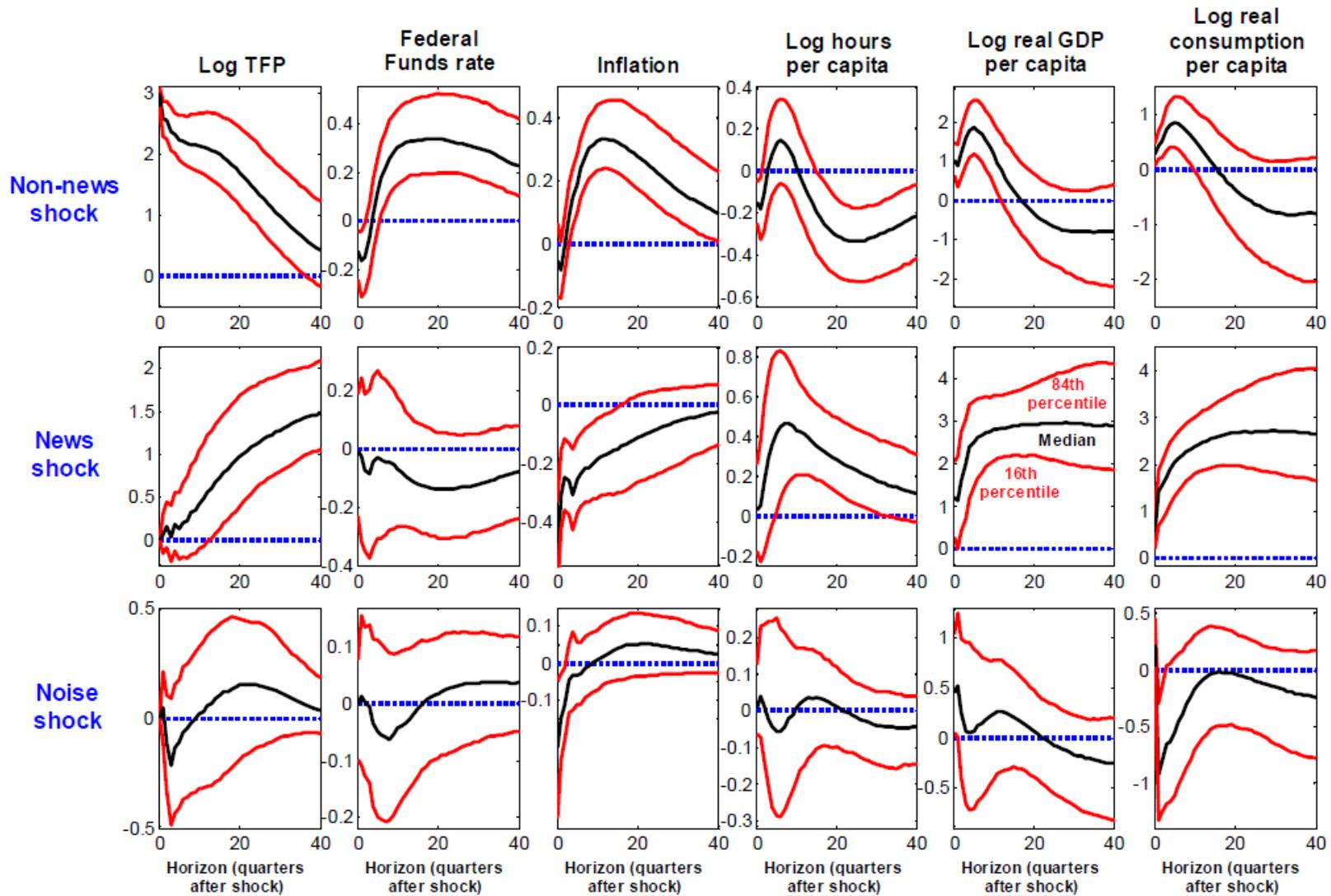


Figure III.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

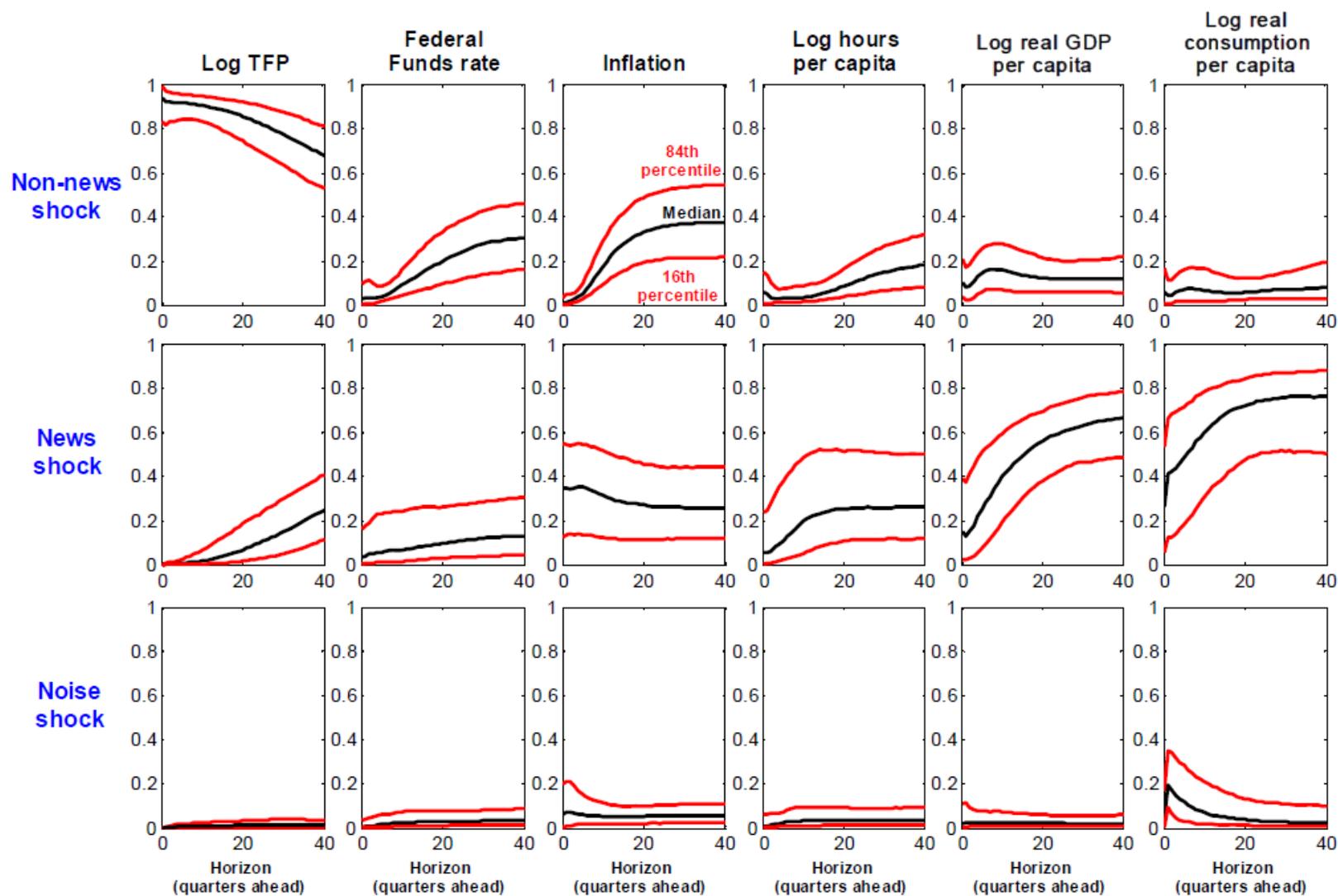


Figure III.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

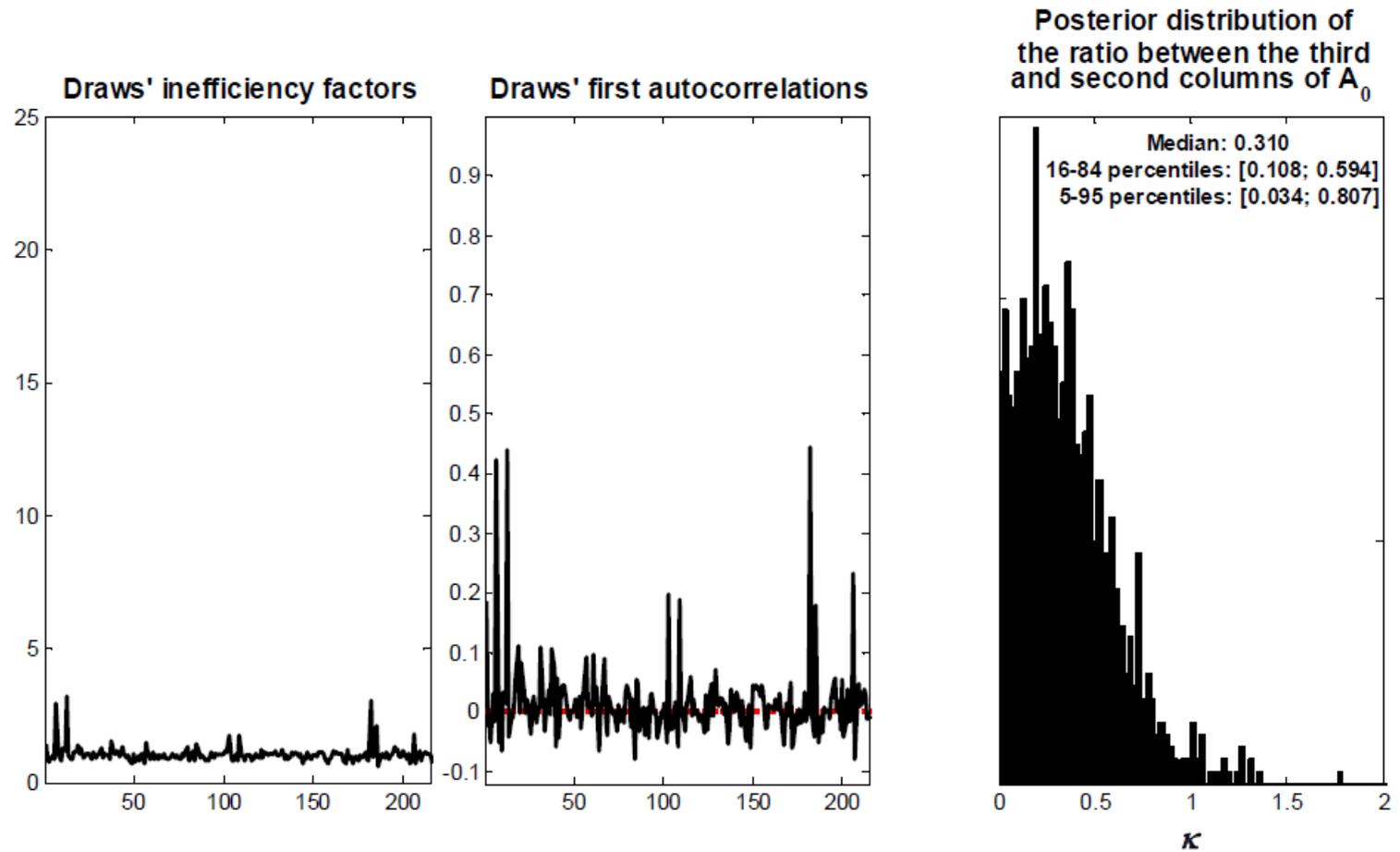


Figure III.4 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,2)

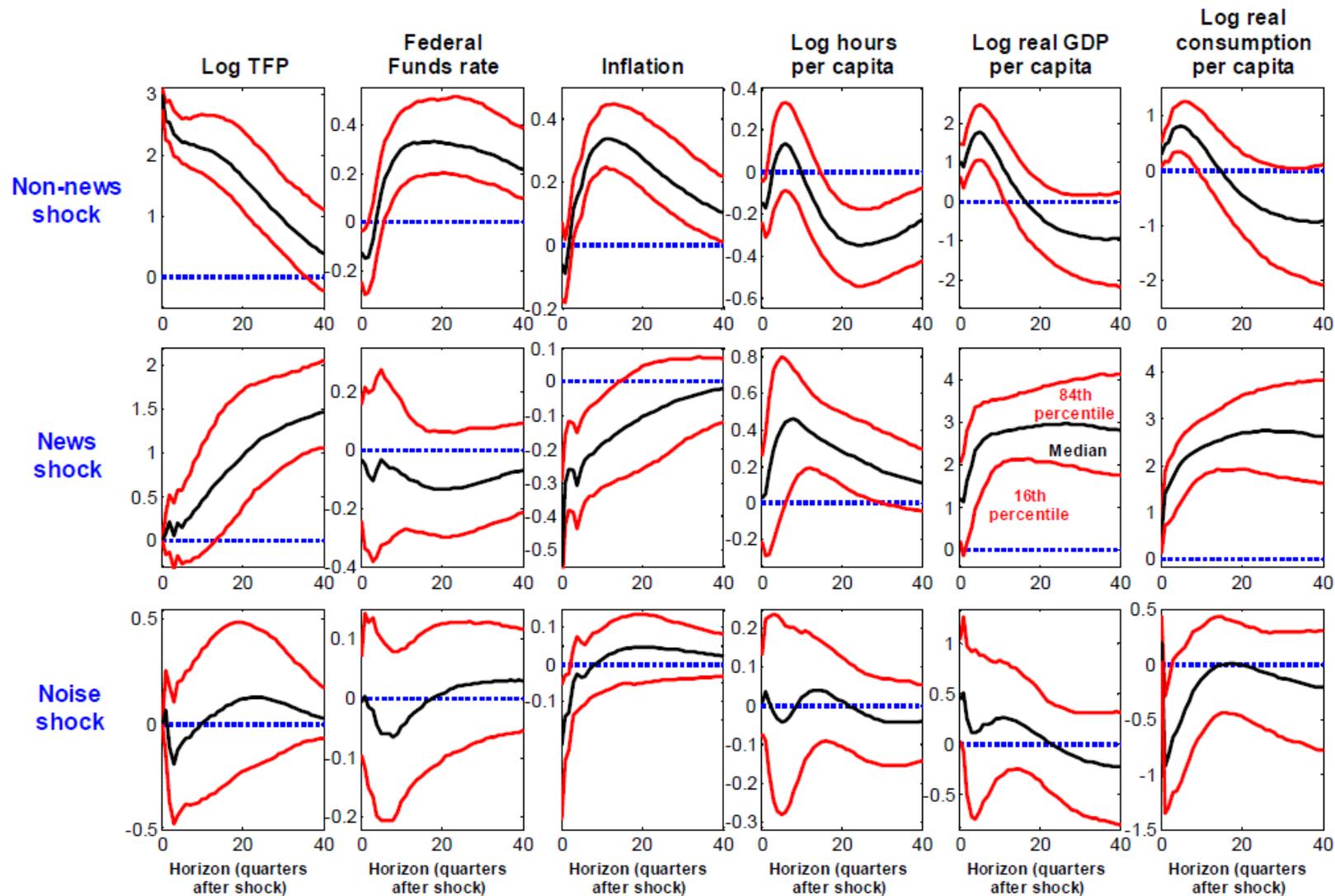


Figure III.5 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,2)

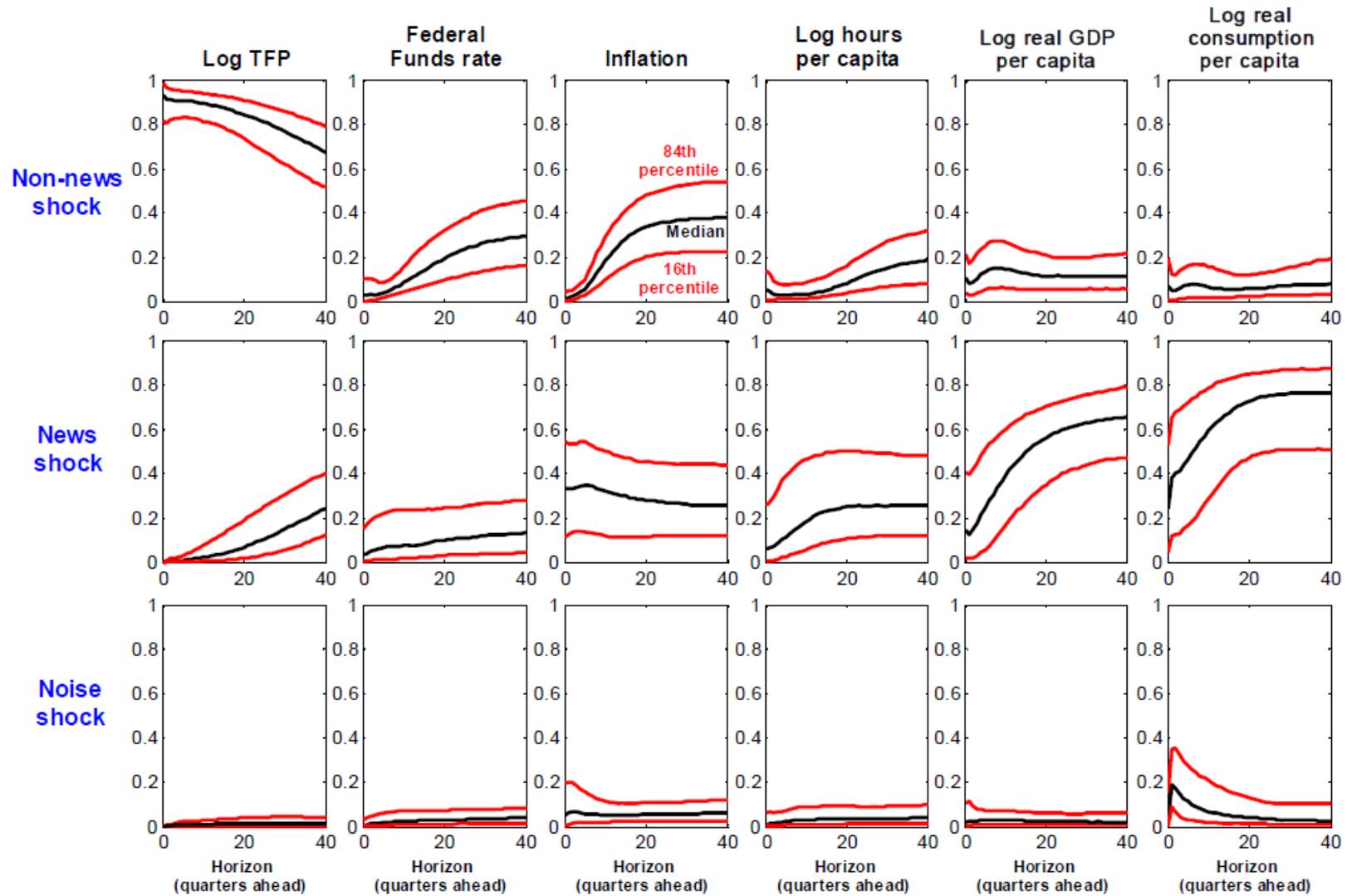


Figure III.6 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,2)

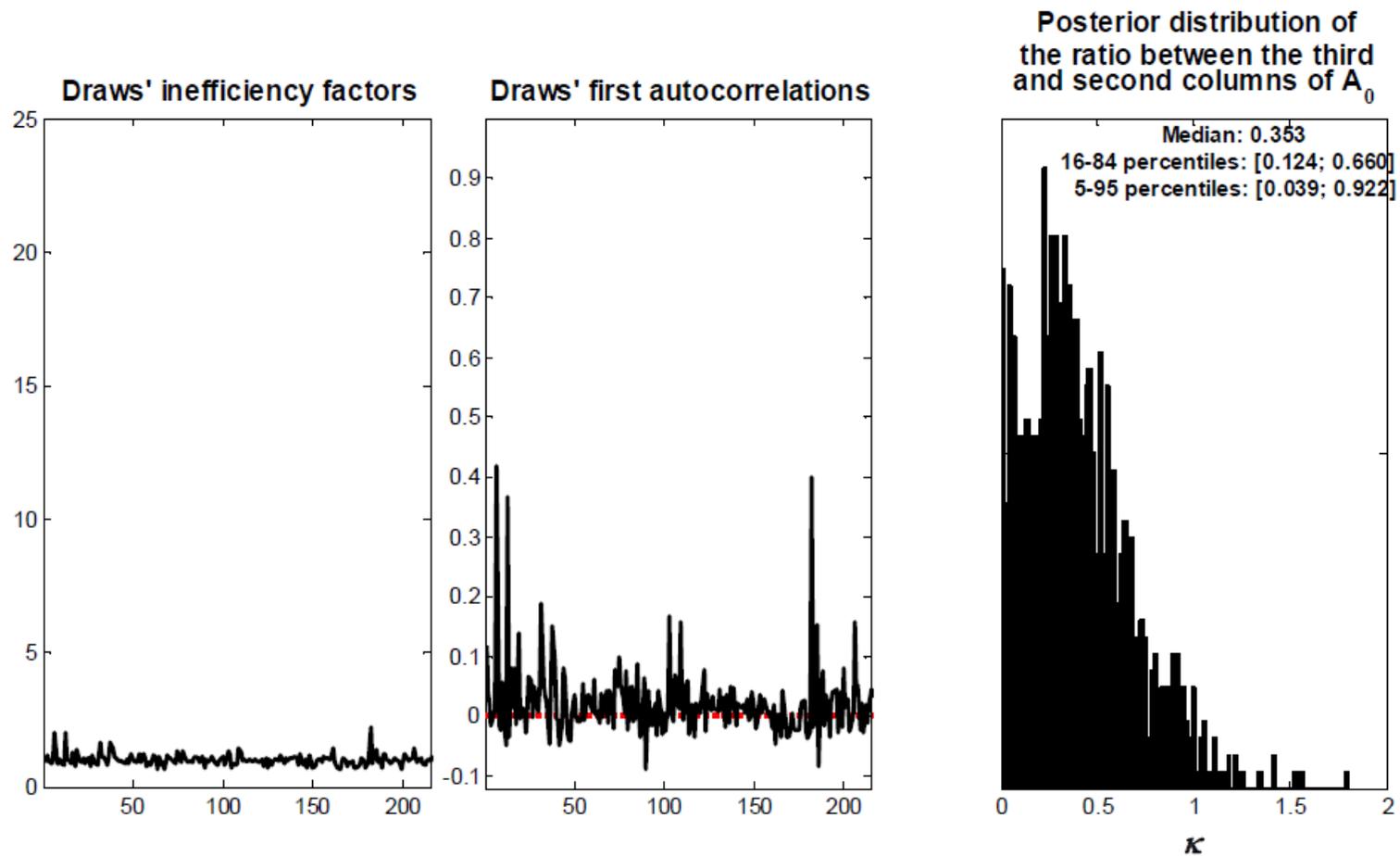


Figure III.7 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,3)

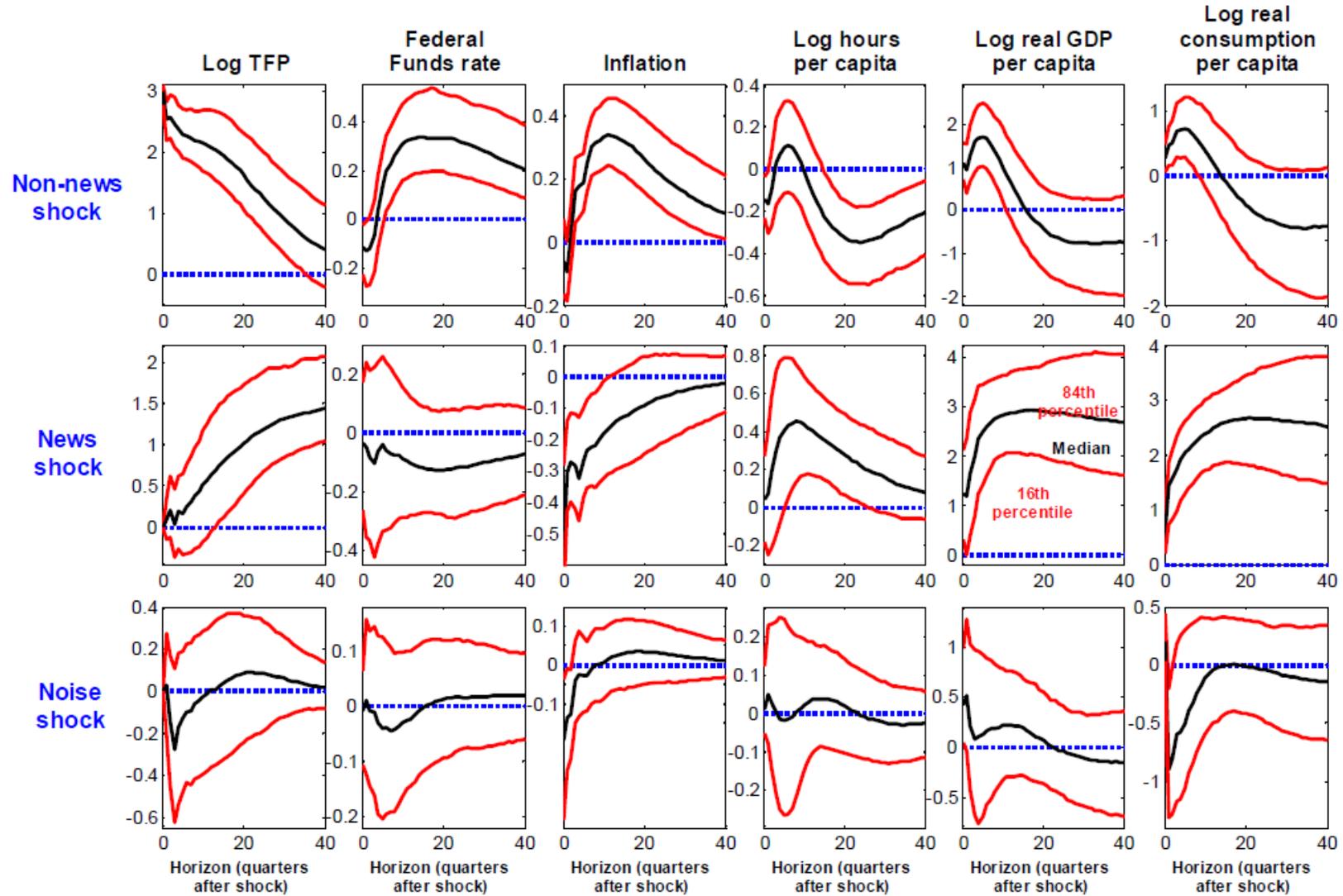


Figure III.8 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,3)

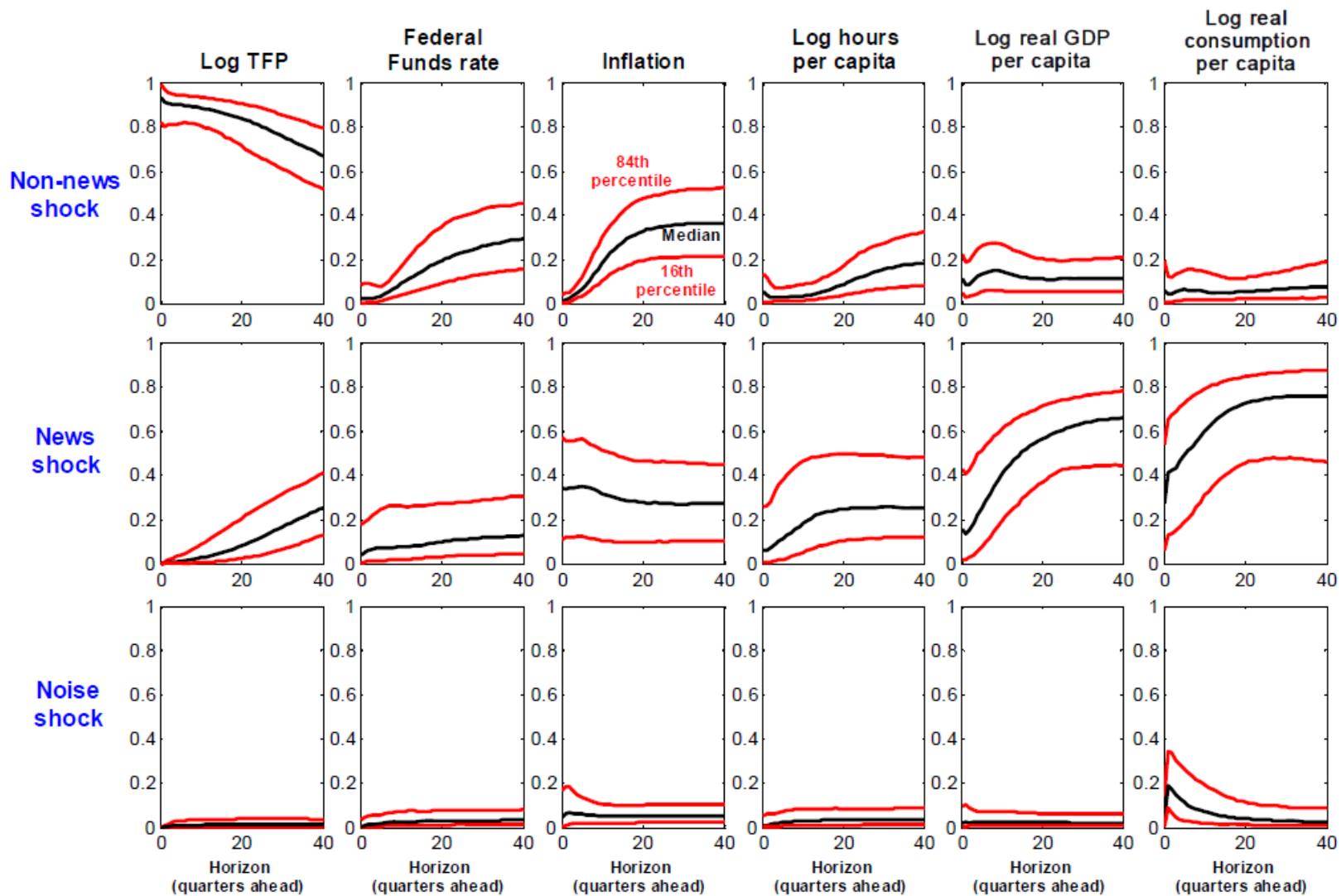


Figure III.9 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,3)

Results based on VARMA with 6 series and set identification, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

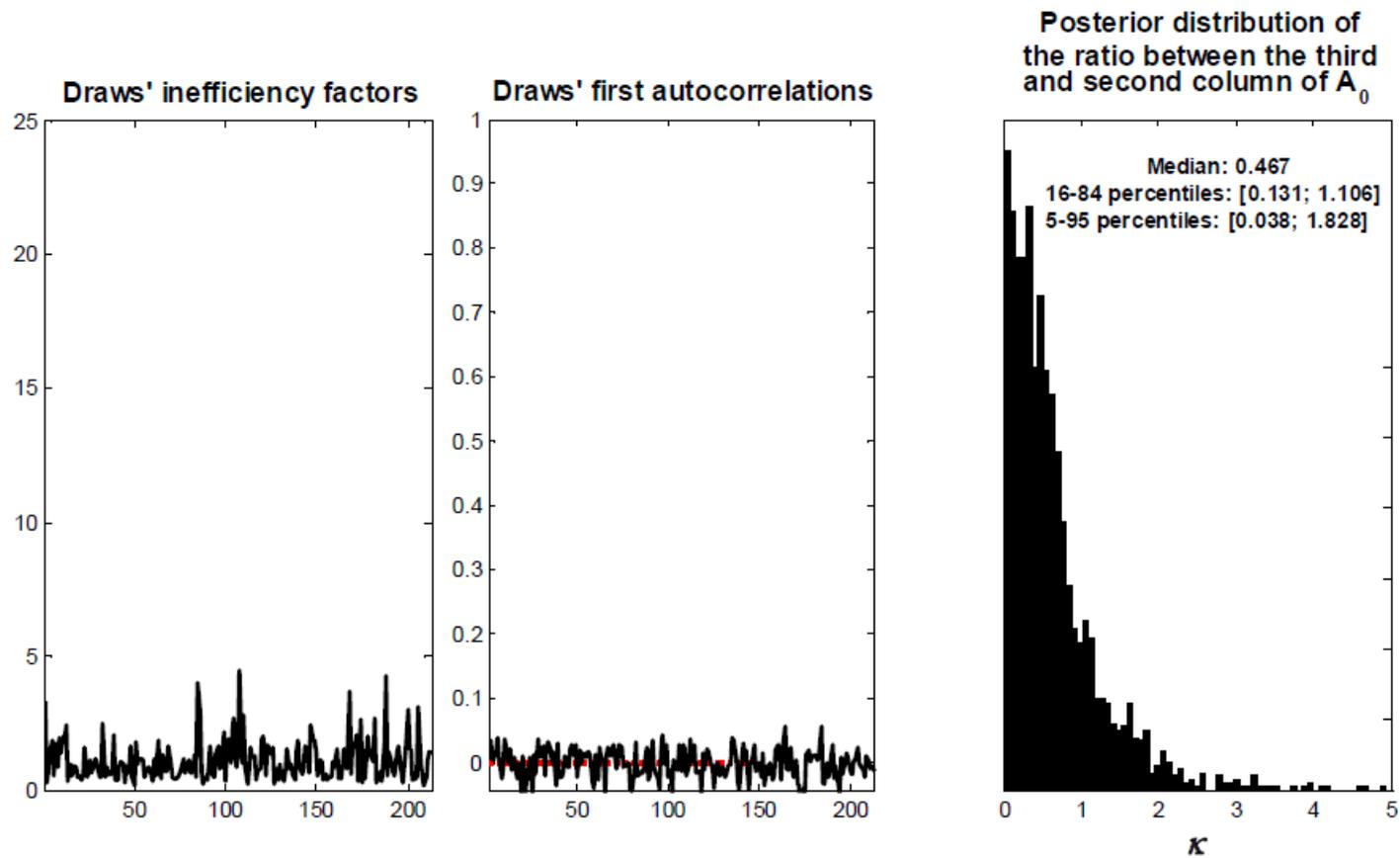


Figure IV.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

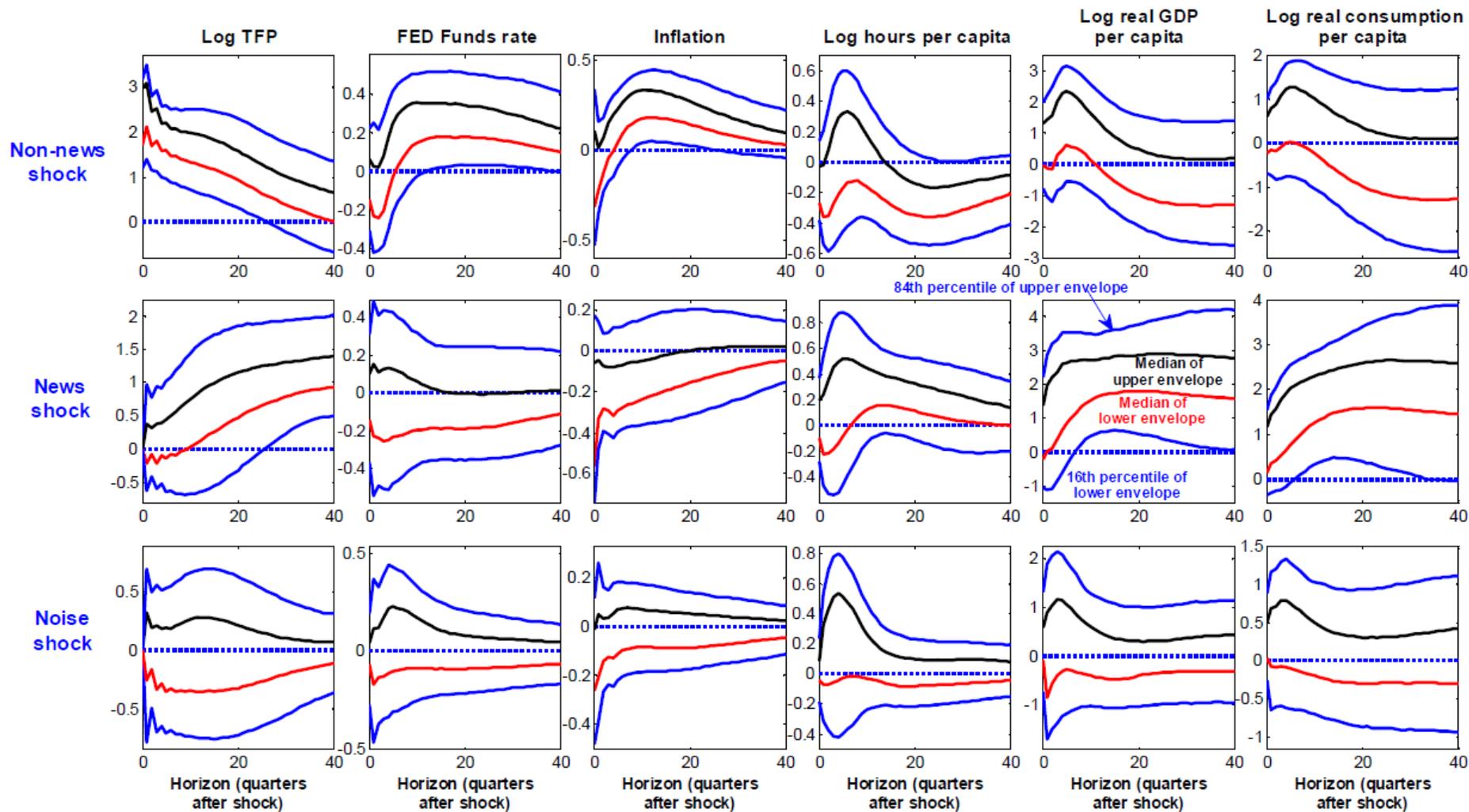


Figure IV.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

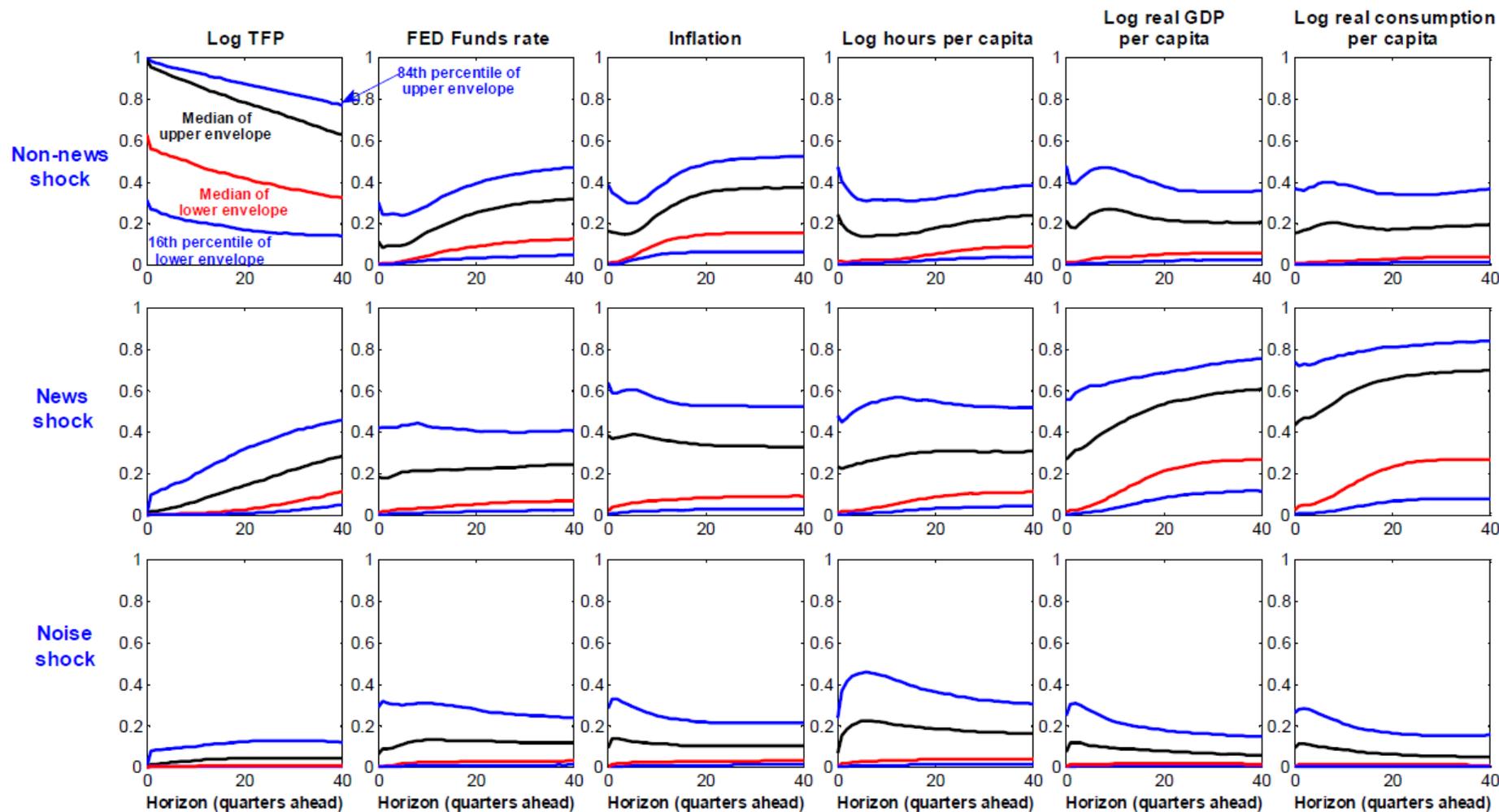


Figure IV.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

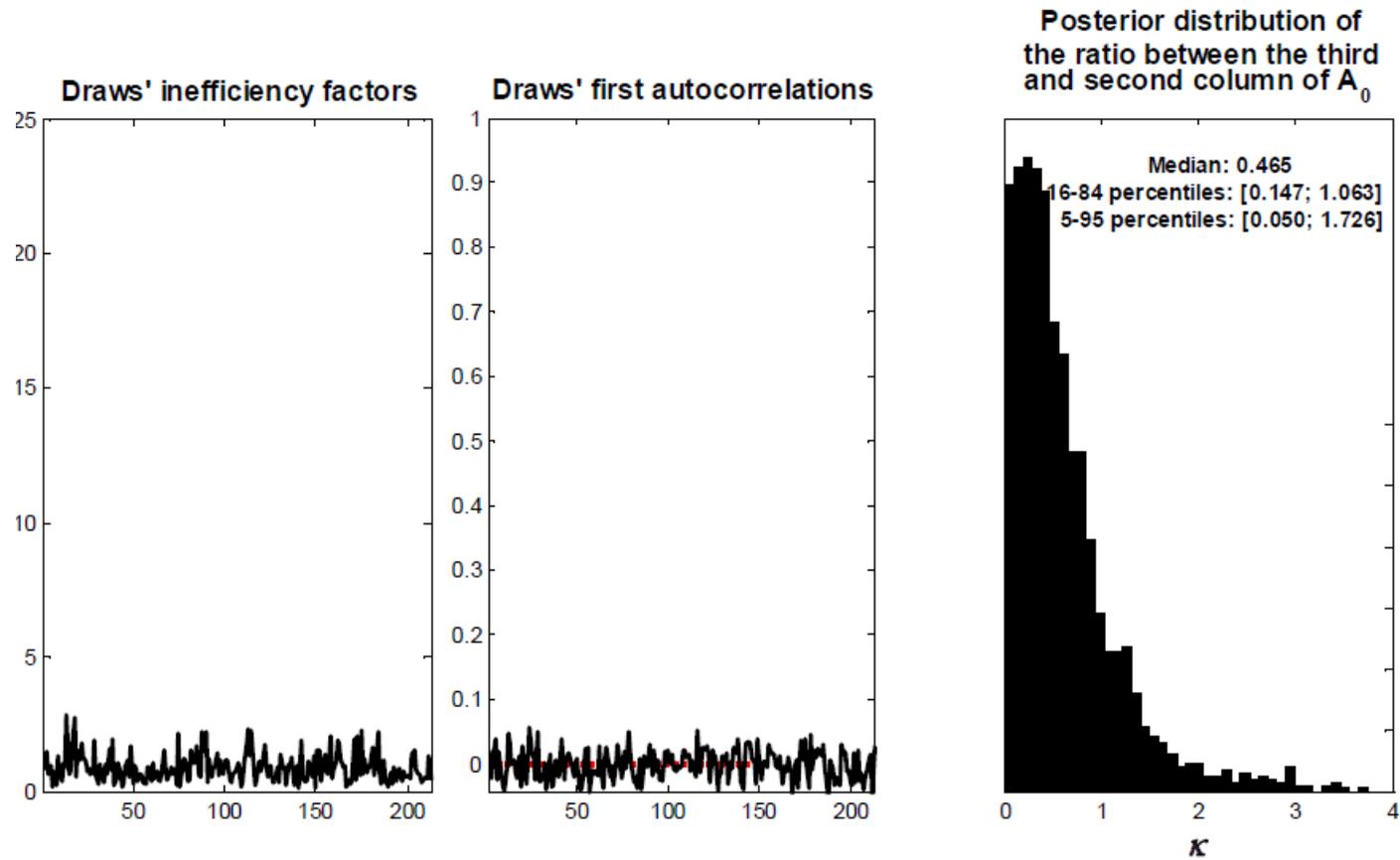


Figure IV.4 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,2)

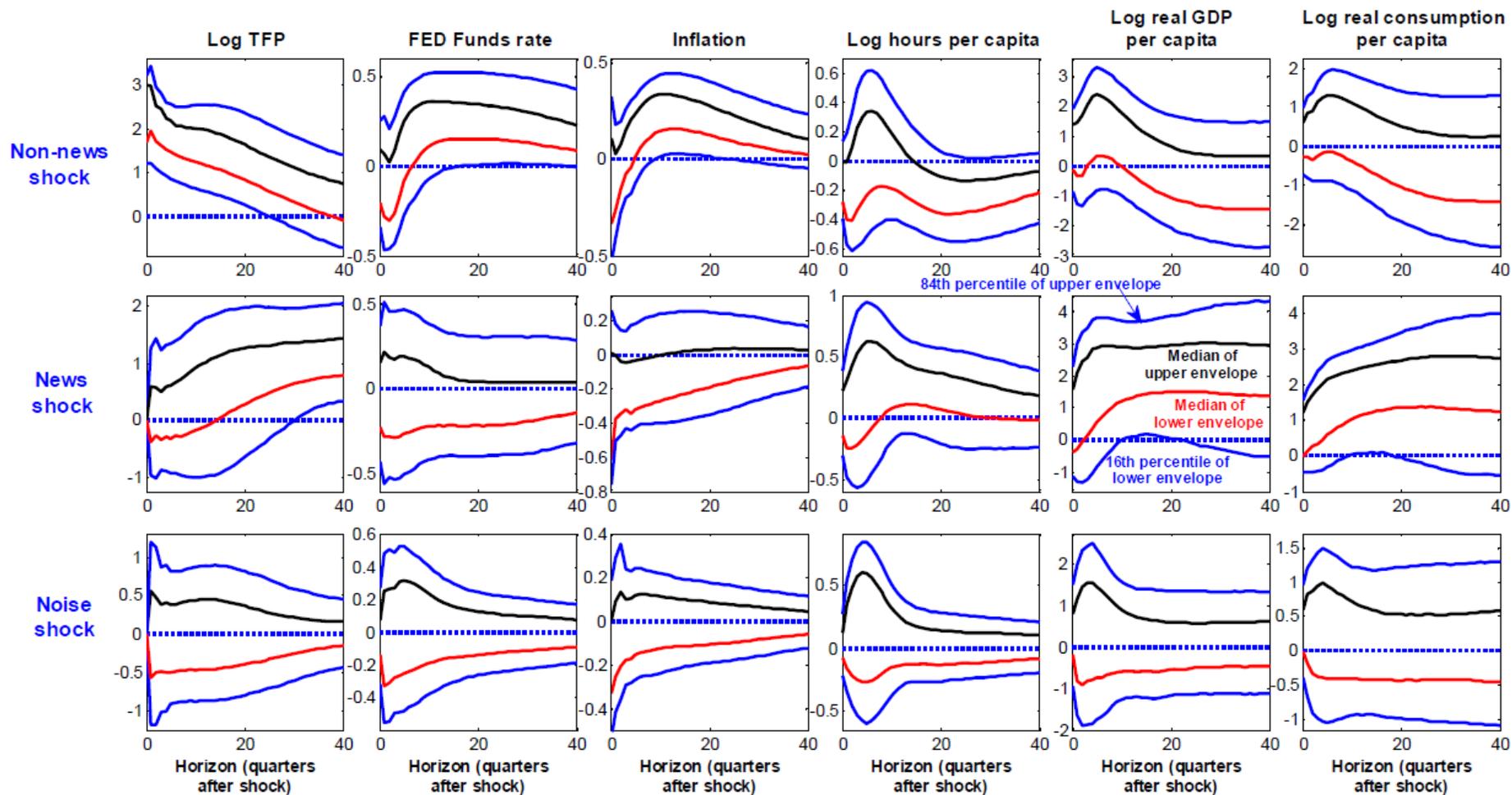


Figure IV.5 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,2)

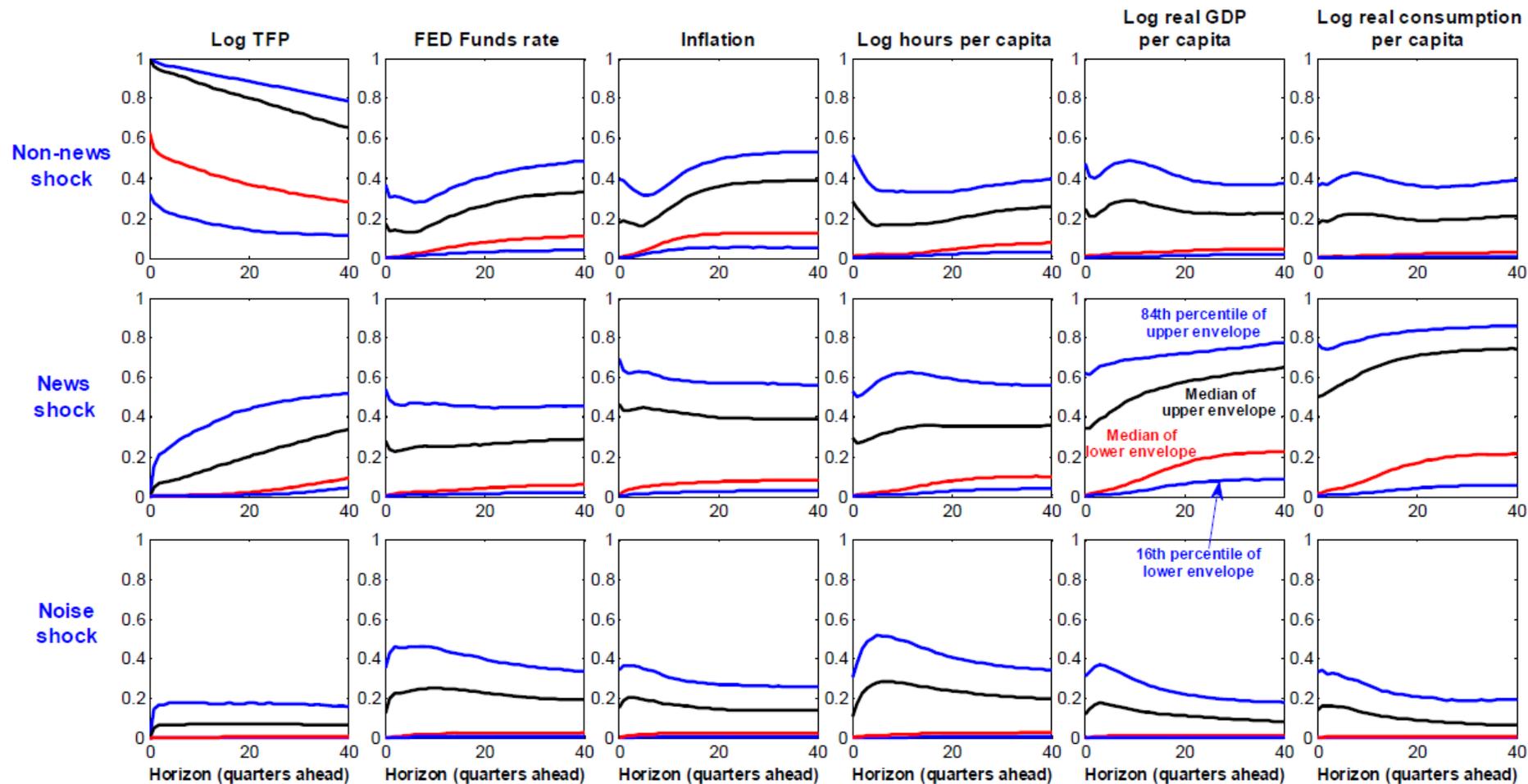


Figure IV.6 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,2)

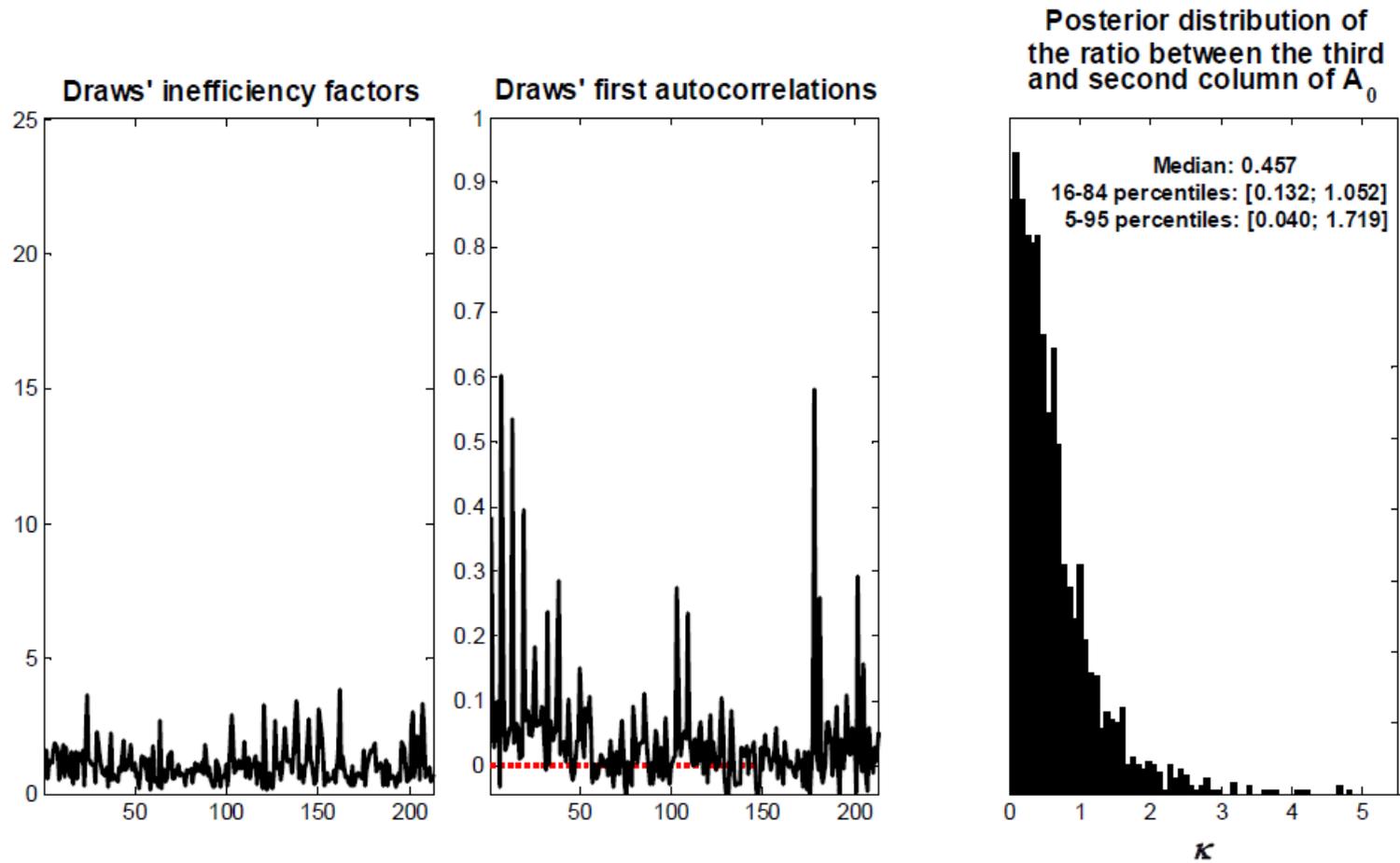


Figure IV.7 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,3)

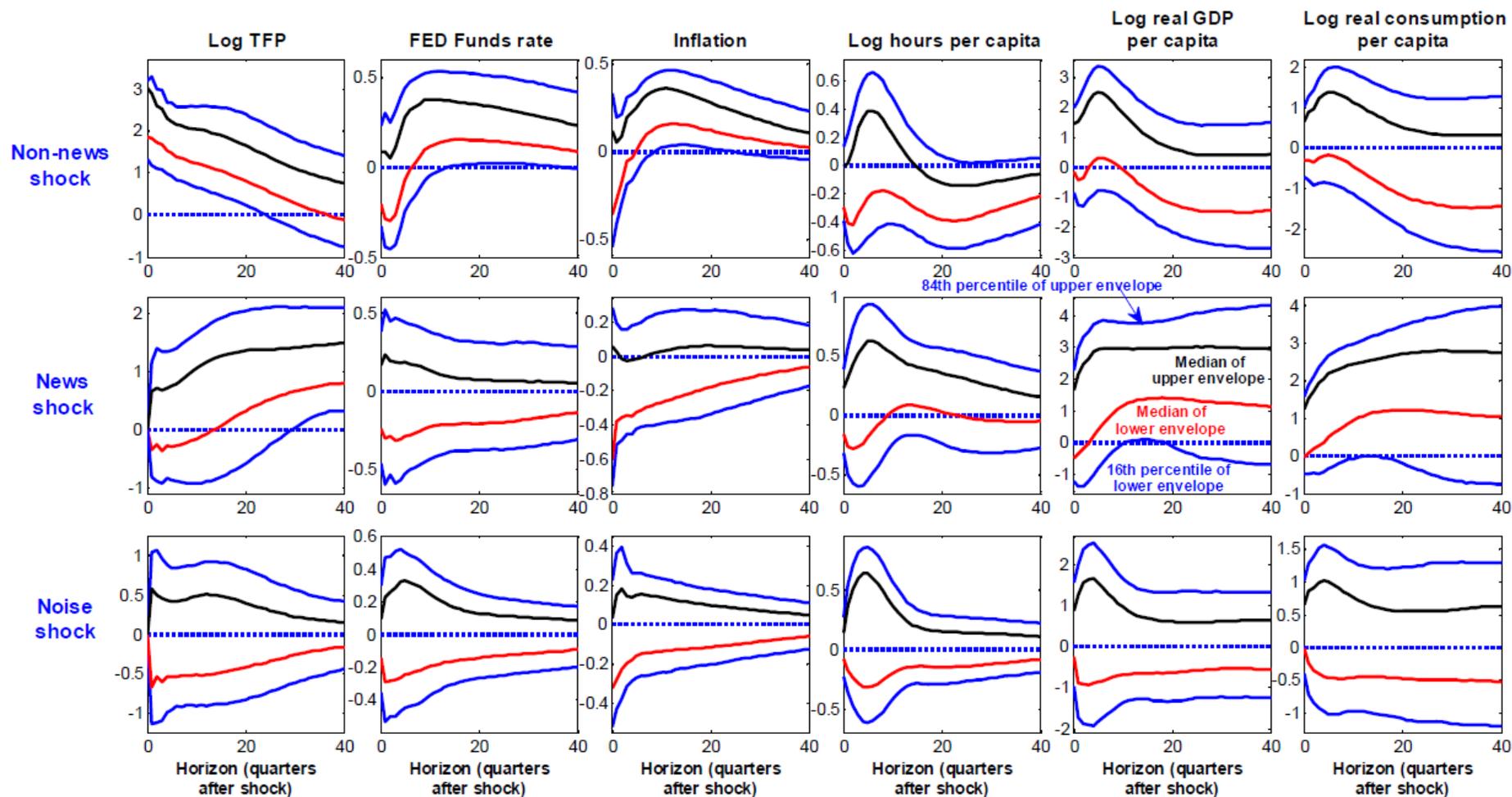


Figure IV.8 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,3)

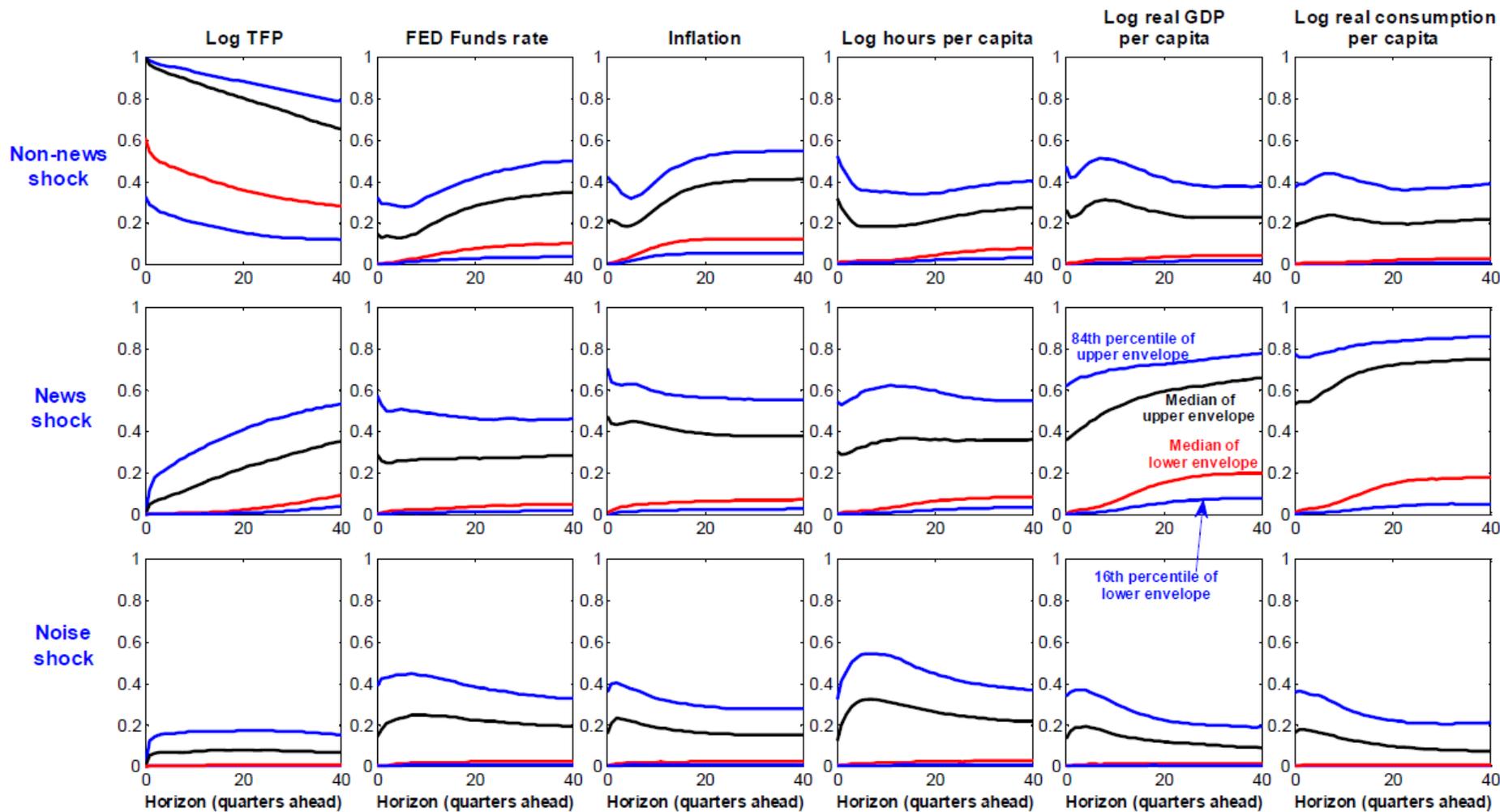


Figure IV.9 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,3)

Results based on VARMA with 8 series and point identification, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

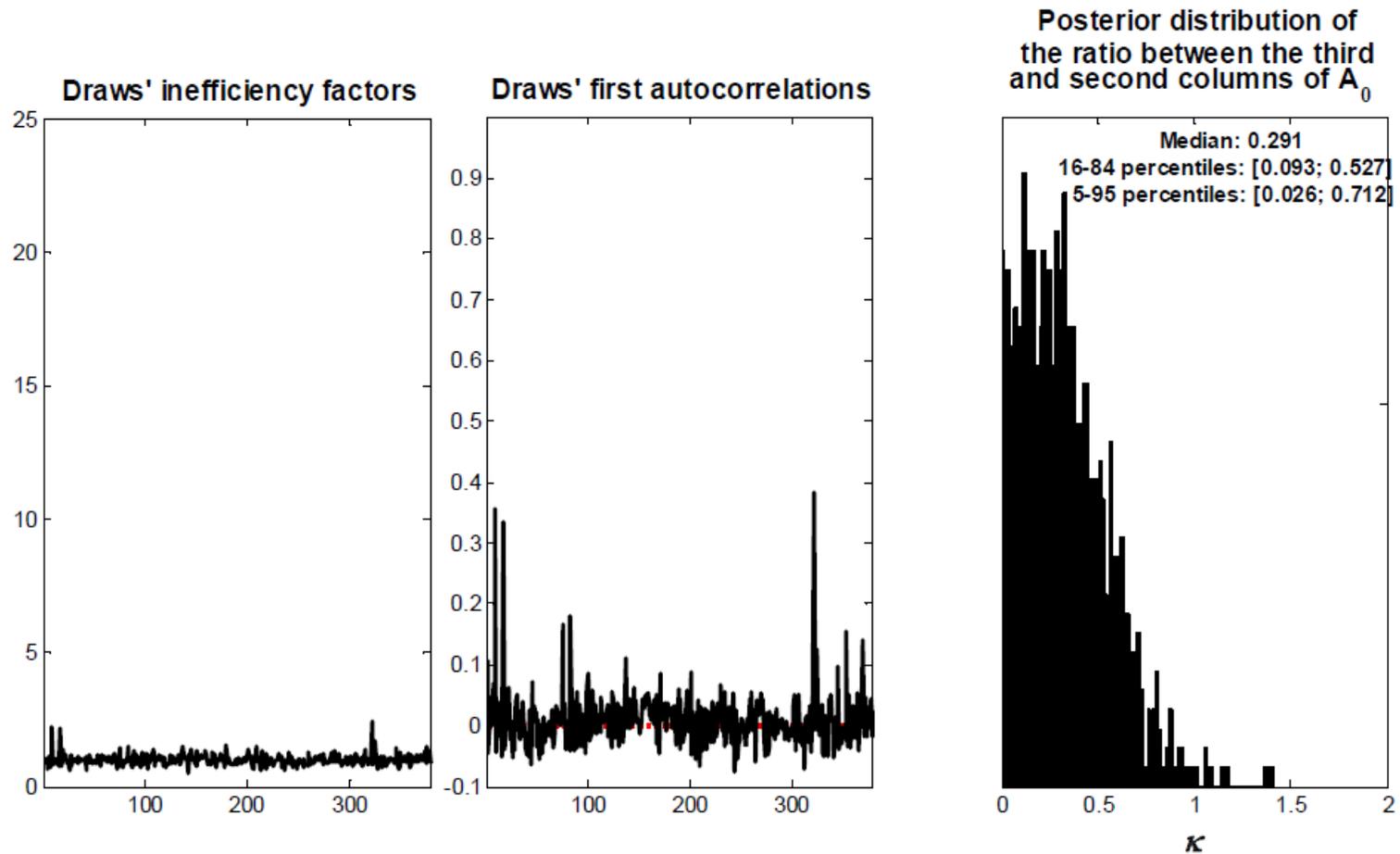


Figure V.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

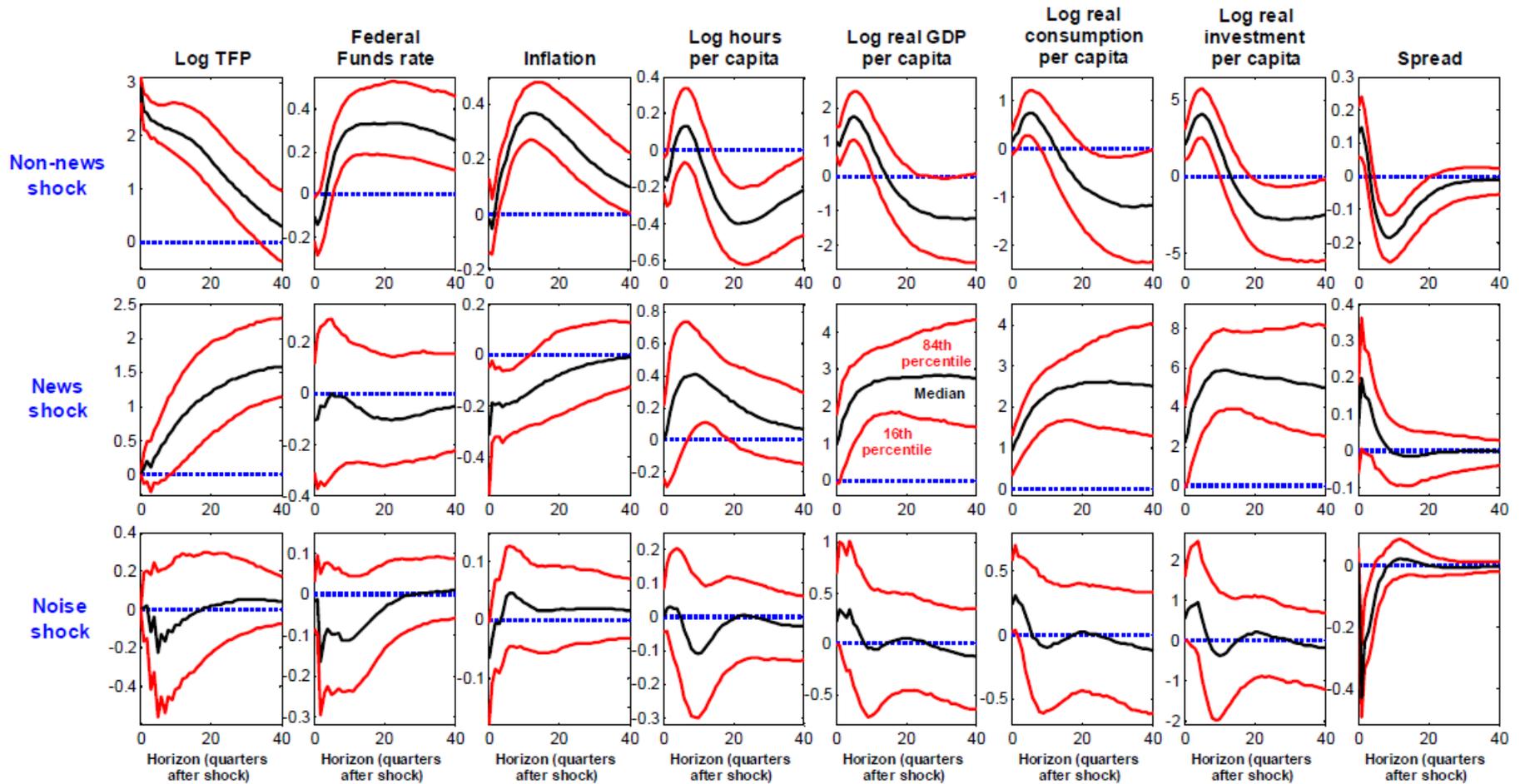


Figure V.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

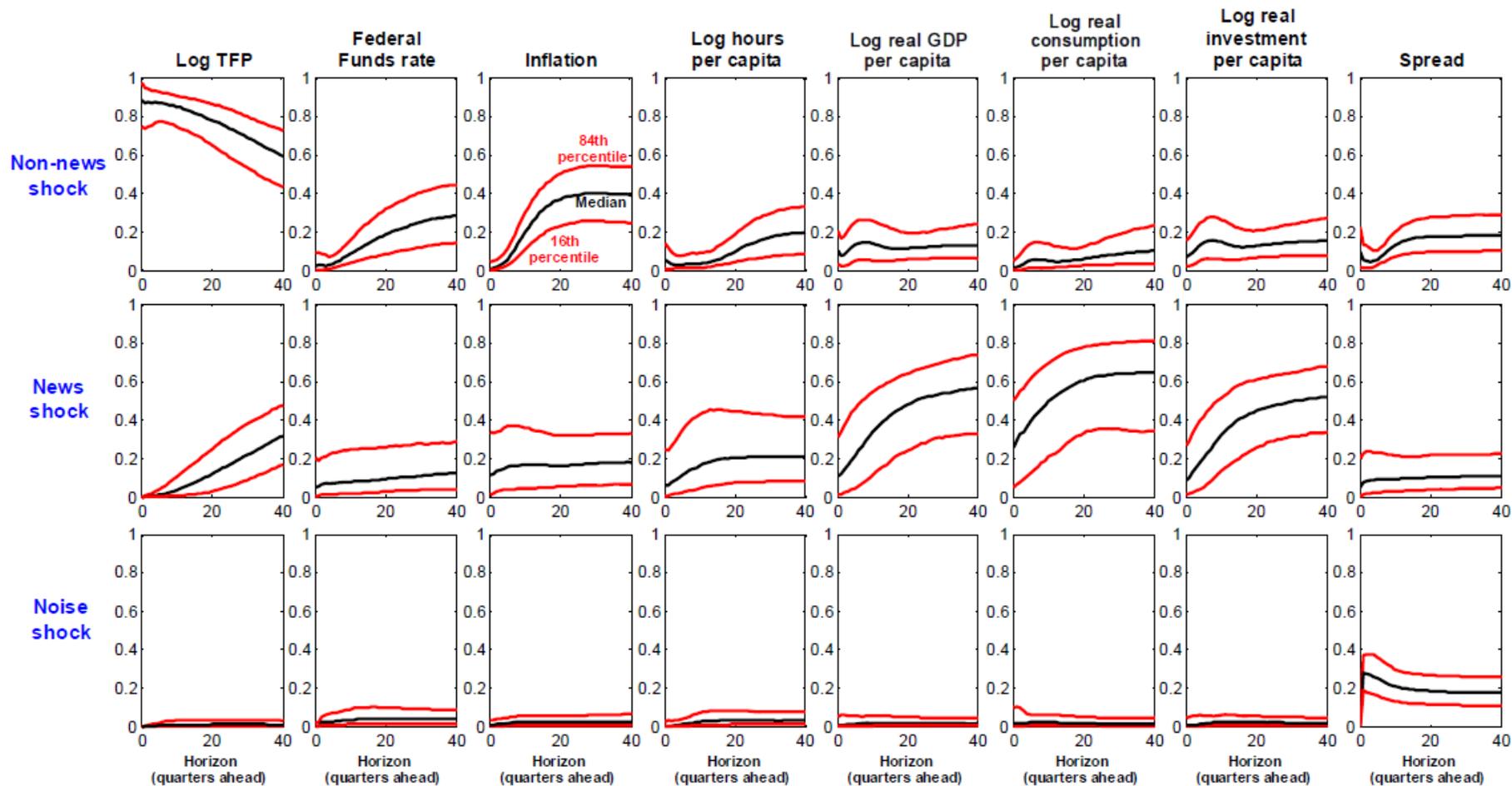


Figure V.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

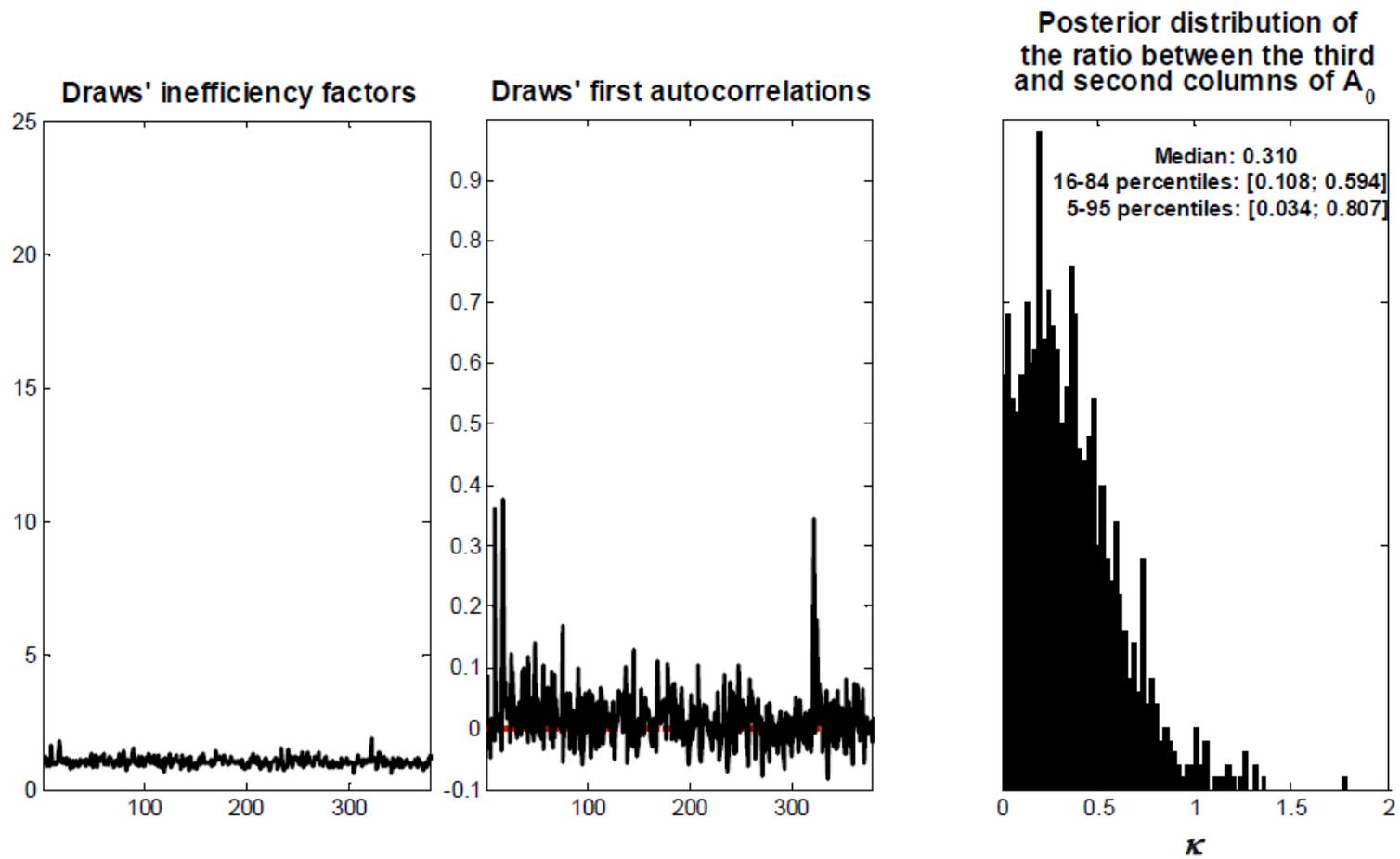


Figure V.4 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,2)

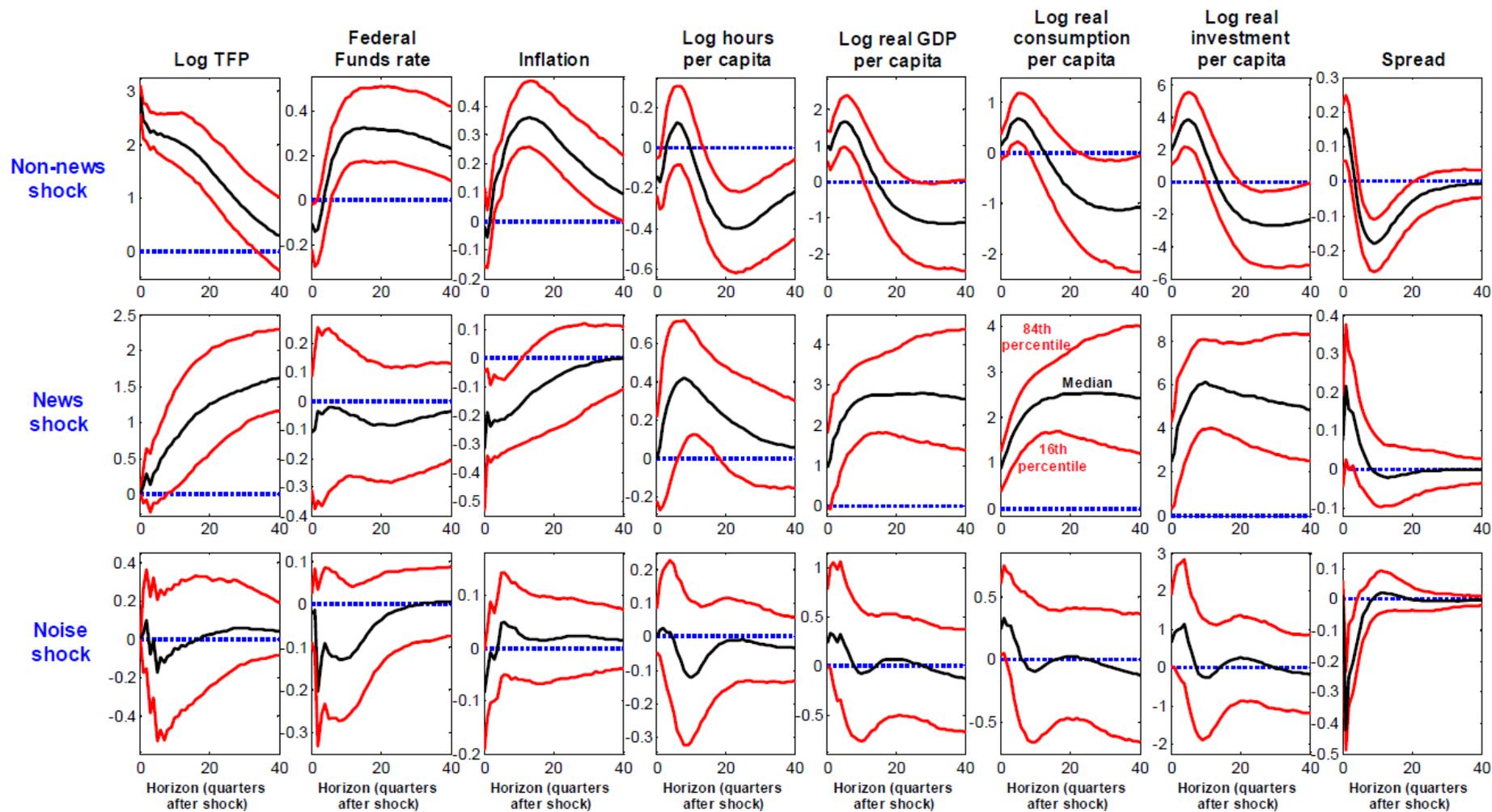


Figure V.5 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,2)

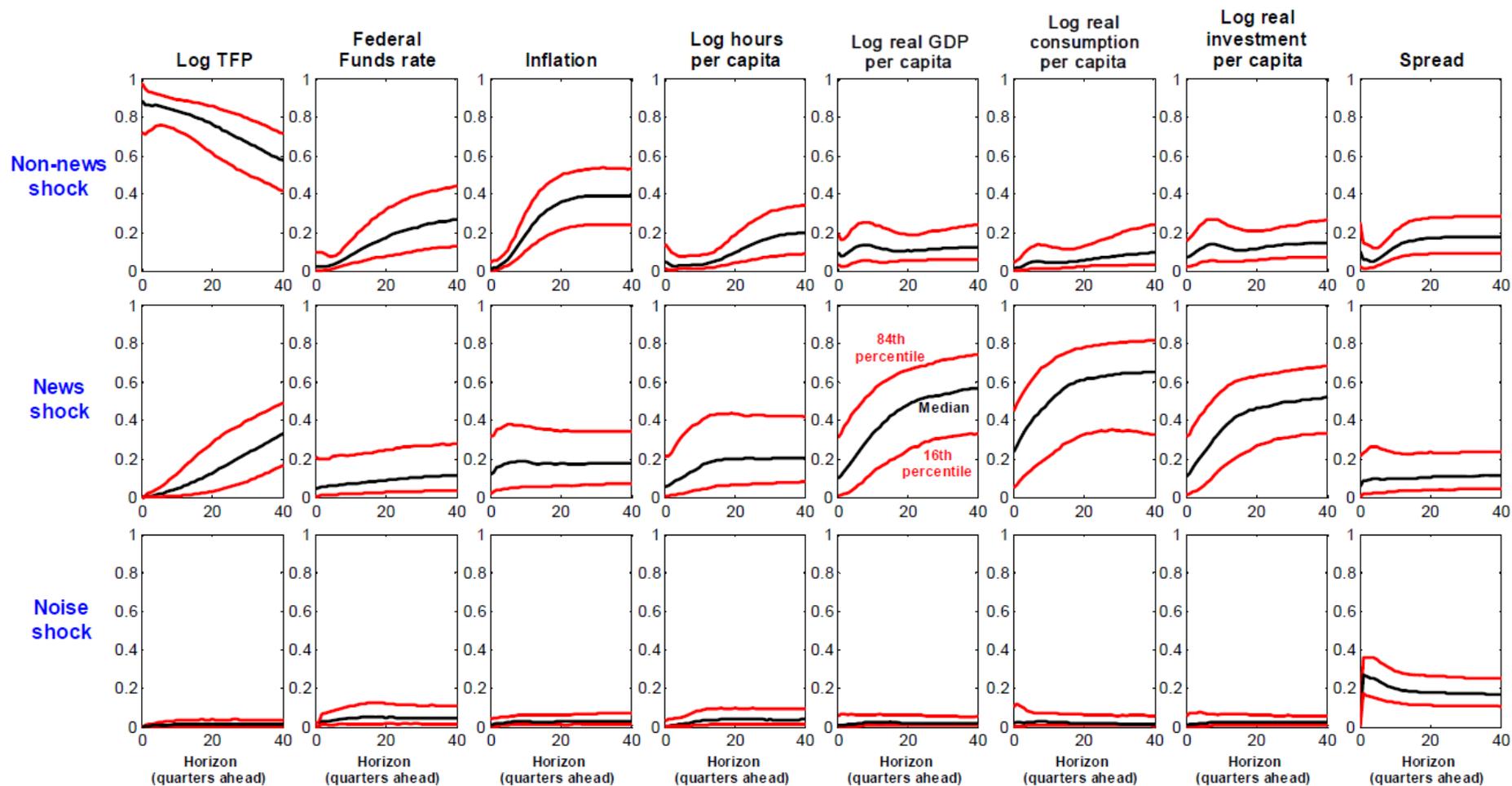


Figure V.6 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,2)

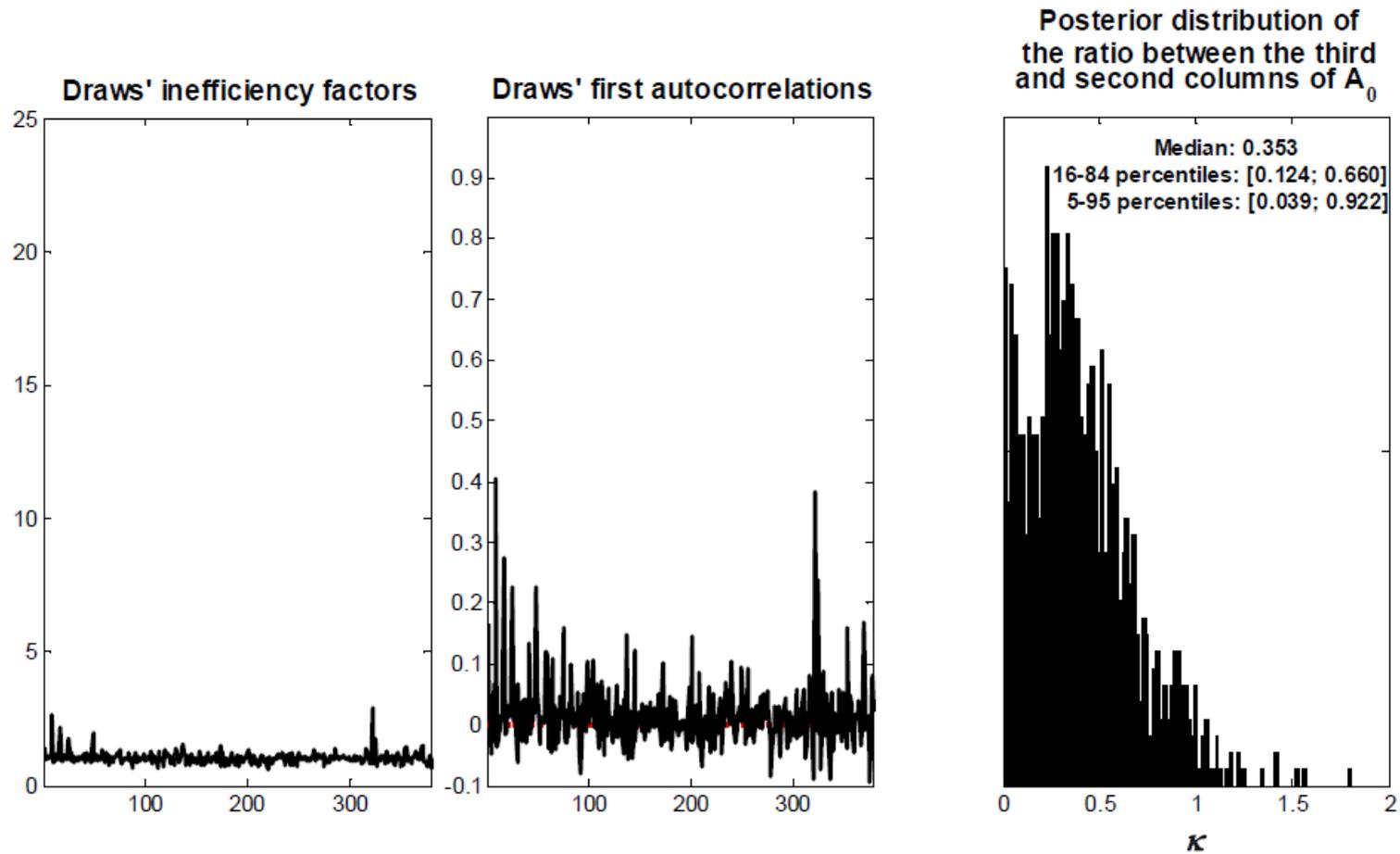


Figure V.7 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,3)

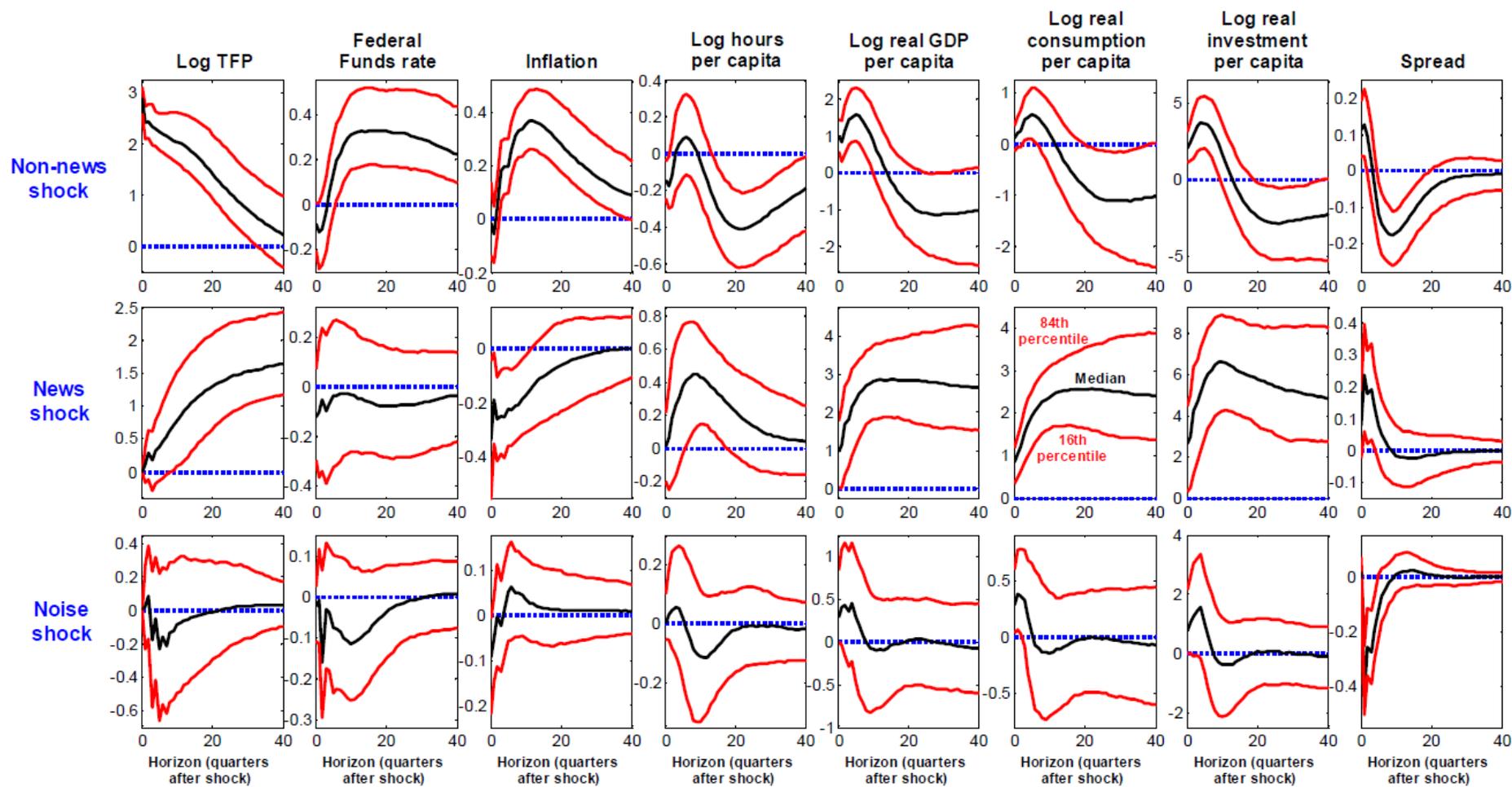


Figure V.8 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,3)

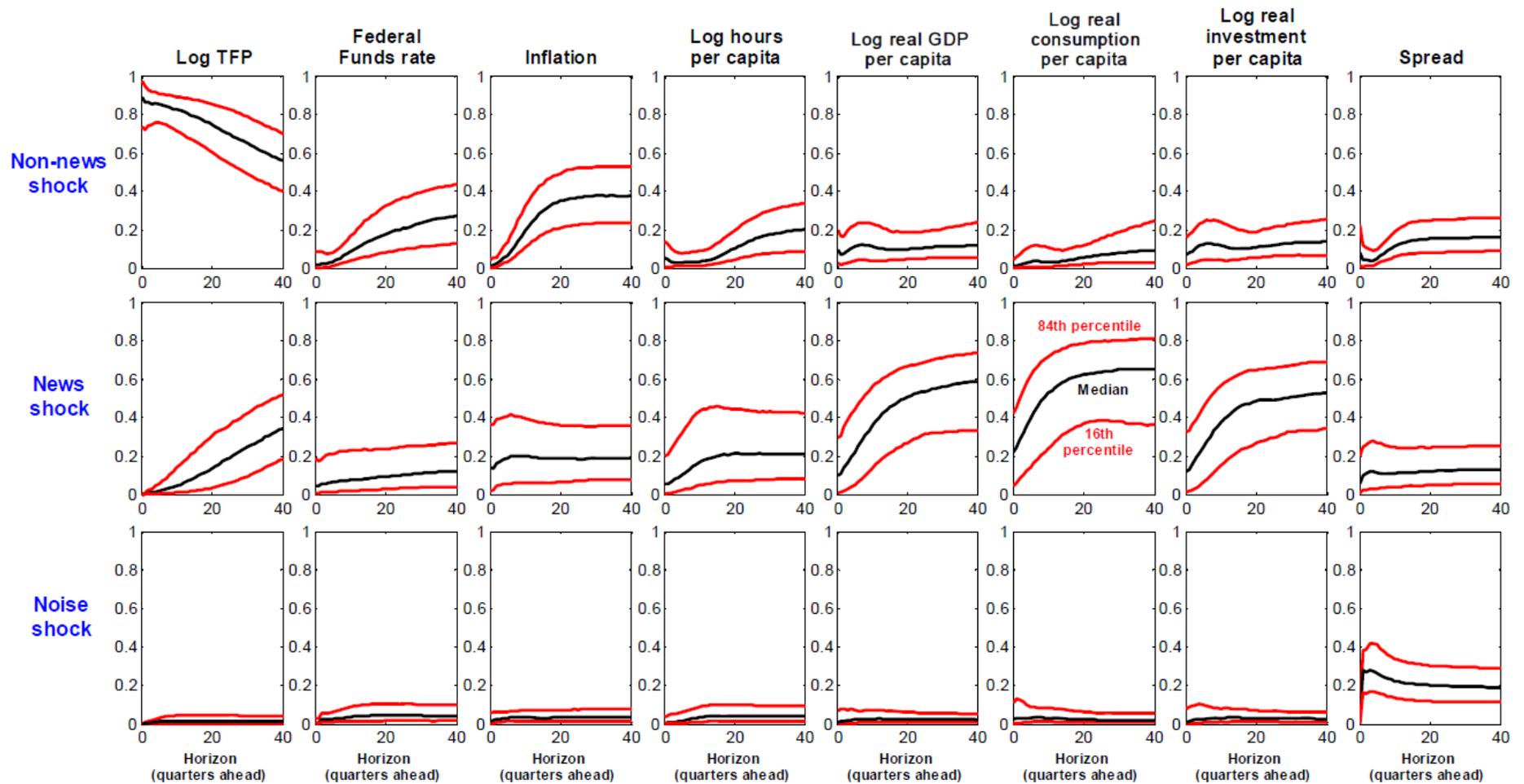


Figure V.9 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,3)

Results based on VARMA with 8 series and set identification, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

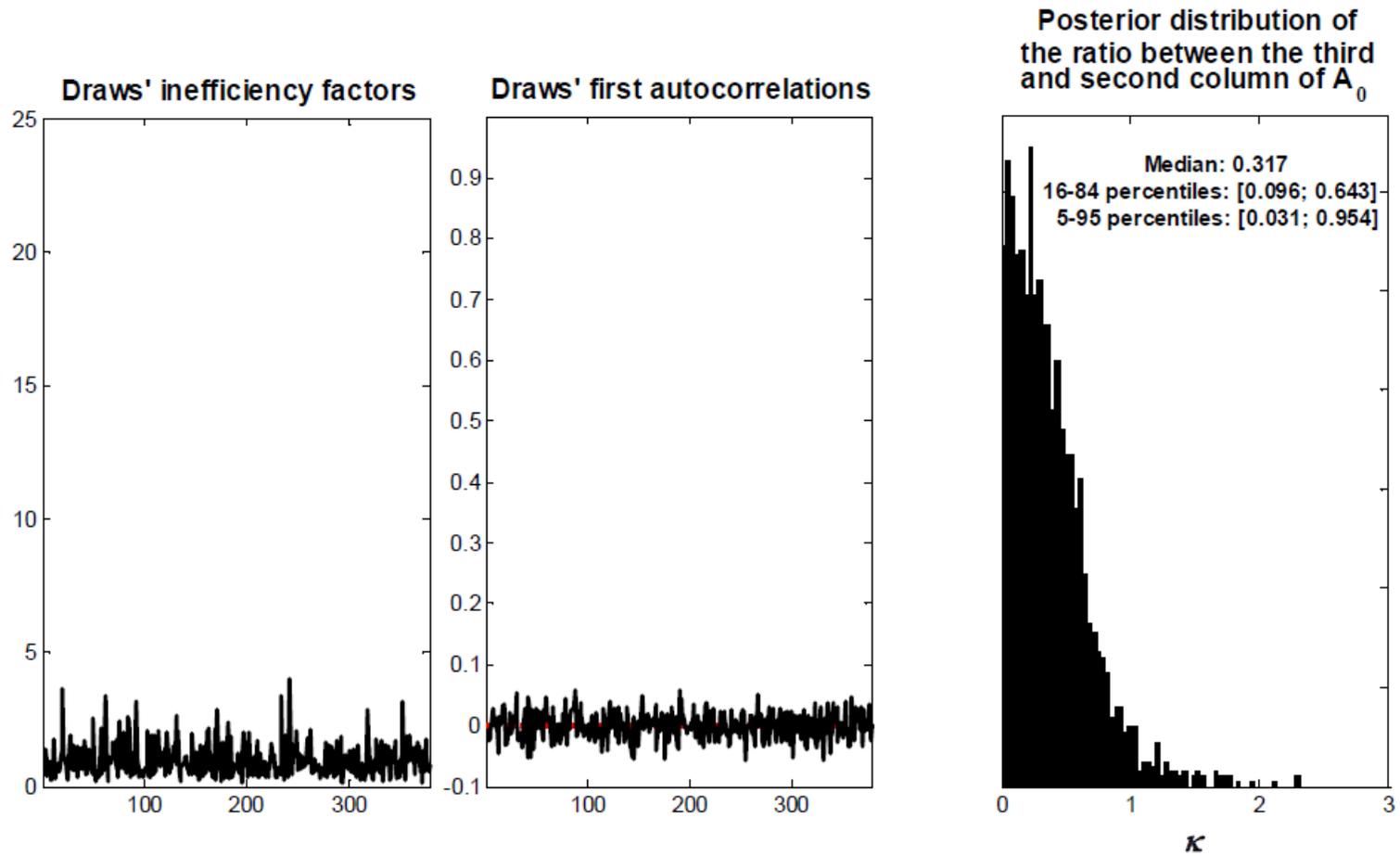


Figure VI.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

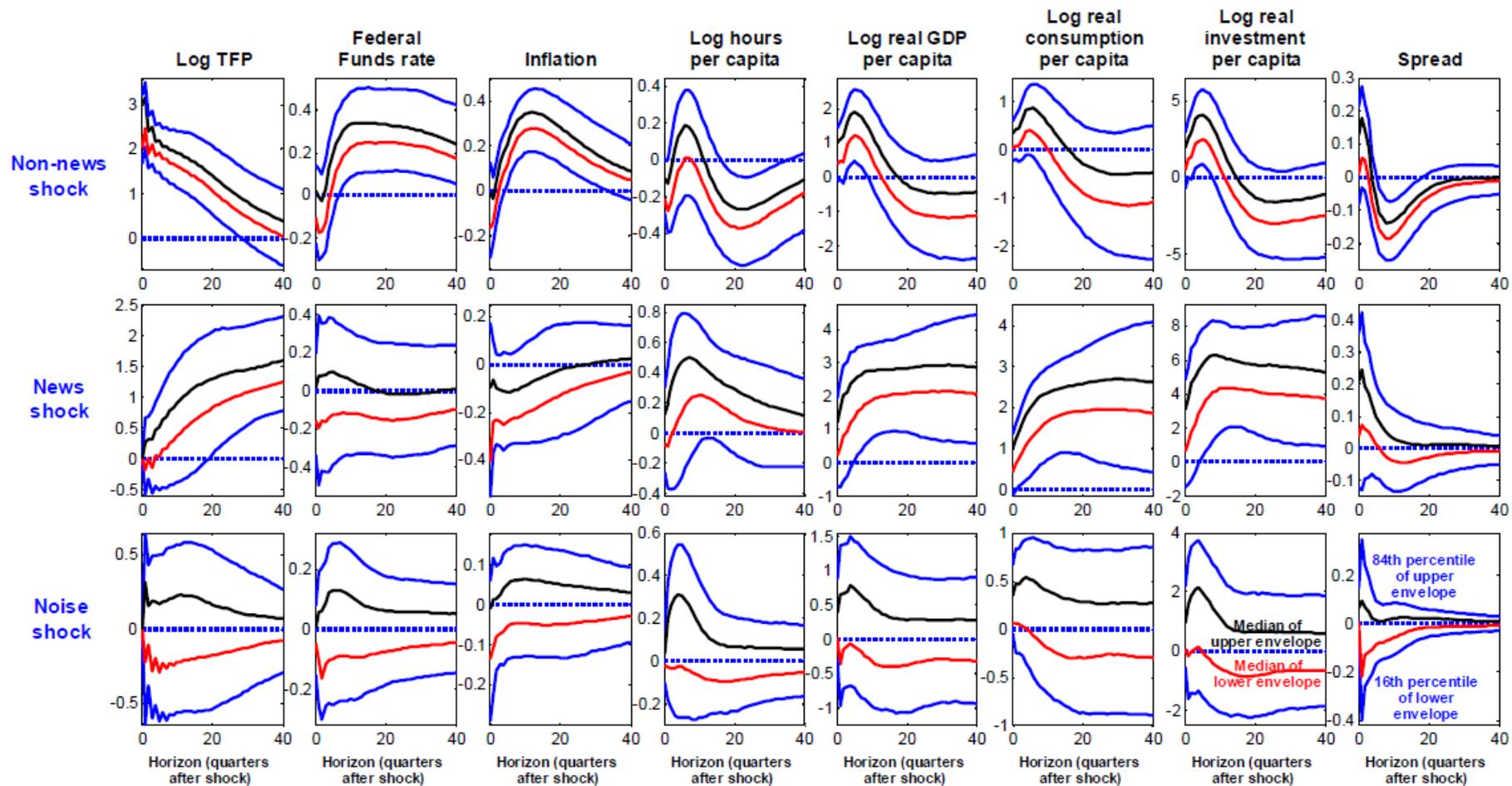


Figure VI.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

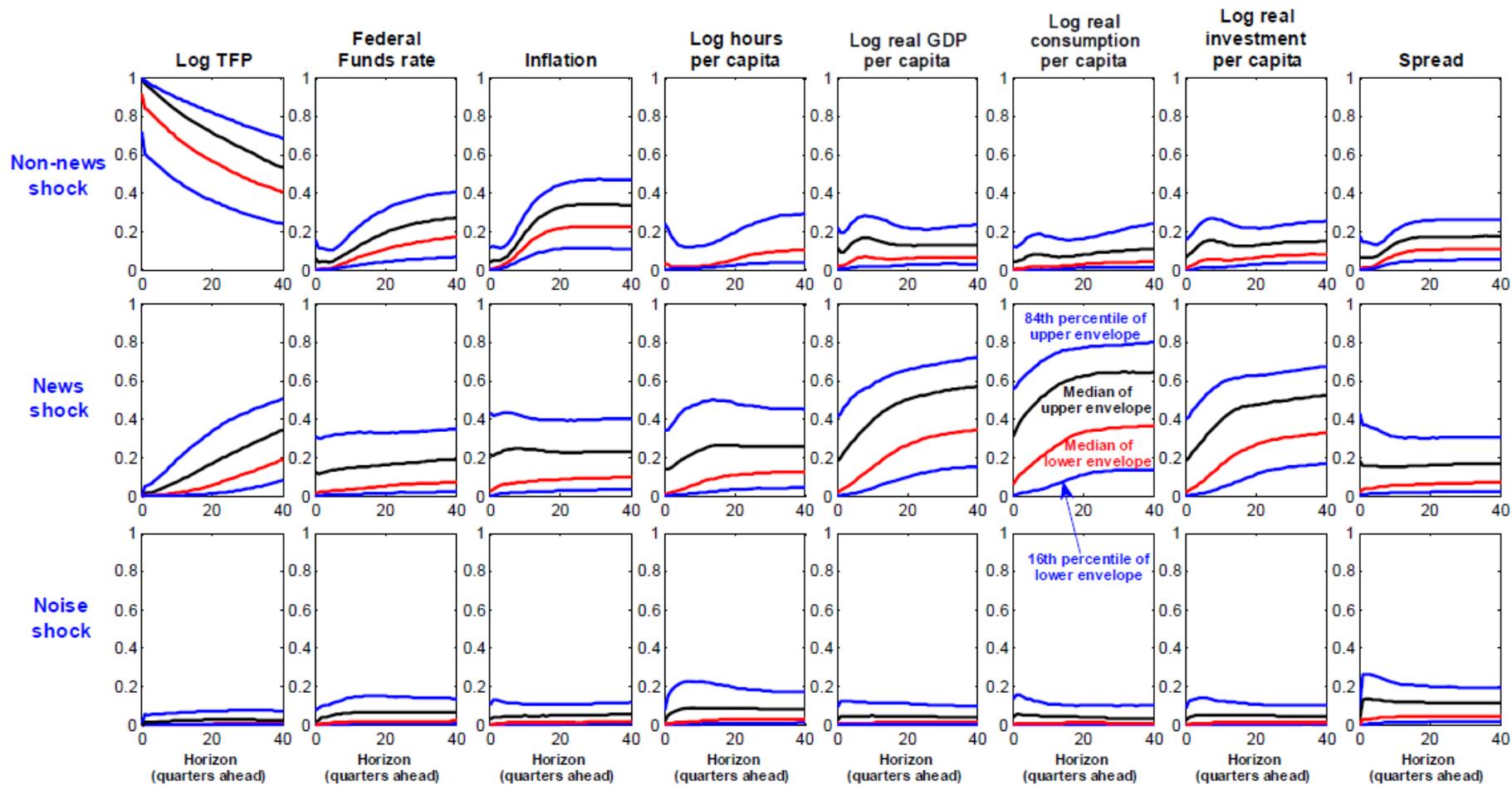


Figure VI.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

Results based on VARMA with 8 series and set identification, imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

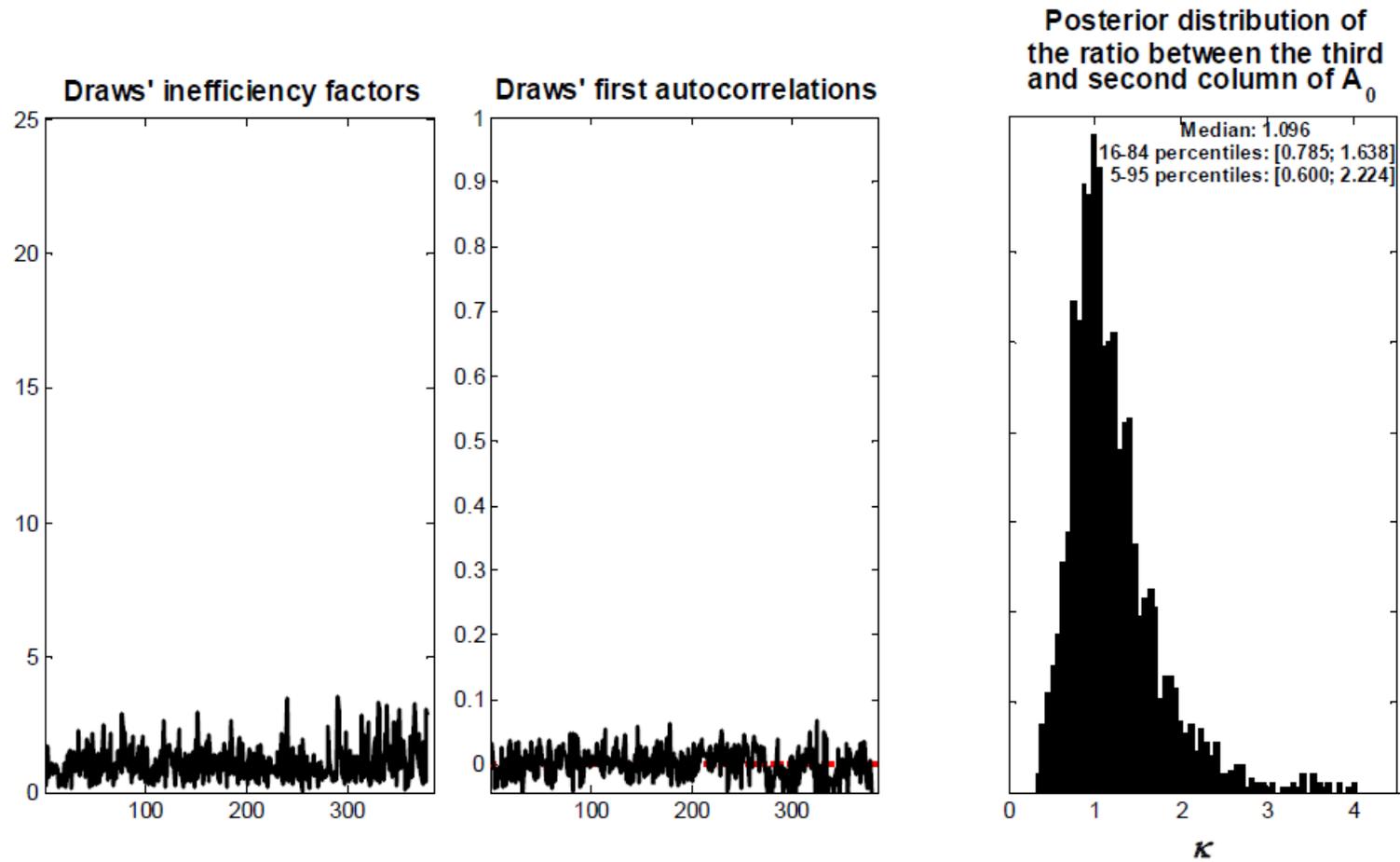


Figure VII.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

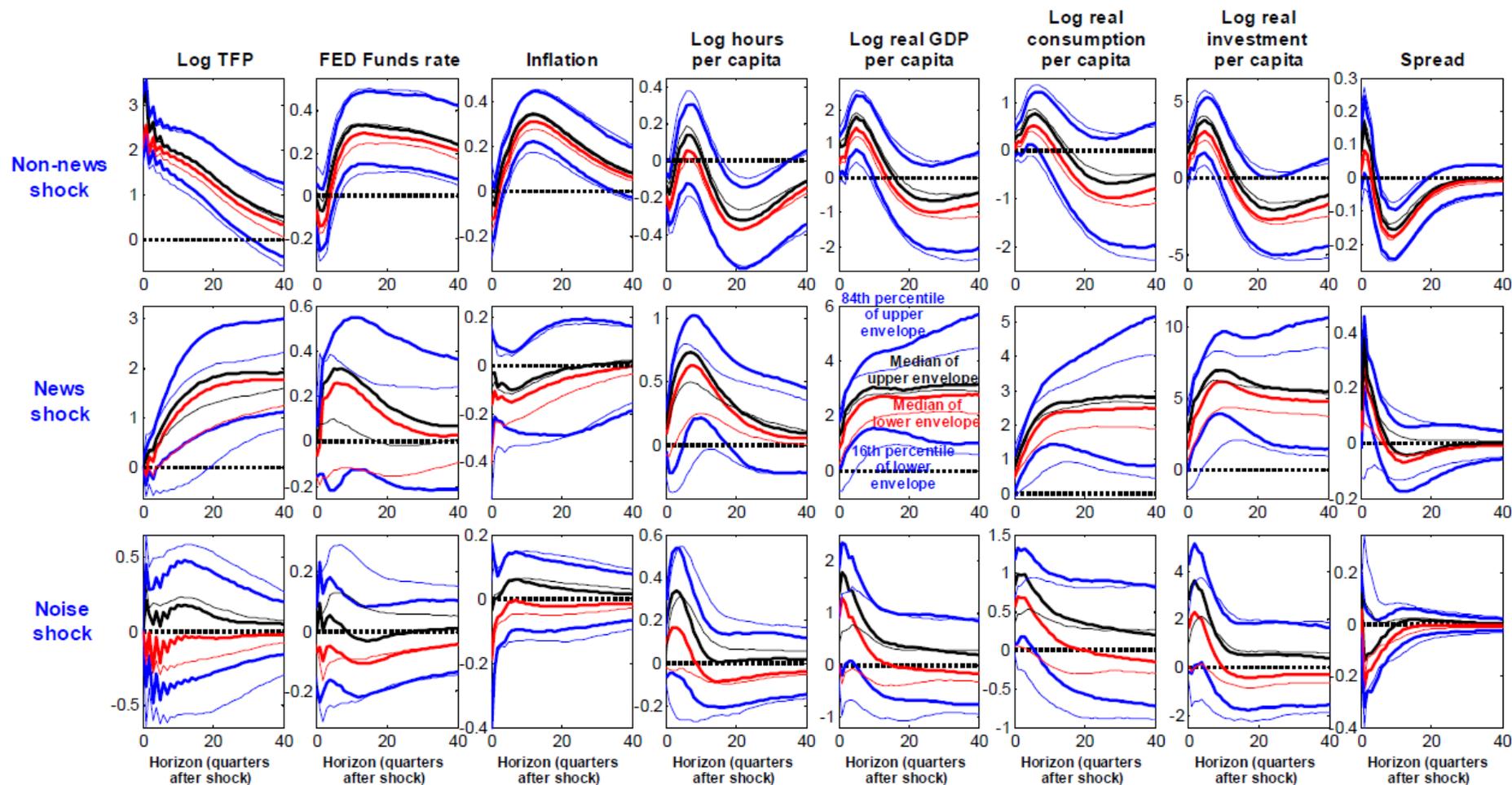


Figure VII.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) (thin lines: corresponding objects obtained without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks)

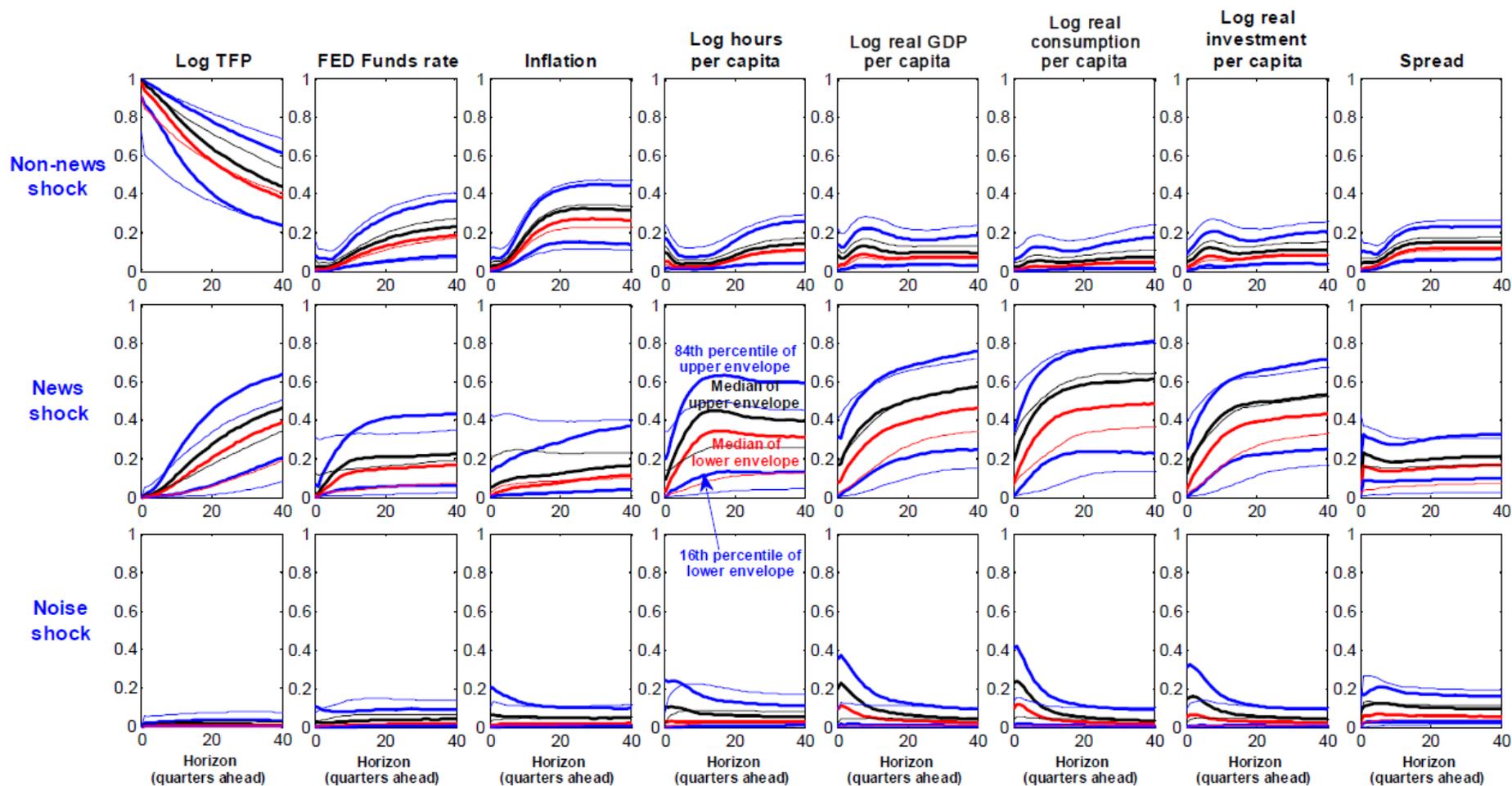


Figure VII.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1) (thin lines: corresponding objects obtained without imposing restrictions on the absolute magnitude of the IRF's to TFP noise shocks)

Results based on VARMA with 10 series and point identification, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

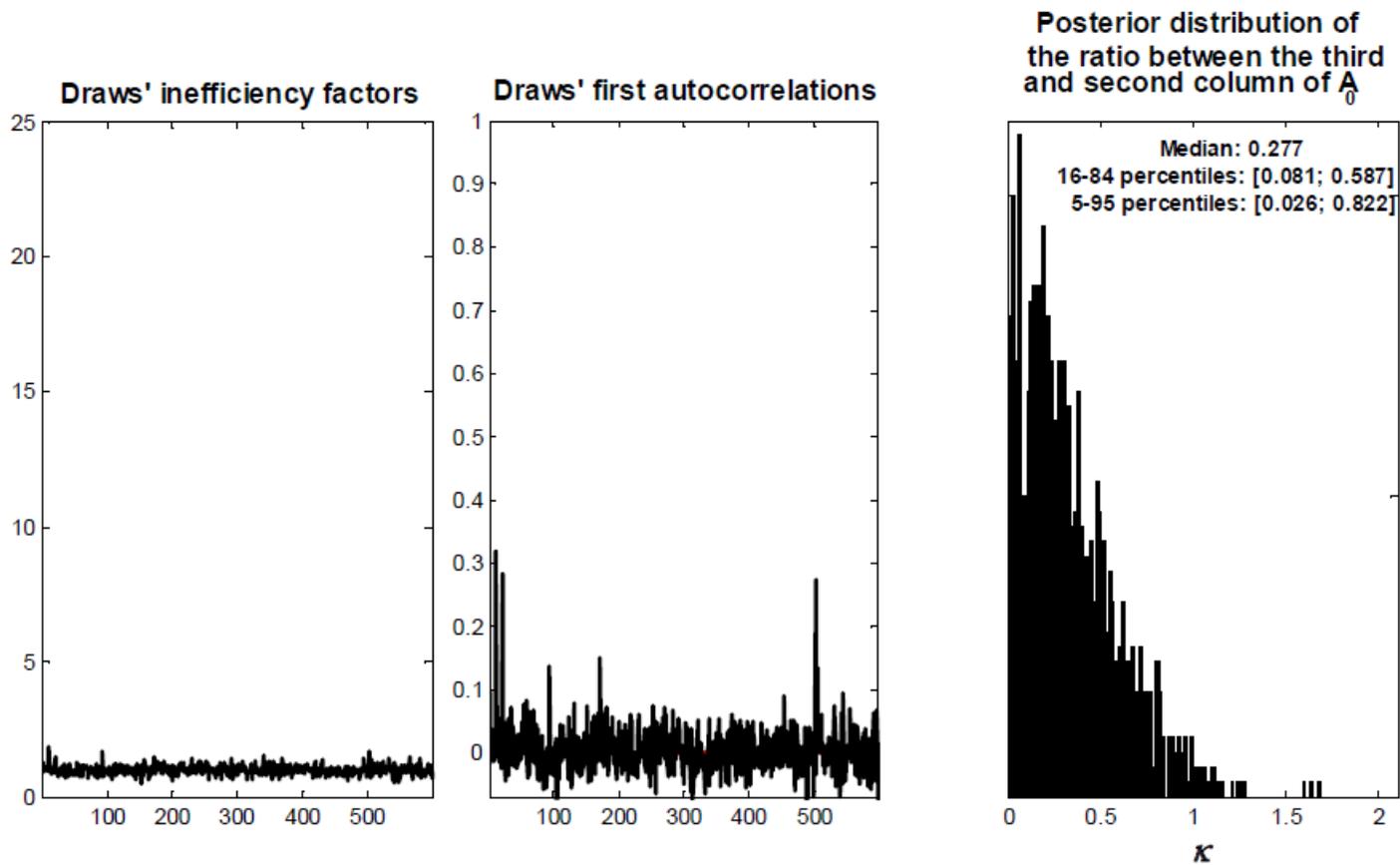


Figure VIII.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

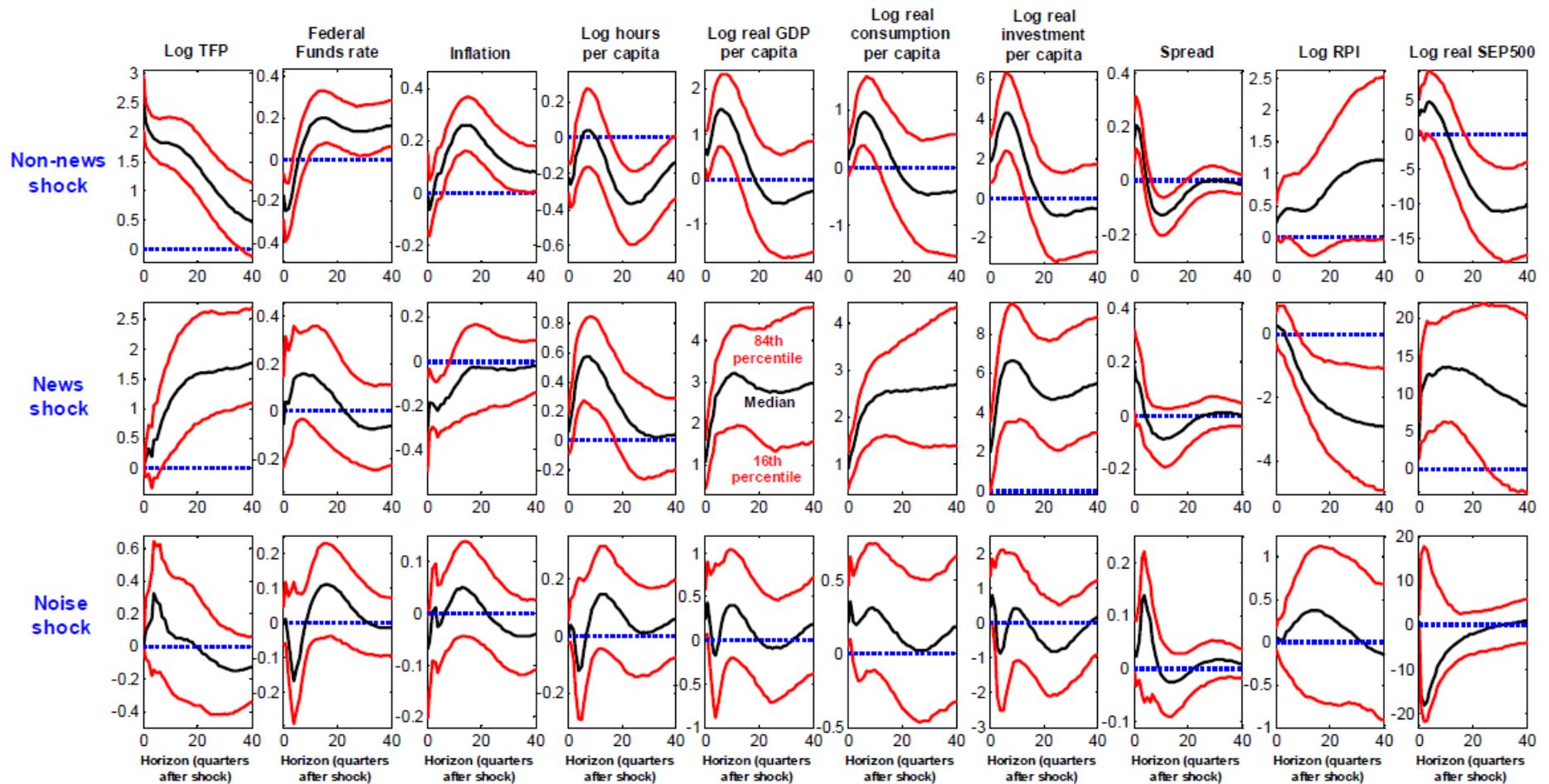


Figure VIII.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

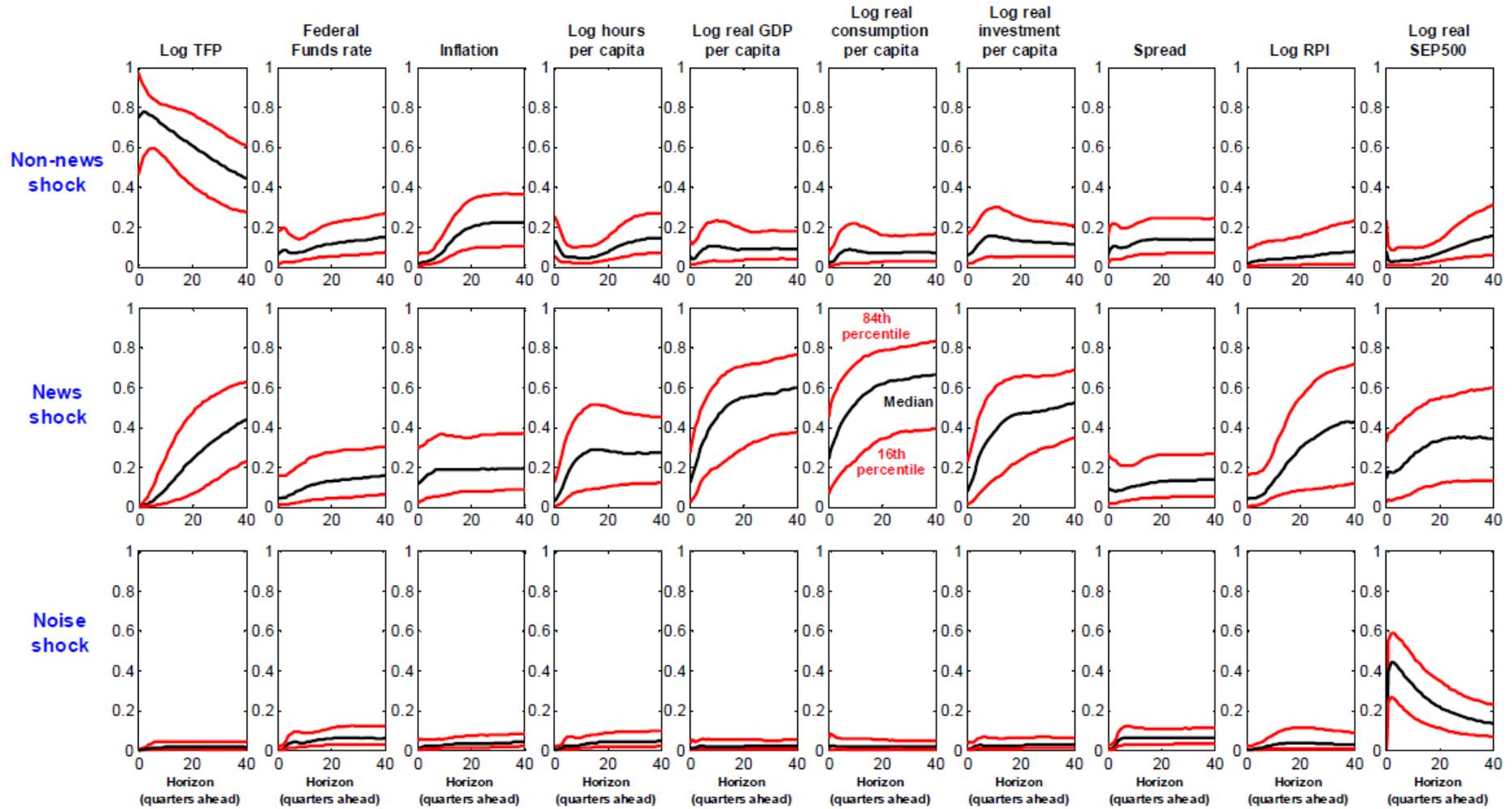


Figure VIII.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

Results based on VARMA with 10 series and set identification, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

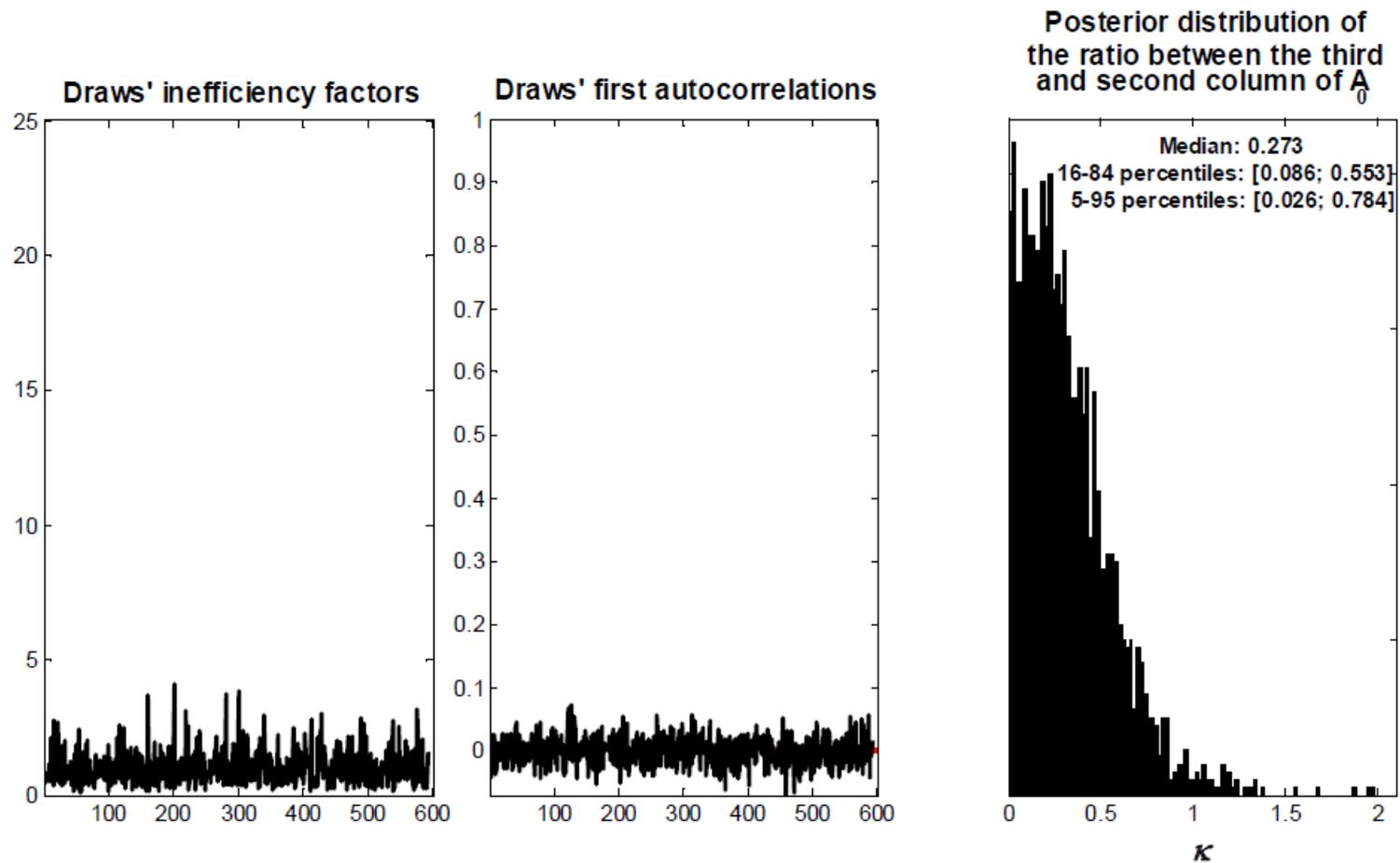


Figure IX.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

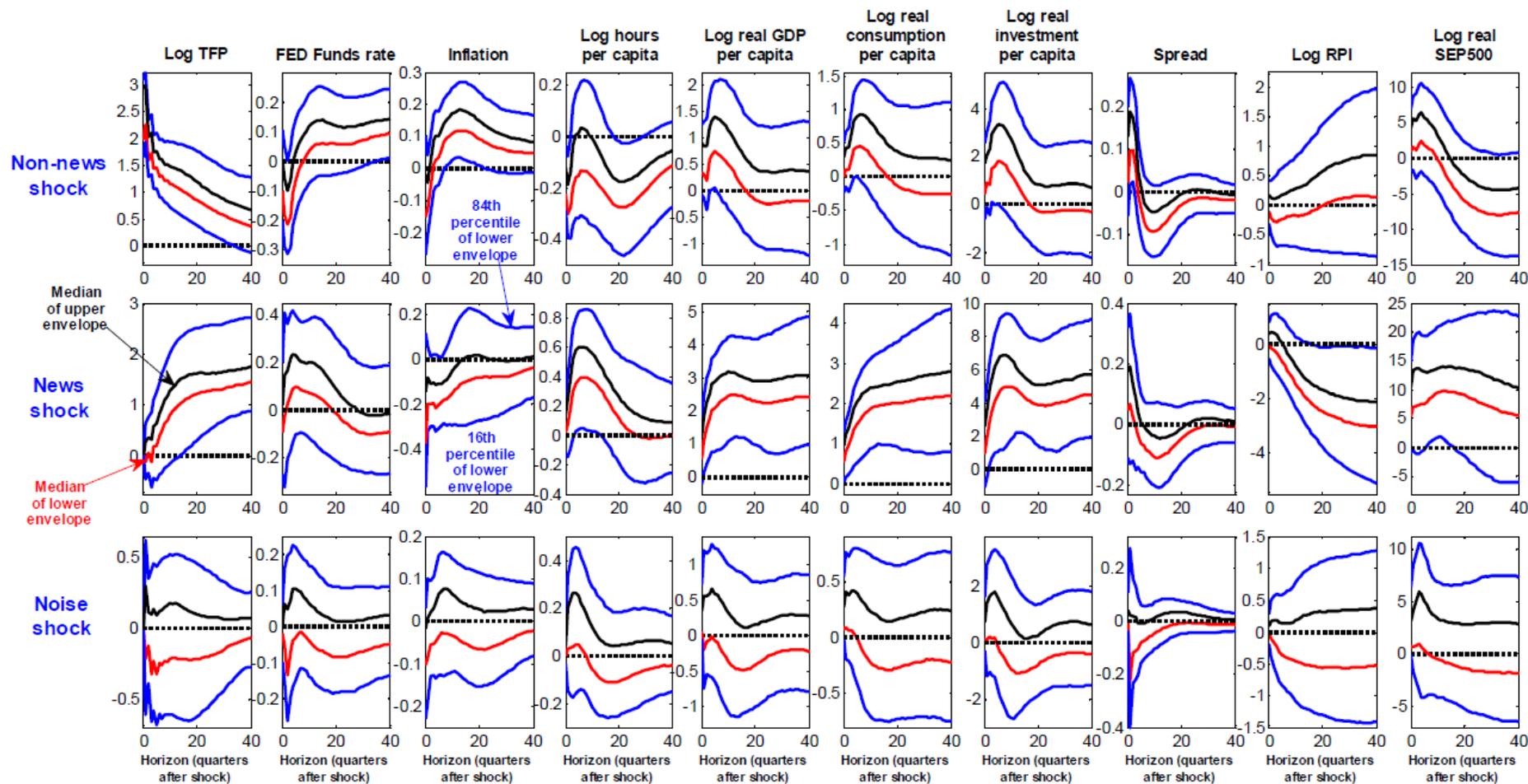


Figure IX.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

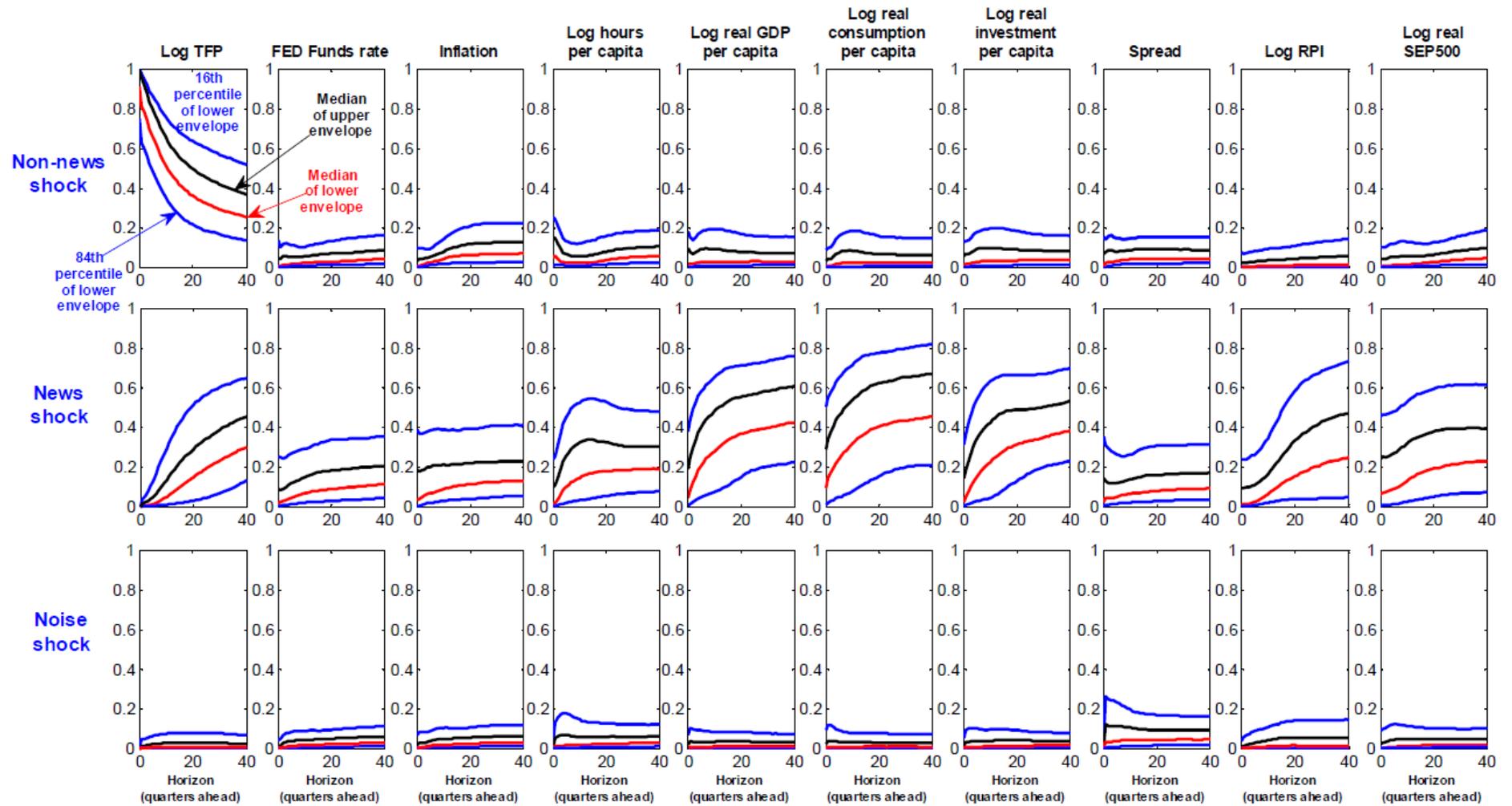


Figure IX.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

Results based on VARMA with 15 series and point identification, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

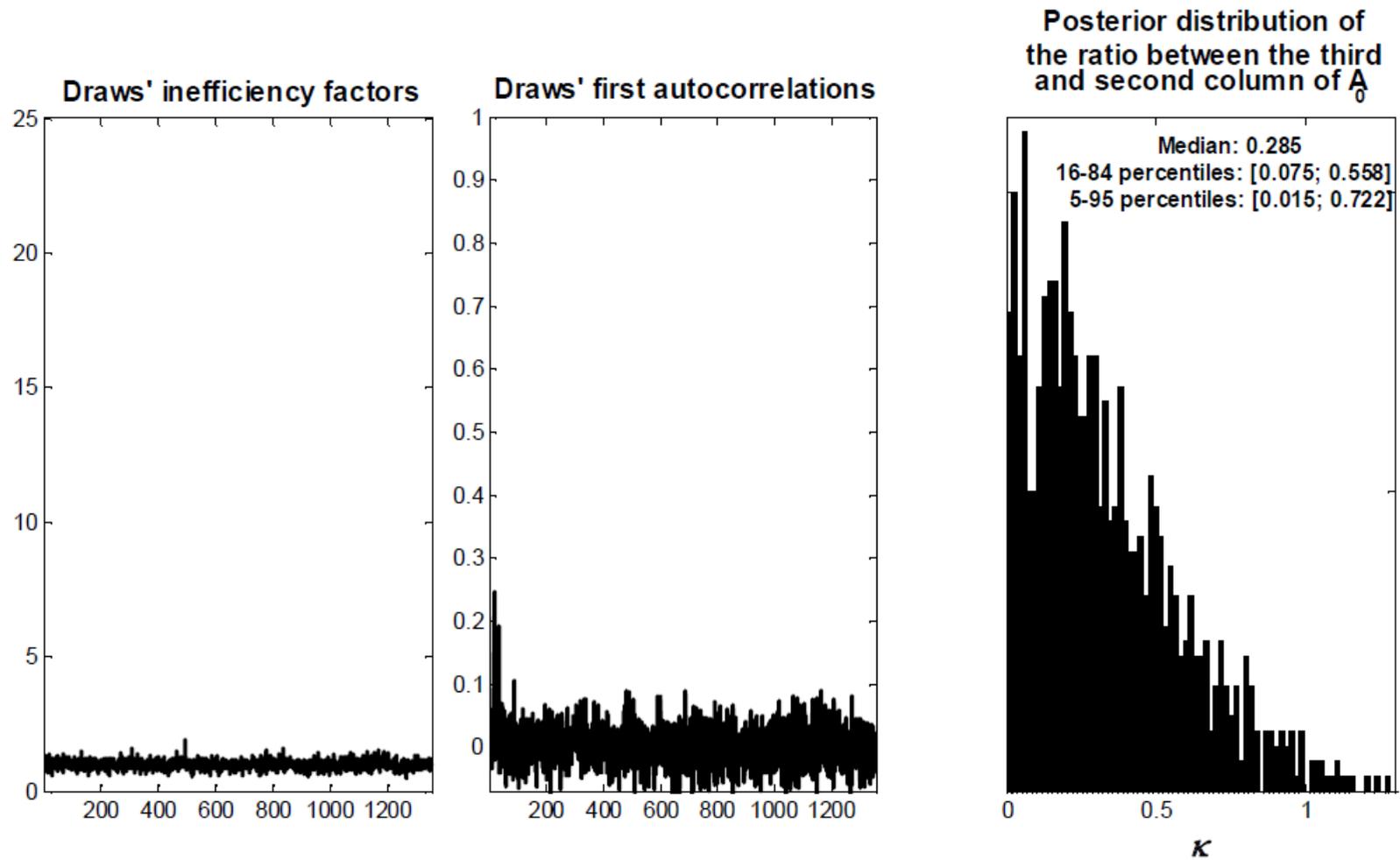


Figure X.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

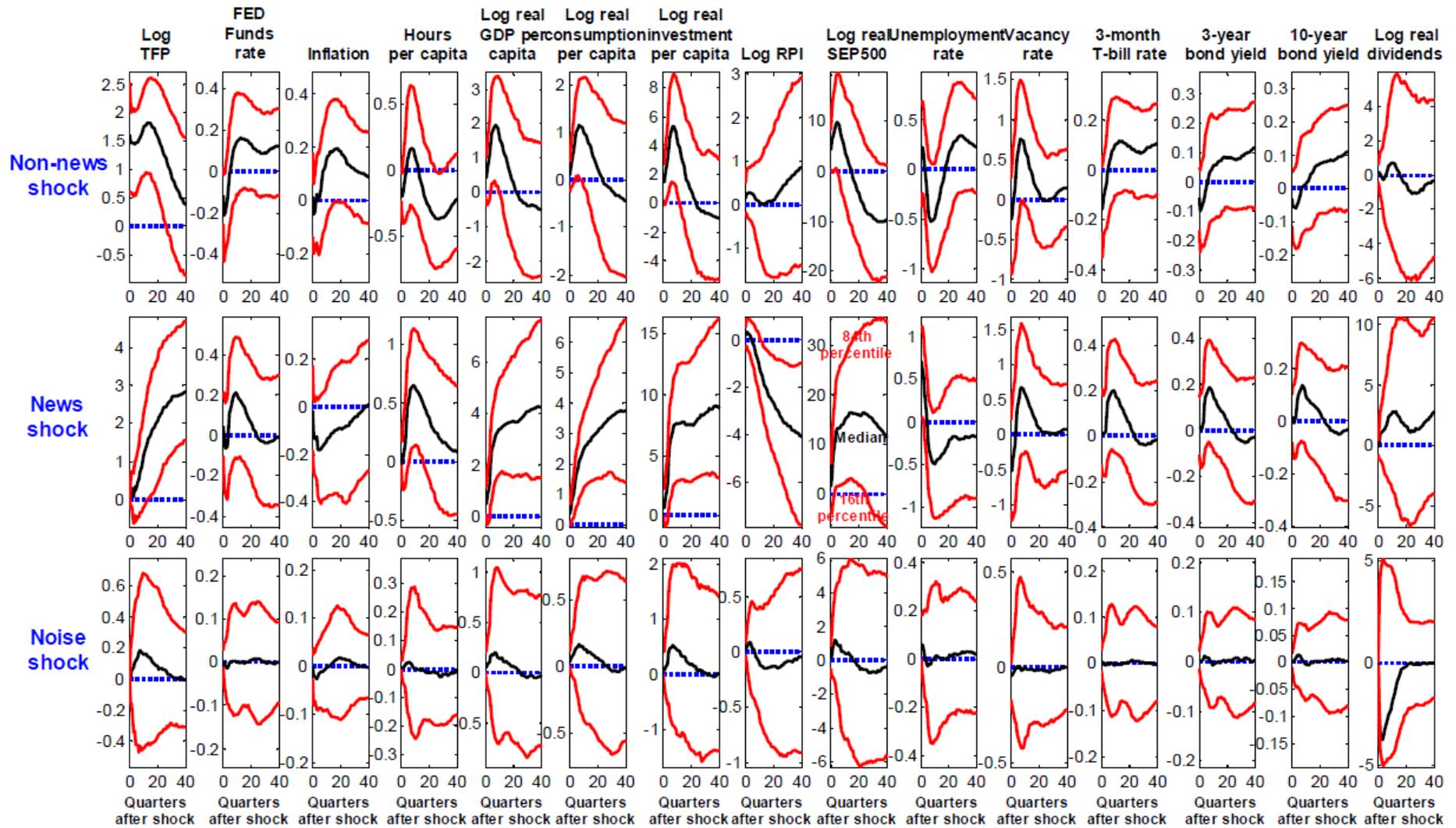


Figure X.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

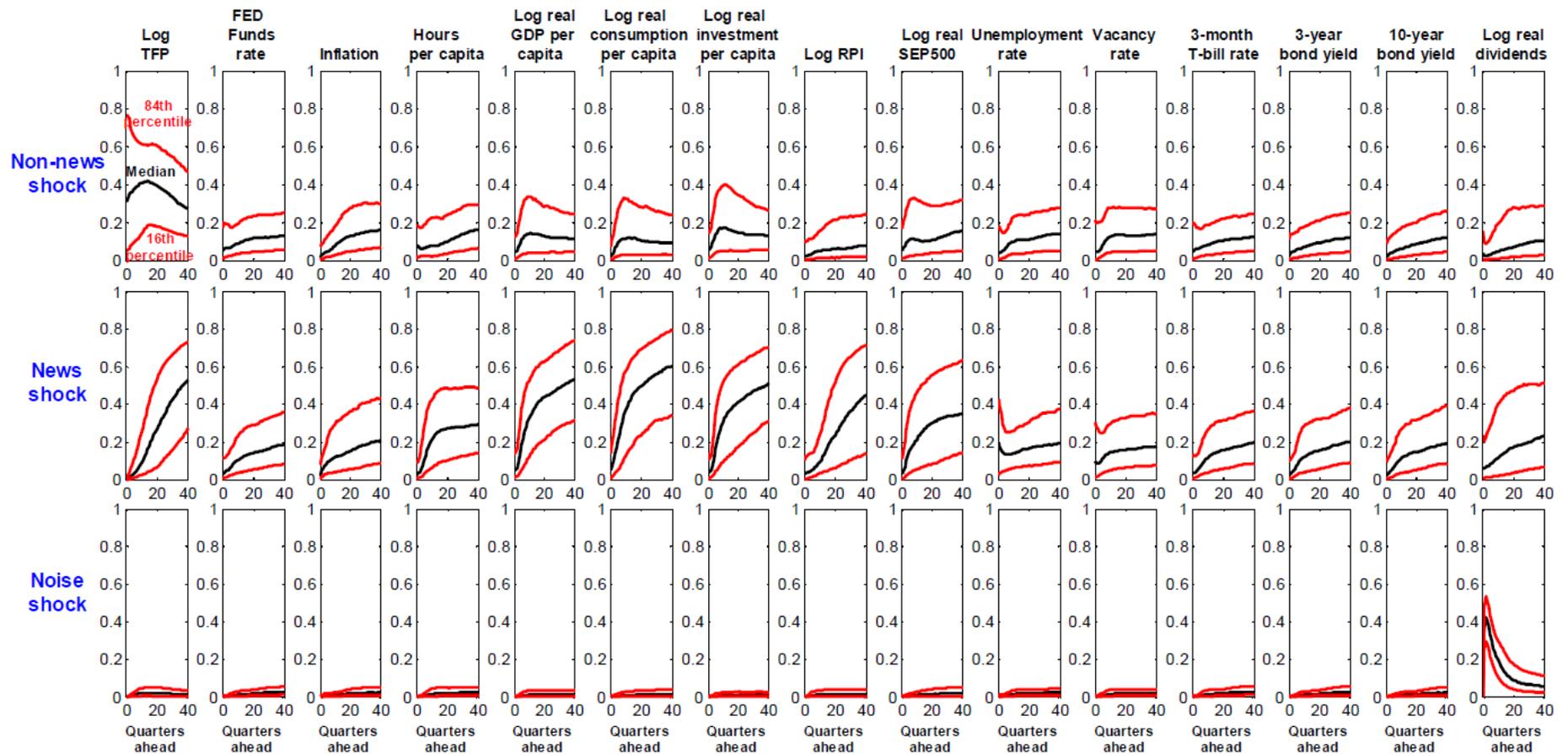


Figure X.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

Results based on VARMA with 15 series and set identification, without imposing restrictions on the absolute magnitude of the IRFs to TFP noise shocks

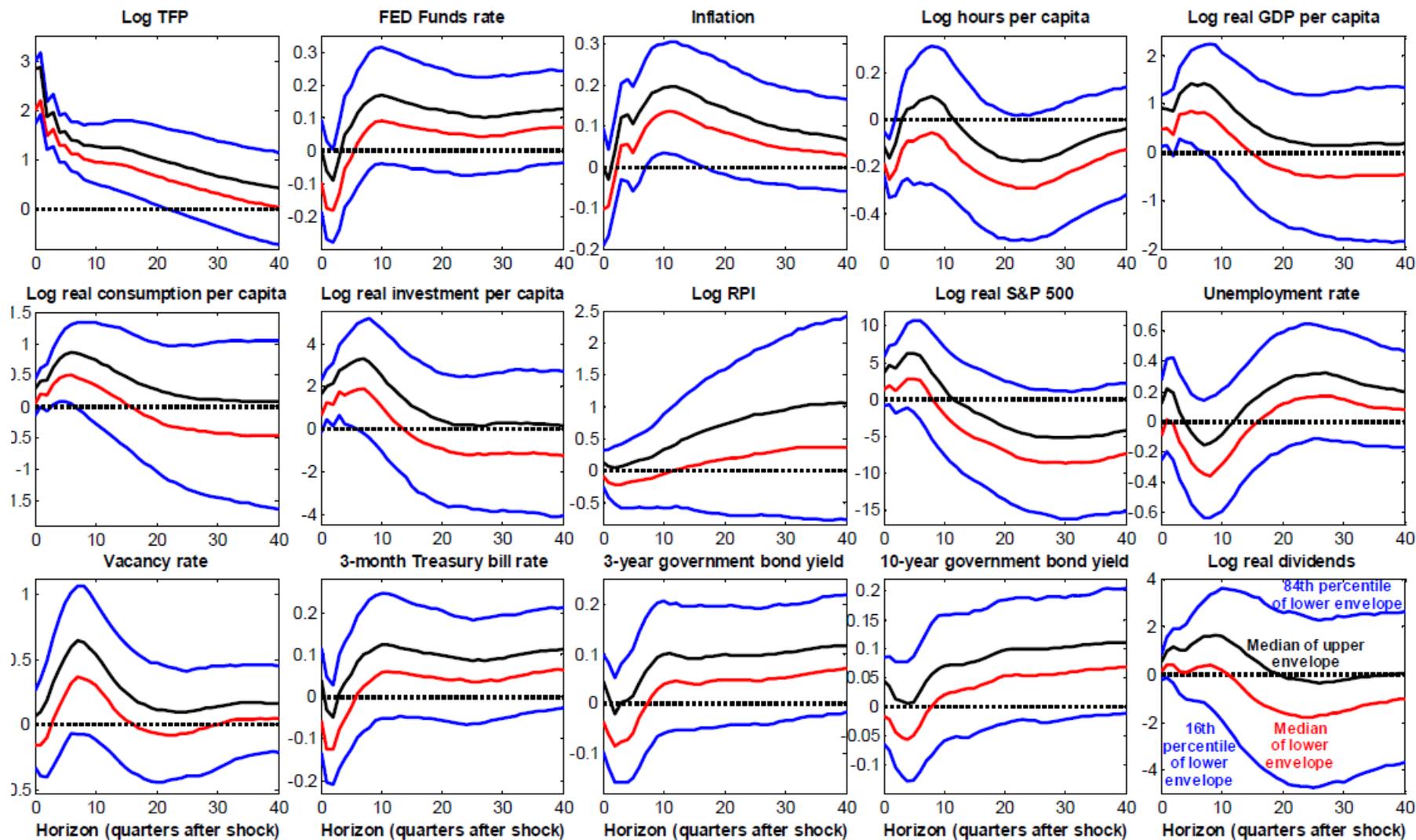


Figure XI.1a Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

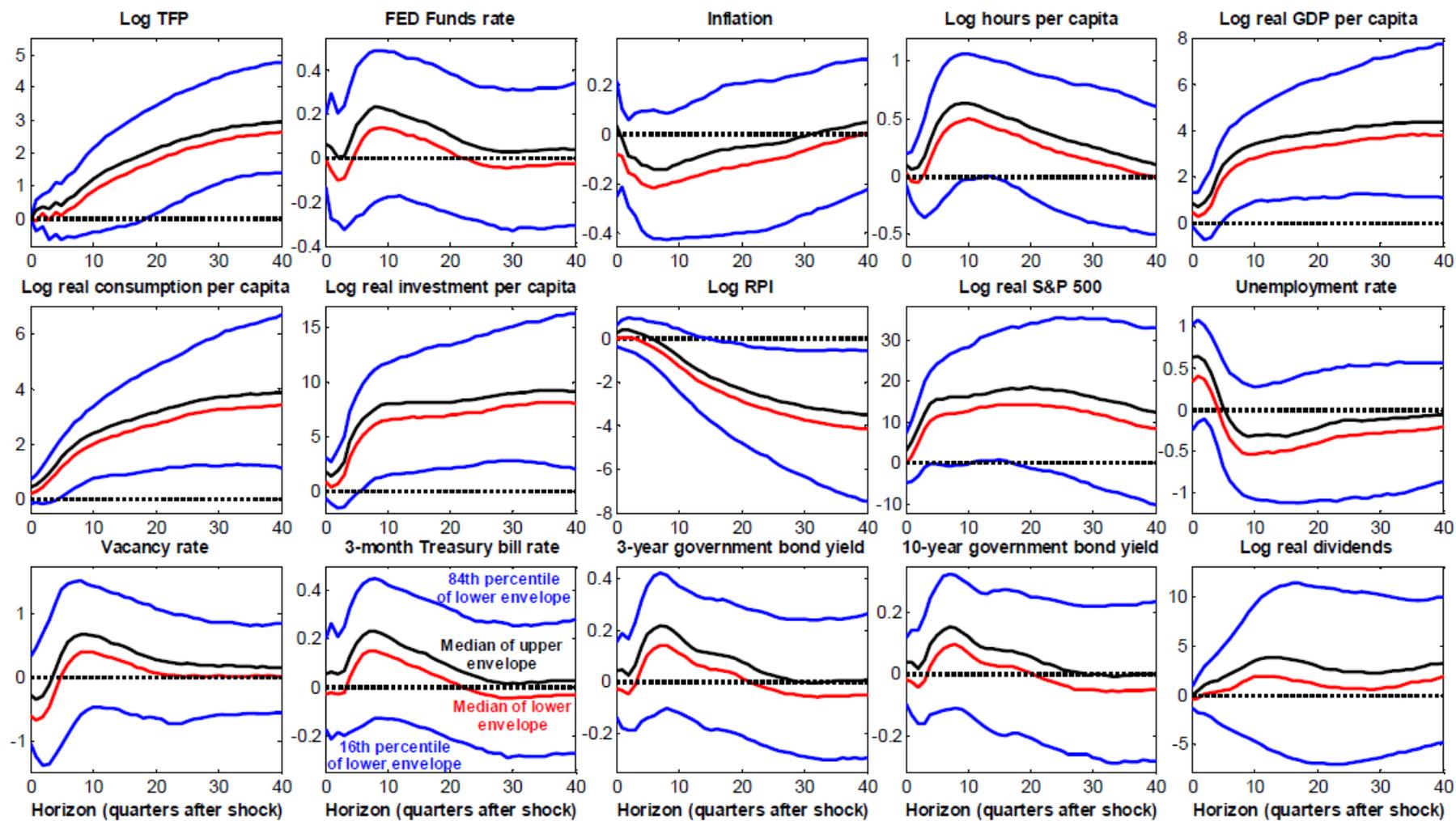


Figure XI.1b Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

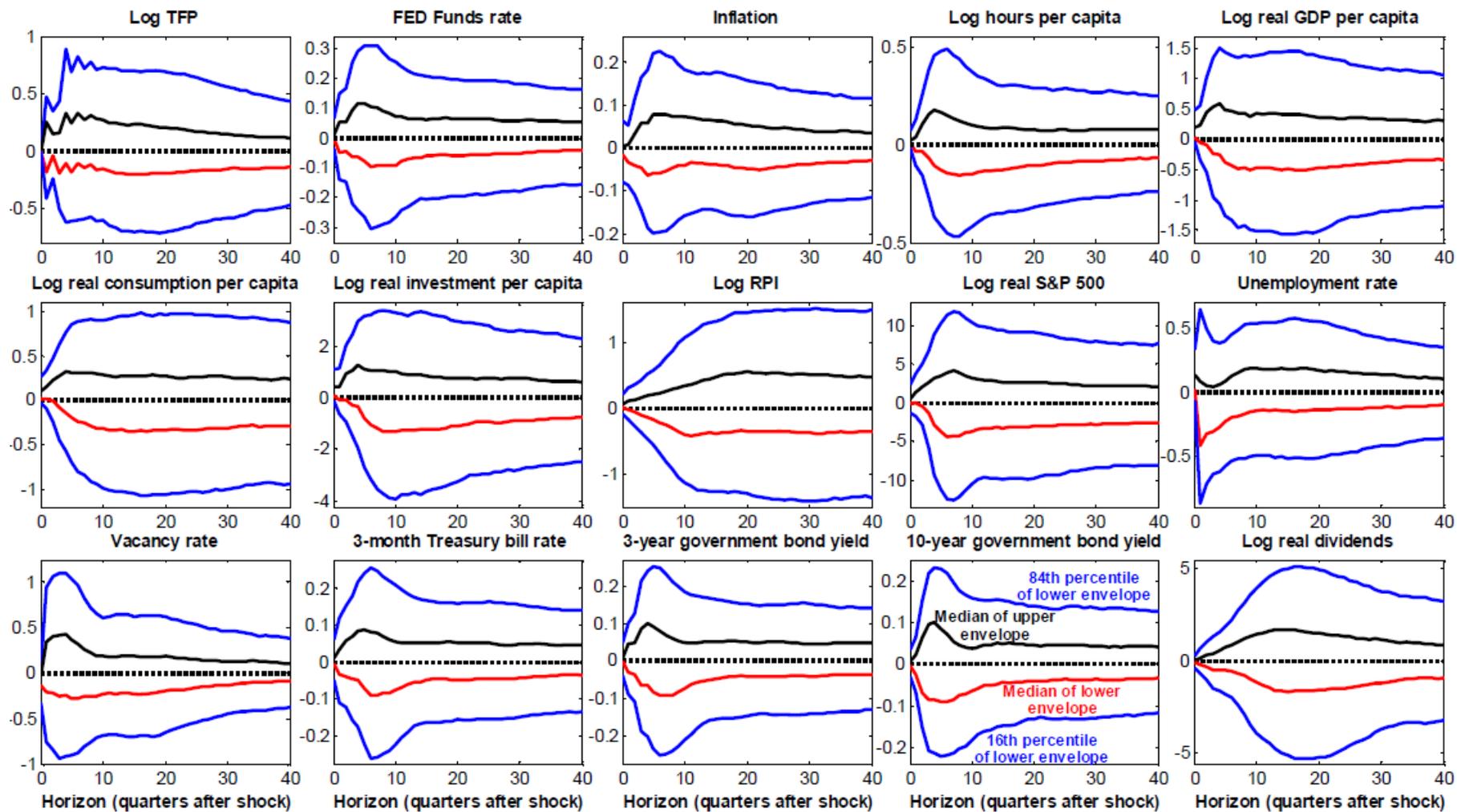


Figure XI.1c Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

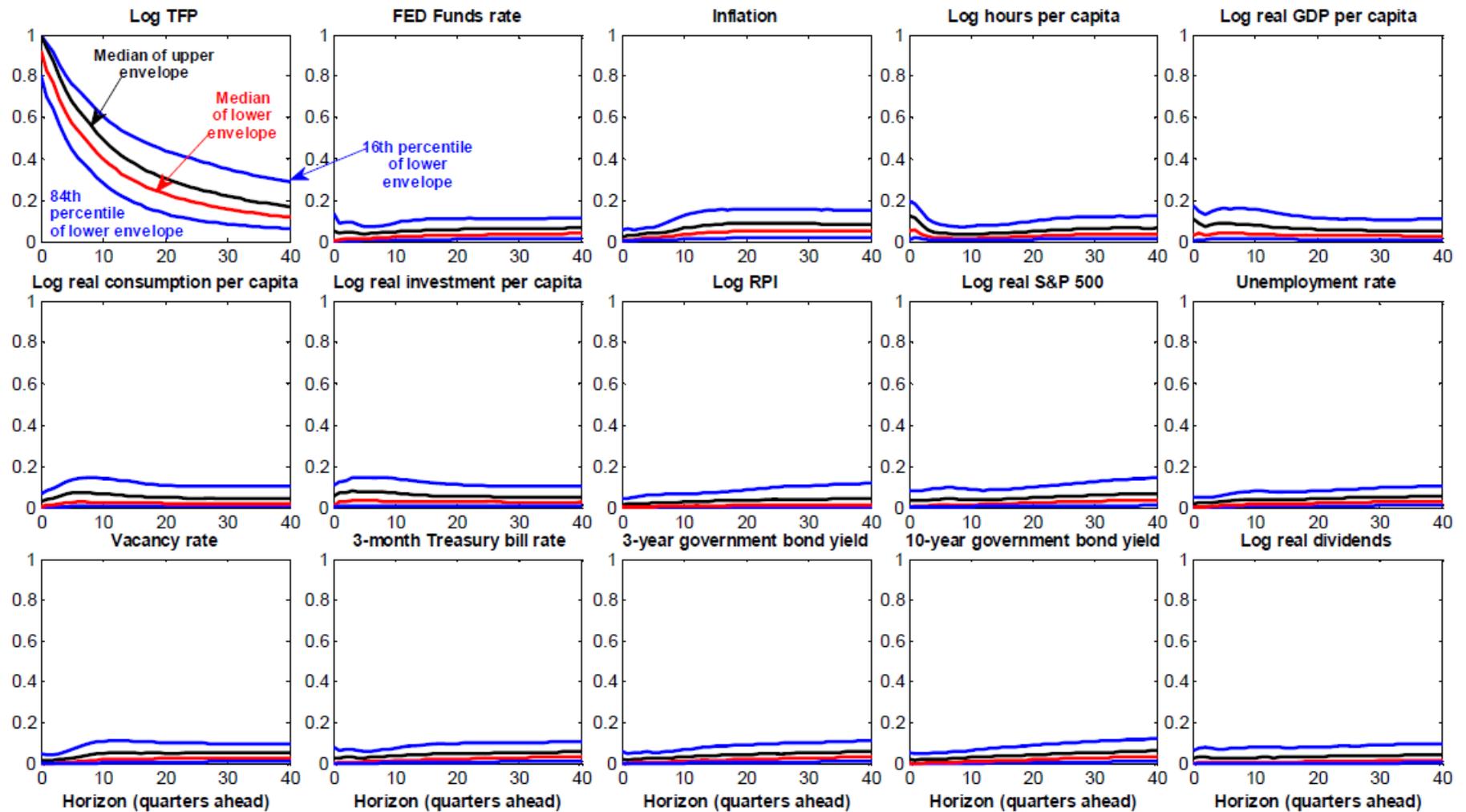


Figure XI.2a Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

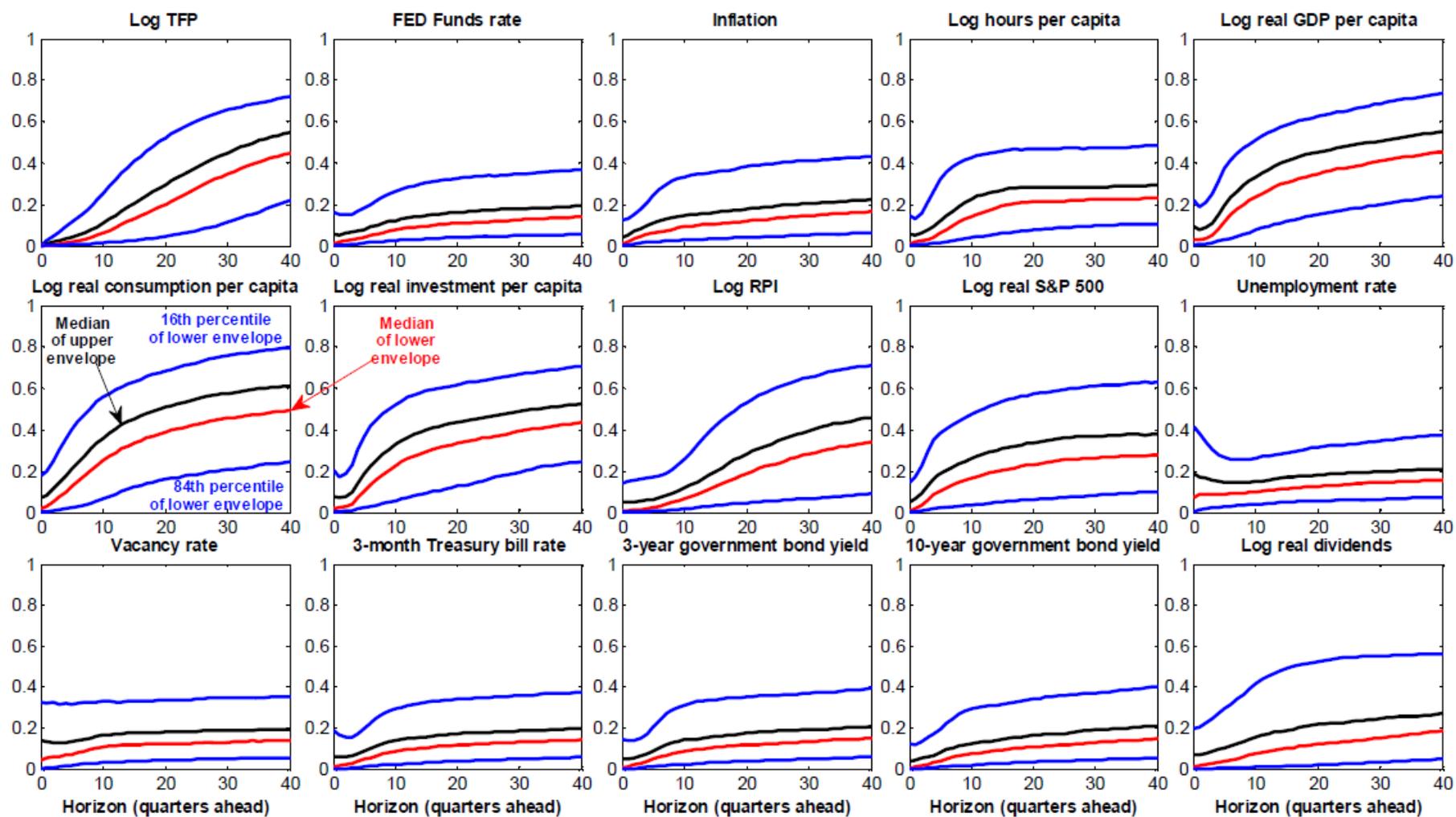


Figure XI.2b Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

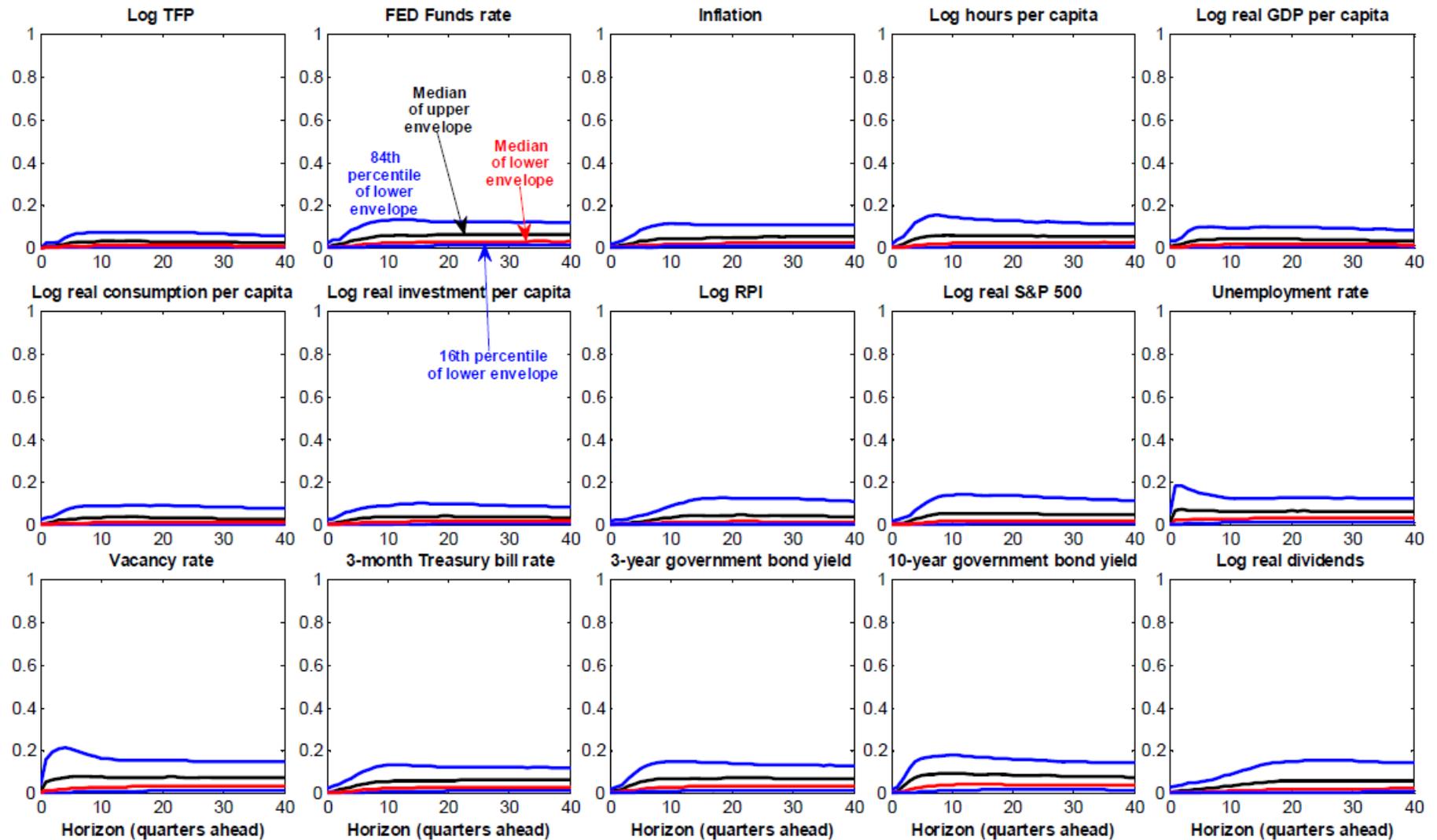


Figure XI.2c Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

Results based on VARMA with 7 series and set identification, without imposing restrictions on the absolute magnitude of the IRFs to noise shocks to real dividends

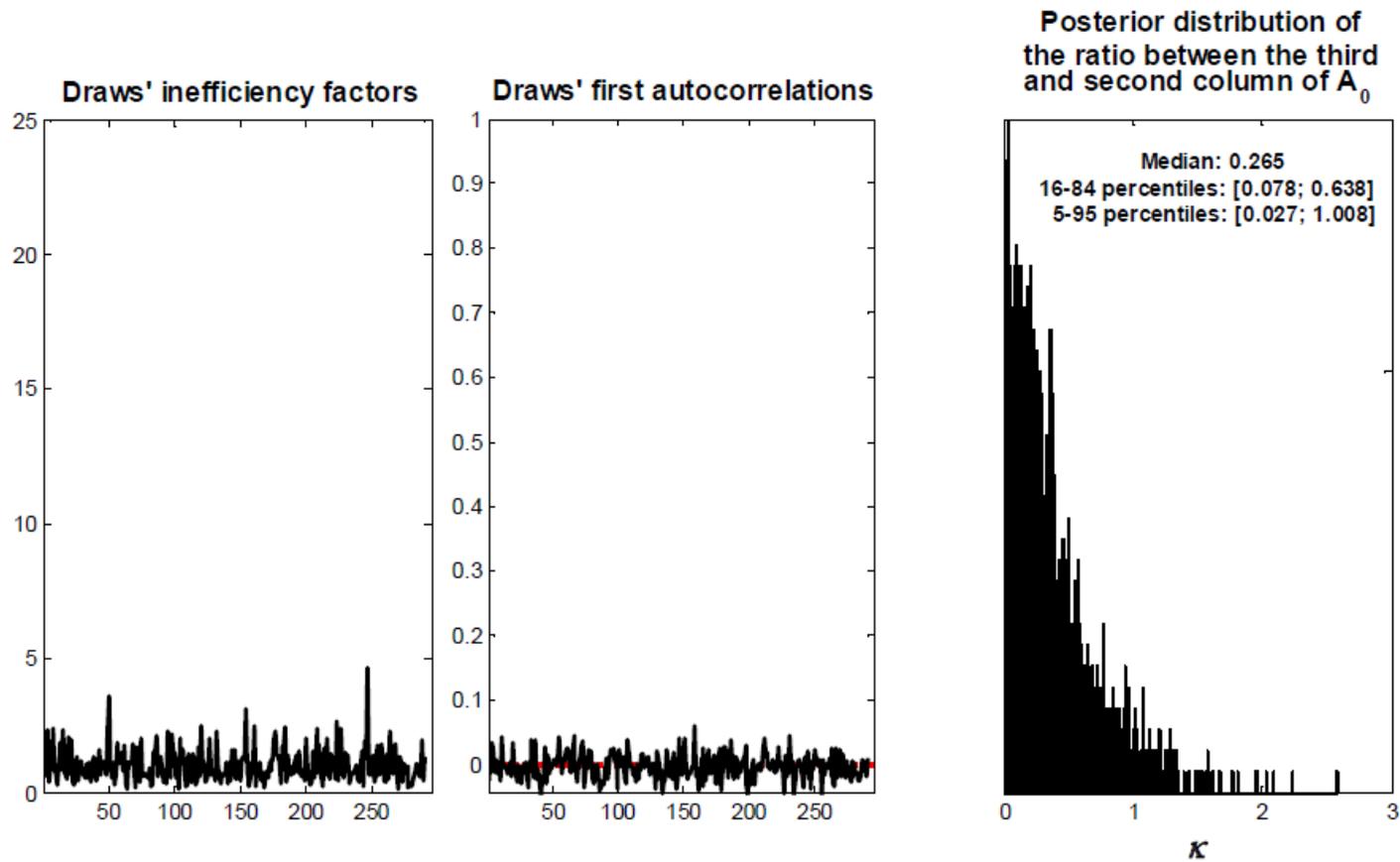


Figure XII.1 Draws' inefficiency factors and first autocorrelations, and posterior distribution of the ratio between the third and second column of  $A_0$ , based on a VARMA(4,1)

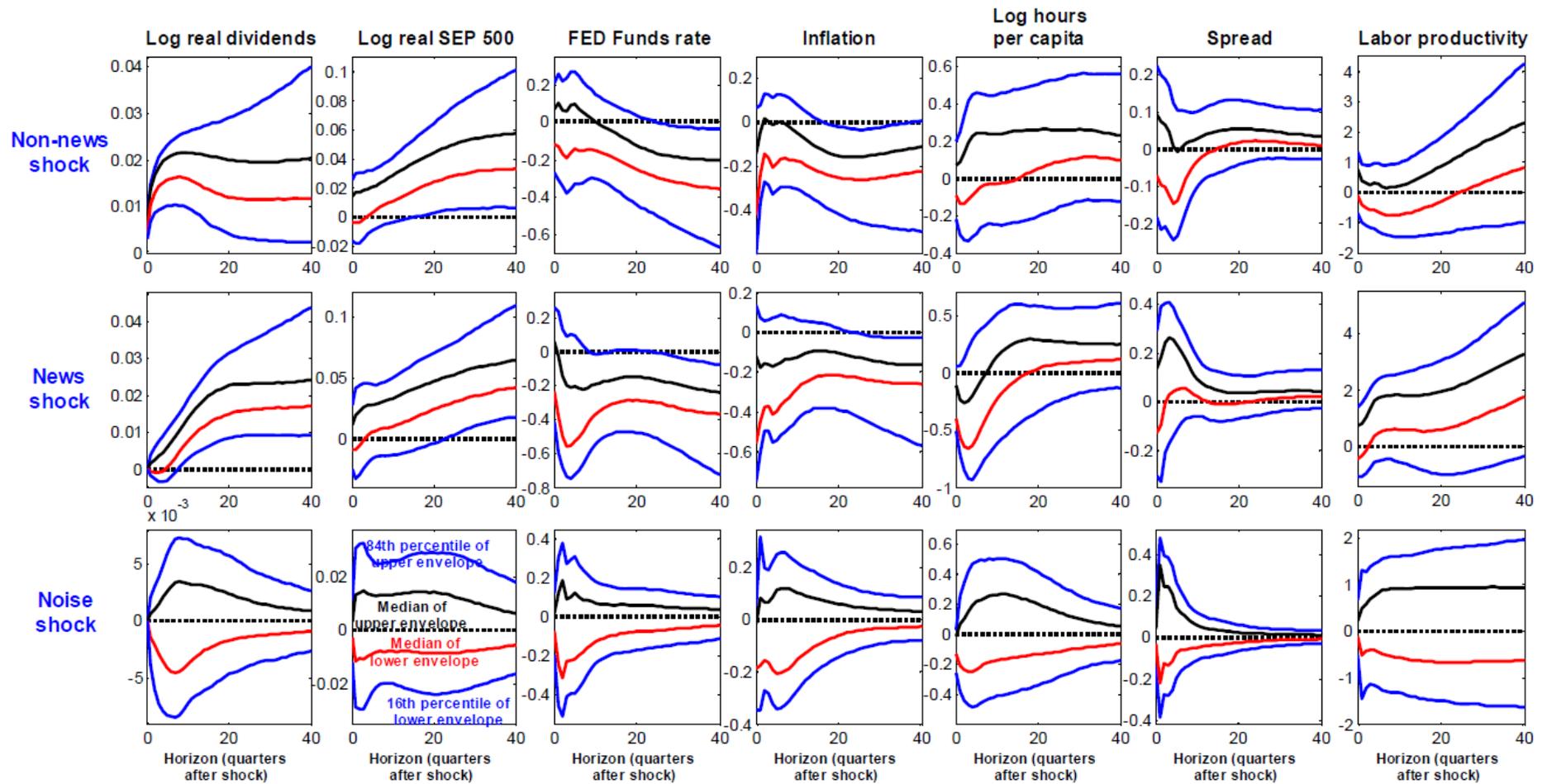


Figure XII.2 Impulse-response functions to non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

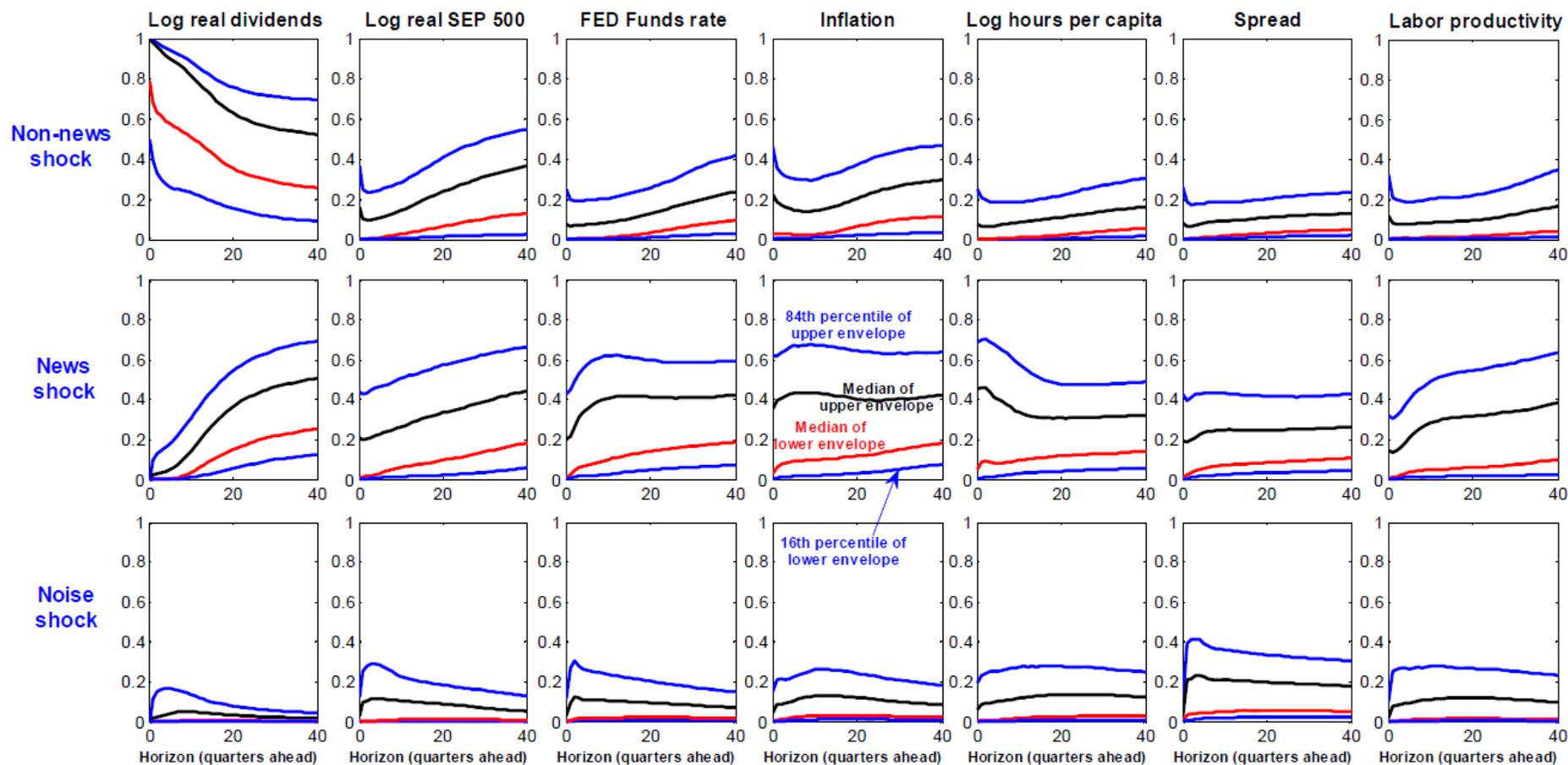


Figure XII.3 Fractions of forecast error variance explained by non-news, news, and noise shocks (median, and 16-84 percentiles of the posterior distribution), based on a VARMA(4,1)

Additional results based on point identification for  
dividends and stock prices and defense expenditure

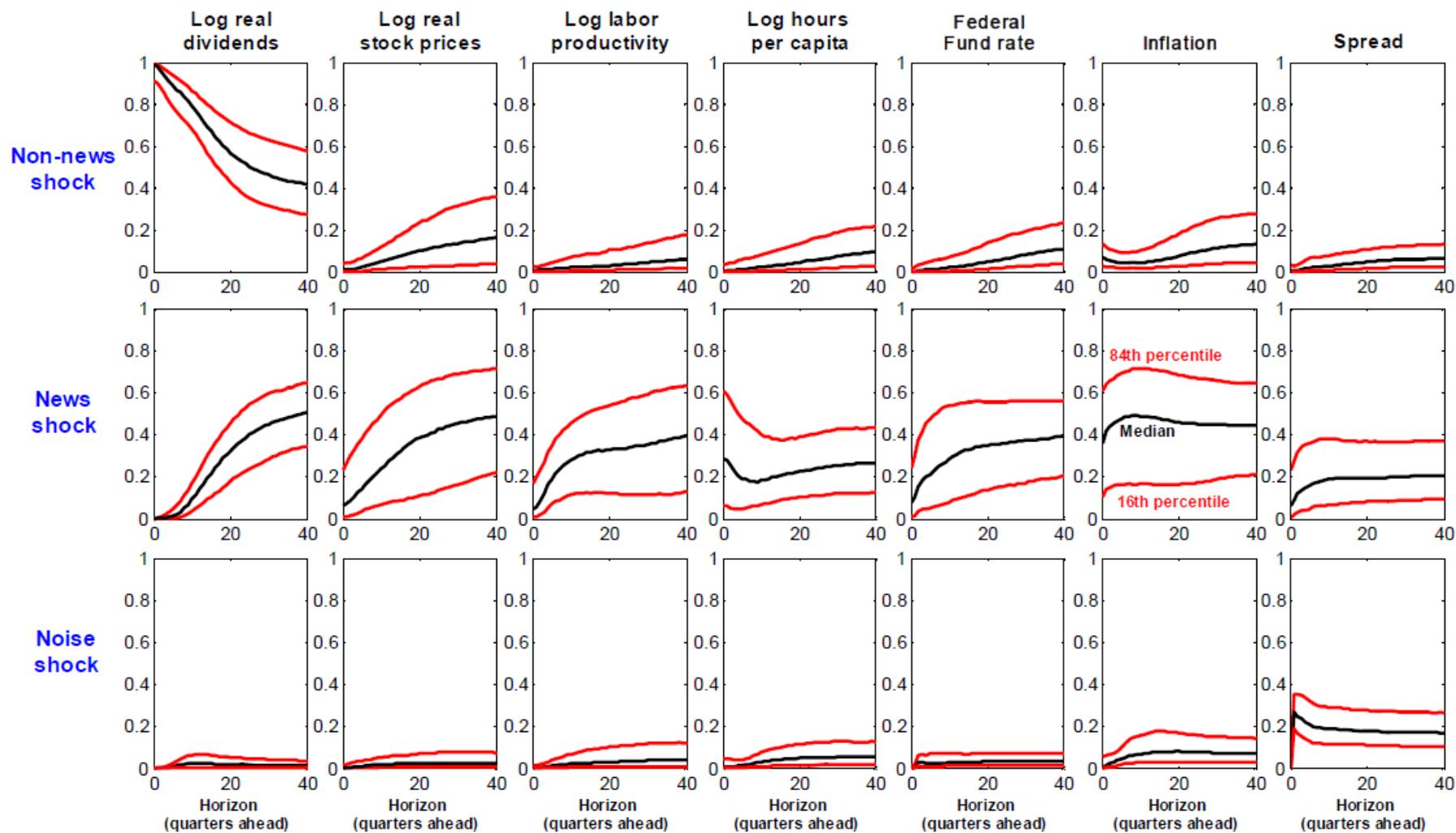


Figure XIII.1 Application with dividends and stock prices: Fractions of forecast error variance explained by non-news, news, and noise shocks, based on a VARMA(4,1), and point-identification

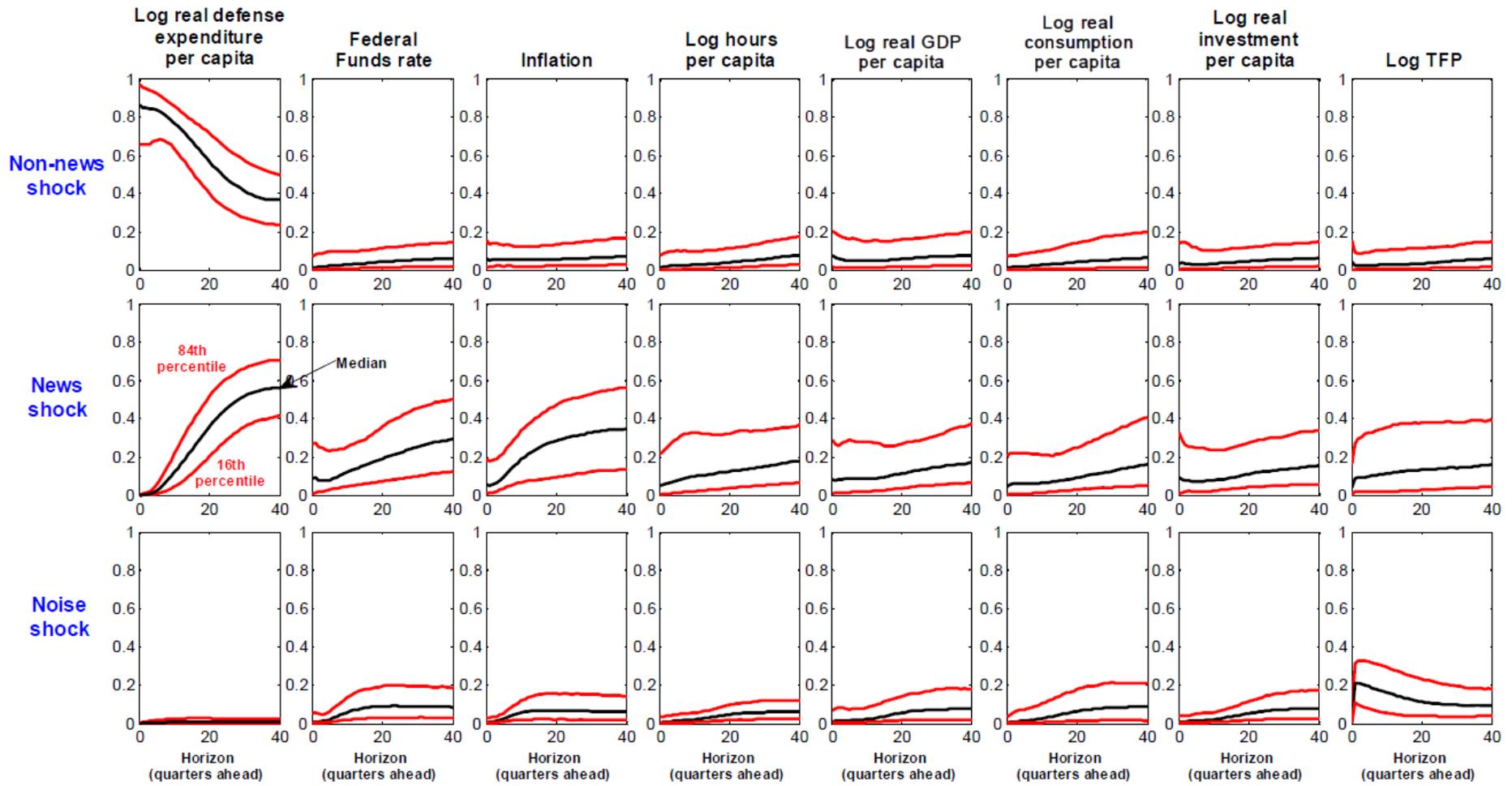


Figure XIII.2 Application with defense expenditure: Fractions of forecast error variance explained by non-news, news, and noise shocks, based on a VARMA(4,1), and point-identification