Online Appendix to "Reducing the State Space Dimension in a Large TVP-VAR"

Joshua C.C. Chan
Purdue University and UTS

Eric Eisenstat
University of Queensland

Rodney W. Strachan*†
University of Queensland

November 4, 2019

1 Obtaining the mode and Hessian for the ARMH step

We use a scoring algorithm to find $\widehat{f_\theta}$ numerically. To this end, the gradient and Hessian of the log of the conditional posterior of f_θ, $\ln p(f_\theta|., y)$, are given by:

$$
\begin{aligned}
d &\equiv \frac{d \ln p(f_\theta|., y)}{(df_\theta)'} = -H'Hf_\theta - \frac{1}{2}(I_r \otimes A_h')\iota_{rT} \\
&\quad + \frac{1}{2}(Z+W)'\Sigma^{-1}(y - X\alpha - Wf_\theta), \\
D &\equiv \frac{d^2 \ln p(f_\theta|., y)}{(df_\theta)(df_\theta)'} = D_1 + D_2 \\
D_1 &= -H'H - \frac{1}{2}Z'\Sigma^{-1}Z - \frac{1}{2}W'\Sigma^{-1}W, \\
D_2 &= -\frac{1}{2}(Z-W)'\Sigma^{-1}W - \frac{1}{2}W'\Sigma^{-1}(Z-W), \\
Z &= Y(I_{rT} \otimes A_h) + W,
\end{aligned}
$$

where $Y = diag\left((y_1 - x_1\alpha_1)', ..., (y_T - x_T\alpha_T)'\right)$, $\Sigma = diag(h_1', \dots, h_T')$,

$$
W = \begin{bmatrix} x_1 A_\theta & & \\ & \ddots & \\ & & x_T A_\theta \end{bmatrix}, \quad H = \begin{bmatrix} I_r & & & \\ -I_r & I_r & & \\ & \ddots & \ddots & \\ & & -I_r & I_r \end{bmatrix},
$$

Observe that given this, a standard Newton-Raphson algorithm could be constructed by updating

$$
\widehat{f}_\theta^{(j+1)} = \widehat{f}_\theta^{(j)} - D^{-1}d.
$$

However, $-D$ is not guaranteed to be positive definite for all f_θ, and in fact, will only be positive definite in a very small neighborhood around $\widehat{f_\theta}$ in many applications. Thus, using the standard

*Corresponding author: School of Economics, University of Queensland, r.strachan@uq.edu.au. Rodney Strachan is a Fellow of the Rimini Centre for Economic Analysis and the Centre for Applied Macroeconomic Analysis.

†This research was supported by funding from the Australian Research Council project DP180102373.

Newton-Raphson scoring algorithm will not work well in practice. Nevertheless, we can construct a similar algorithm by replacing D with D_1.

The advantage of this approach is that D_1 is guaranteed to be positive definite for all f_θ, and therefore, an update from any f_θ will always be an ascent direction. The disadvantage, of course, is that in the neighborhood around the mode where D is positive definite, the convergence may be theoretically slower than what is achieved by standard Newton-Raphson. However, even this drawback may be small to the extent that $E_y(D_2) = 0$. In fact, D_1 is closely related to the *Fisher information matrix*

$$F = -H'H - \frac{1}{2}\left(I_{rT} \otimes A_h' A_h\right) - W'\Sigma^{-1}W,$$

which is sometimes used to construct scoring algorithms. Using either F or D will guarantee positive ascent for any value of f_θ; we prefer D_1 as it appears to yield faster convergence in practice. Finally, note that D, D_1 and F are all sparse, banded matrices which results in fast computation of updates even in large dimensions.

2 Derivation of posterior terms

In this appendix we define the terms in the conditional posteriors presented in Section 3 for a_α and a_h in both specifications and f_h and f_α in Specification 2. Each of these parameters has a normal prior of the form $a_\alpha \sim \mathcal{N}\left(0, \underline{V}_\alpha\right)$, $f_\alpha \sim \mathcal{N}\left(0, \underline{V}_{f,\alpha}\right)$, $a_h \sim \mathcal{N}\left(0, \underline{V}_h\right)$, and $f_h \sim \mathcal{N}\left(0, \underline{V}_{h,\alpha}\right)$, and a conditional normal posterior.

For Specification 2, recall the model specification in (??) and (??) reproduced here:

$$\begin{gathered} y_t = x_t\alpha + x_t A_\alpha f_{\alpha,t} + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \Sigma_t\right), \\ f_{\alpha t} = f_{\alpha,t-1} + z_{\alpha,t}, \quad z_{\alpha,t} \sim N\left(0, I_{r_\alpha}\right), \quad f_{\alpha,0} = 0, \\ \Sigma_t = diag\left(e^{h_{1,t}}, \ldots, e^{h_{n,t}}\right) \qquad h_t = \left(h_{1,t}, , \ldots, h_{n,t}\right)' \\ h_t = h + A_h f_{h,t}, \\ f_{h,t} = f_{h,t-1} + z_{h,t} \quad z_{h,t} \sim N\left(0, I_{r_h}\right), \quad f_{h,0} = 0, \end{gathered}$$

To obtain a simple form for the posterior for $a_\alpha = \left(\alpha', vec\left(A_\alpha\right)'\right)'$ we use

$$\begin{aligned} y_t &= x_t\alpha + \left(f'_{\alpha,t} \otimes x_t\right) vec\left(A_\alpha\right) + \varepsilon_t \\ &= \left[x_t \quad \left(f'_{\alpha,t} \otimes x_t\right)\right] a_\alpha + \varepsilon_t \end{aligned}$$

Stack y_t over time to form the $Tn \times 1$ vector y, stack the matrices $\left[x_t \quad \left(f'_{\alpha,t} \otimes x_t\right)\right]$ into the $Tn \times kr_\alpha$ matrix X, and similarly stack ε_t into the $Tn \times 1$ vector ε. We can now write the measurement equation as

$$y = Xa_\alpha + \varepsilon \text{ where } \varepsilon \sim N\left(0, \Sigma\right). \tag{1}$$

Σ is the diagonal matrix in which the $(t+i, t+i)^{th}$ element is the variance of the i^{th} element of ε_t where $i \in \{1, \ldots, n\}$. With a prior of the form $\mathcal{N}\left(0, \underline{V}_\alpha\right)$, the posterior has the form $\mathcal{N}\left(\overline{a}_\alpha, \overline{V}_\alpha\right)$ where $\overline{V}_\alpha = \left[X'\Sigma^{-1}X + \underline{V}_\alpha^{-1}\right]^{-1}$ and $\overline{a}_\alpha = \overline{V}_\alpha X'\Sigma^{-1}y$.

To define the terms in the posterior for the factors $f_{\alpha,t}$ and $f_{h,t}$, we first define the form of the prior covariance matrices $\underline{V}_{f,\alpha}$ and $\underline{V}_{f,h}$. Let the $(r \times 1)$ vector f_t be either $f_{\alpha,t}$ or $f_{h,t}$ such that $r = r_\alpha$ or $r = r_h$ respectively. In generic form then, the state equation for the factors can be written as

$$f_t = f_{t-1} + z_t, \quad z_t \sim N\left(0, I_r\right), \quad f_0 = 0.$$

Stack the f_t into the $Tr \times 1$ vector f, similarly stack the z_t into z, and let R be the $(Tr \times Tr)$ differencing matrix,

$$R_r = \begin{bmatrix} I_r & 0 & 0 & 0 \\ -I_r & I_r & 0 & 0 \\ 0 & -I_r & I_r & 0 \\ & & & \\ 0 & 0 & 0 & I_r \end{bmatrix}.$$

We can now write

$$R_r f = z \quad z \sim \mathcal{N}\left(0, I_{Tr}\right) \text{ and so } f = R_r^{-1} z \sim \mathcal{N}\left(0, \left(R'_r R_r\right)^{-1}\right).$$

This shows that the priors covariance matrices for $f_{\alpha,t}$ and $f_{h,t}$ are $\underline{V}_{f,\alpha} = \left(R'_{r_\alpha} R_{r_\alpha}\right)^{-1}$ and $\underline{V}_{f,h} = \left(R'_{r_h} R_{r_h}\right)^{-1}$ respectively.

To define the terms in the posterior for the $f_{\alpha,t}$, $\mathcal{N}\left(\overline{f}_{\alpha},\overline{V}_{f,\alpha}\right)$, we write the measurement equation as

$$\begin{aligned} y_t - x_t\alpha &= x_t A_\alpha f_{\alpha,t} + \varepsilon_t \\ &= \overline{x}_t a_\alpha + \varepsilon_t \text{ where } \overline{x}_t = \begin{bmatrix} x_t & \left(f'_{\alpha,t}\otimes x_t\right)\end{bmatrix}. \end{aligned}$$

Stacking $y_t^f = y_t - x_t\alpha$ over time to form the $Tn \times 1$ vector y^f and defining

$$X^f = \begin{bmatrix} x_1A_\alpha & 0 & & 0 \\ 0 & x_2A_\alpha & \cdots & 0 \\ & \vdots & \ddots & \\ 0 & 0 & & x_TA_\alpha \end{bmatrix}$$

we can write $\overline{V}_{f,\alpha} = \left[X'\Sigma^{-1}X + R'_{r_\alpha}R_{r_\alpha}\right]^{-1}$ and $\overline{f}_\alpha = \overline{V}_{f,\alpha}X^{f\prime}\Sigma^{-1}y^f$.

For the vector $a_h = \left(h' \quad vec\left(A_h\right)'\right)'$ with posterior $\mathcal{N}\left(\overline{a}_h,\overline{V}_h\right)$, we apply the transformation from Kim, Shephard and Chib (1998) and condition upon the states s_h to obtain the measurement equation as

$$\begin{aligned} y_t^* &= \ln\left(\varepsilon_t^2 + \overline{c}\right) - m_t = h + A_h f_{h,t} + \varepsilon_t^* \\ &= x_t^* a_h + \varepsilon_t^* \text{ where } x_t^* = \begin{bmatrix} I_n & \left(f'_{h,t}\otimes I_n\right)\end{bmatrix}. \end{aligned}$$

The term $\varepsilon_t^* + m_t$ is normal with mean vector m_t.[1] Let $y^* = \{y_t^*\}$ be the $Tn \times 1$ vector of stacked y_t^*, X^* be the $Tn \times nr_h$ matrix of stacked x_t^*. Finally, let Σ_h be the diagonal matrix in which the $(t+i,t+i)^{th}$ element is the variance of the i^{th} element of ε_t^* where $i \in \{1,\ldots,n\}$. Combining the likelihood with the prior $\mathcal{N}\left(0,\underline{V}_{a_h}\right)$ we can write $\overline{V}_h = \left[X^{*\prime}\Sigma_h^{-1}X^* + \underline{V}_{a_h}^{-1}\right]^{-1}$ and $\overline{a}_h = \underline{V}_h X^{*\prime}\Sigma_h^{-1}y^*$.

Finally we define the terms in the conditional posterior for $f_{h,t}$, $\mathcal{N}\left(\overline{f}_h,\overline{V}_{f,h}\right)$. Again conditional upon the states identified in s_h, the measurement equation can be written

$$y_t^{**} = \ln\left(\varepsilon_t^2 + \overline{c}\right) - m_t - h = A_h f_{h,t} + \varepsilon_t^*.$$

Let y^{**} be the $Tn \times 1$ vector of stacked y_t^{**}, X^{**} be the $Tn \times Tr_h$ matrix $\left(I_T \otimes A_h\right)$. Finally, let Σ_h be the diagonal matrix in which the $(t+i,t+i)^{th}$ element is the variance of the i^{th} element of $\varepsilon_t^*; i = 1,\ldots,n$. Combining the likelihood with the prior $\mathcal{N}\left(0,\left(R'_{r_h}R_{r_h}\right)^{-1}\right)$ we can write $\overline{V}_{f,h} = \left[X^{**\prime}\Sigma_h^{-1}X^{**} + R'_{r_h}R_{r_h}\right]^{-1}$ and $\overline{f}_h = \underline{V}_{f,h}X^{**\prime}\Sigma_h^{-1}y^{**}$.

In Specification 1, the conditional posteriors for a_α and a_h have the same form as that in Specification 2 except with $f_{\alpha,t}$ and $f_{h,t}$ replaced by $f_{\theta,t}$.

[1] The means and variances of the elements of $\varepsilon_t^* + m_t$ depend upon the states in s_h and are presented in Table 4 of Kim, Shephard and Chib (1998).

3 Tables and Figures

In this section we present a complete set of results obtained for the macroeconomic application discussed in Section 4 of the main text.

Table 1: DICs for models specified with $n = 8$ and various combinations of r_α and r_h. All values are relative to the DIC of the constant coefficient model (i.e. $r_\alpha = r_h = 0$).

3 states			5 states			7 states			10 states			12 states		
r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC
3	0	-402	5	0	-422	7	0	-351	10	0	-102	12	0	88
2	1	-443	4	1	-452	6	1	-349	8	2	-269	8	4	-200
1	2	-414	3	2	-486	4	3	-478	6	4	-358	7	5	-333
0	3	-334	2	3	-490	3	4	-475	5	5	-446	6	6	-360
			1	4	-408	1	6	-410	4	6	-483	5	7	-384
			0	5	-338	0	7	-336	2	8	-463	4	8	-442
shared		-263	shared		-68	shared		199	shared		441	shared		769

Table 2: DICs for models specified with $n = 15$ and various combinations of r_α and r_h. All values are relative to the DIC of the constant coefficient model (i.e. $r_\alpha = r_h = 0$).

3 states			5 states			7 states			10 states			12 states		
r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC
3	0	-764	5	0	-766	7	0	-742	10	0	-366	12	0	-140
2	1	-771	4	1	-816	6	1	-688	8	2	-486	8	4	-573
1	2	-711	3	2	-887	4	3	-892	6	4	-697	7	5	-655
0	3	-562	2	3	-851	3	4	-888	5	5	-854	6	6	-800
			1	4	-756	1	6	-698	4	6	-876	5	7	-792
			0	5	-583	0	7	-565	2	8	-800	4	8	-840
									0	10	-545	0	12	-577
shared		-770	shared		-835	shared		-719	shared		-418	shared		199

Figure 1: Impulse-response functions to non-news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

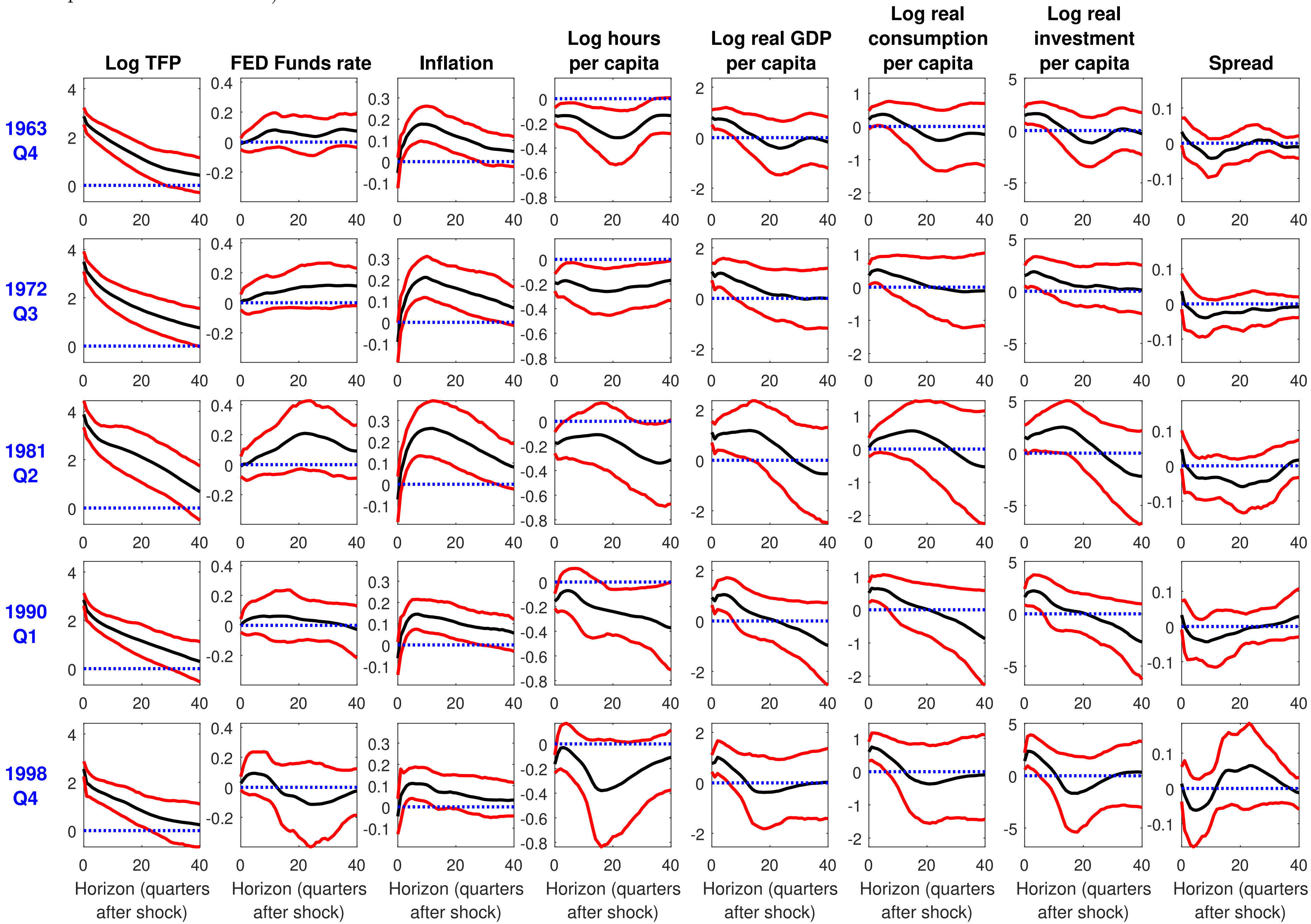

Figure 2: Impulse-response functions to news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

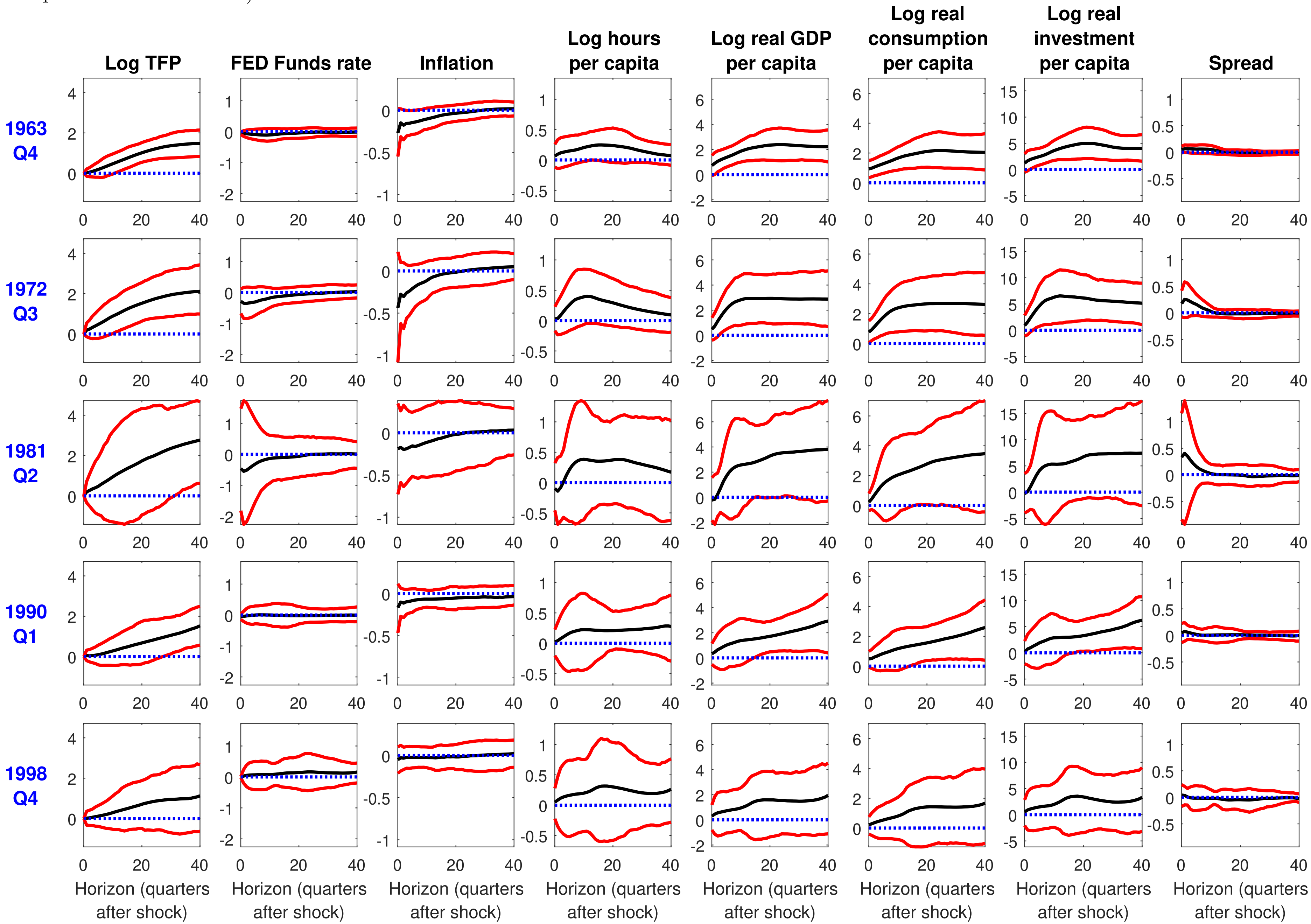

Figure 3: Time-varying responses to non-news and news shocks on impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

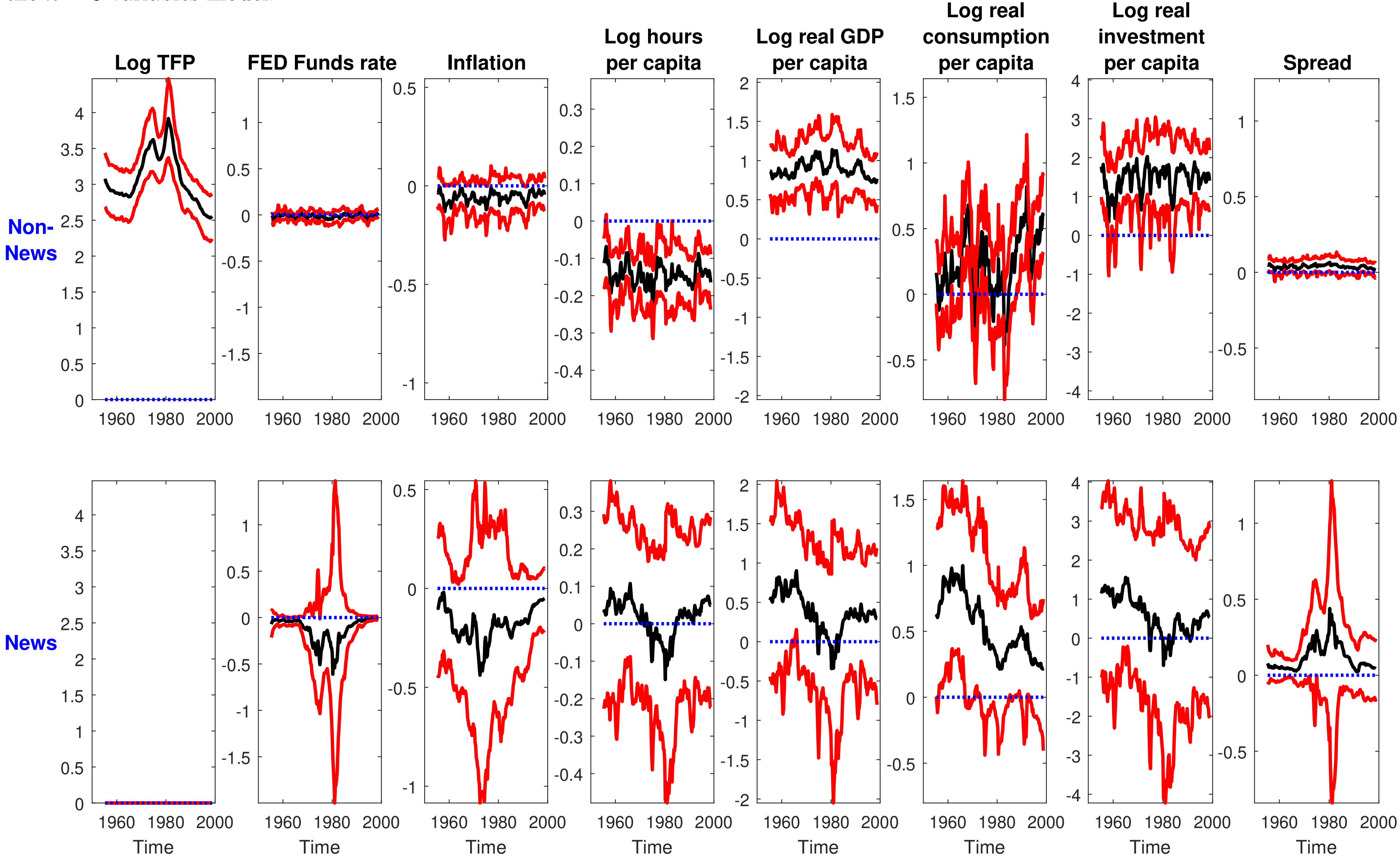

Figure 4: Time-varying responses to non-news and news shocks at 40 quarters after impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

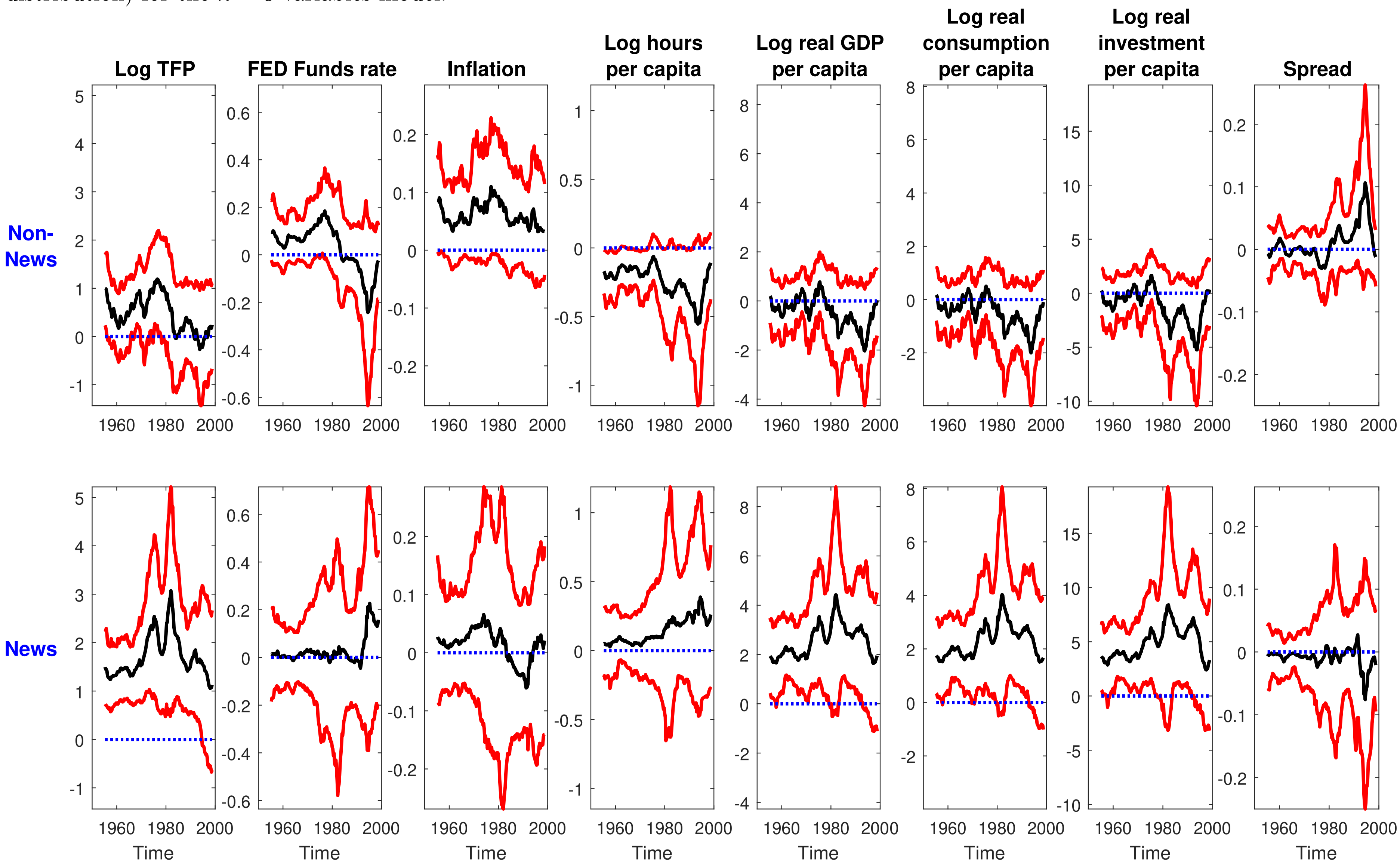

Figure 5: Fractions of forecast error variance explained by non-news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

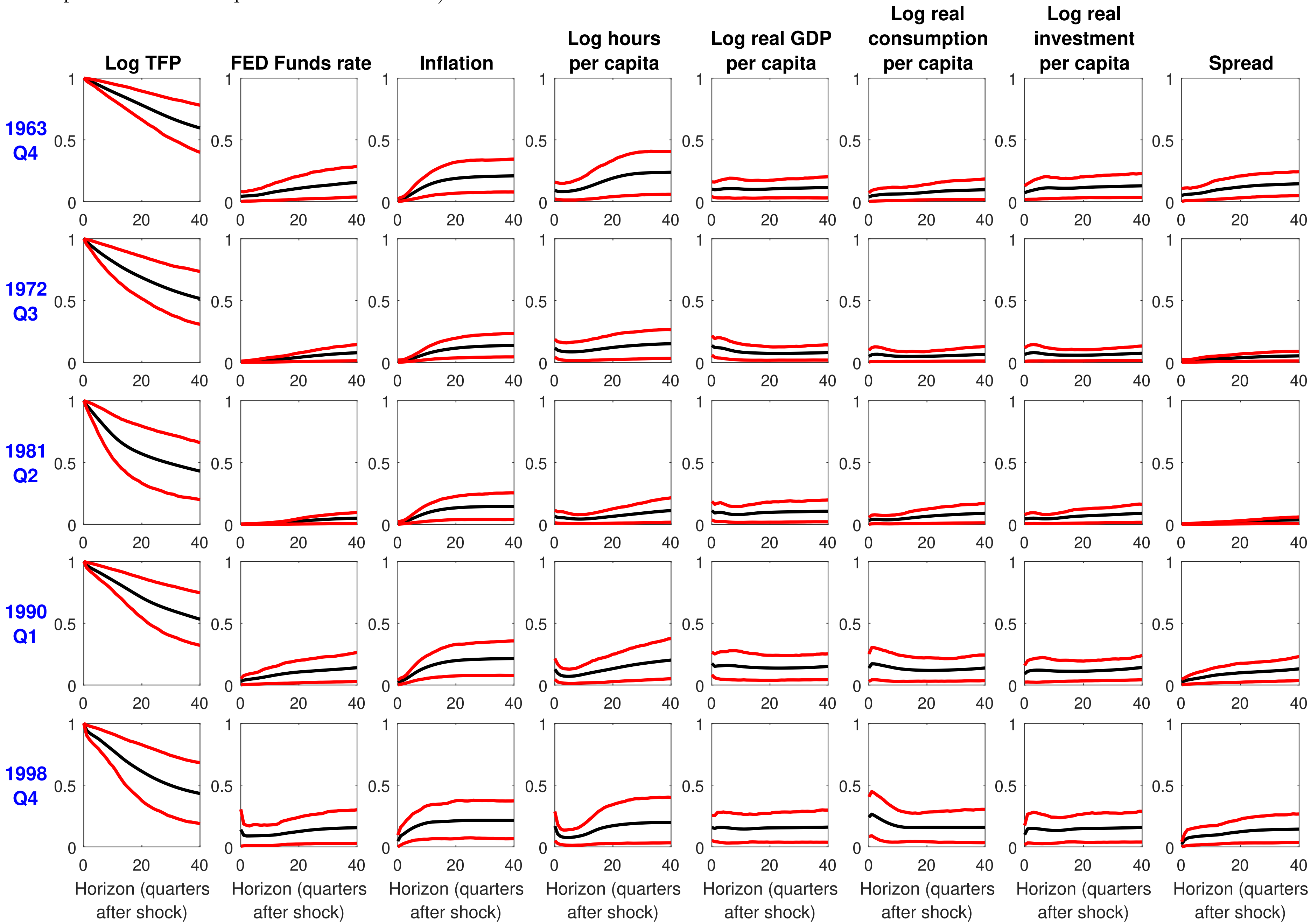

Figure 6: Fractions of forecast error variance explained by news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

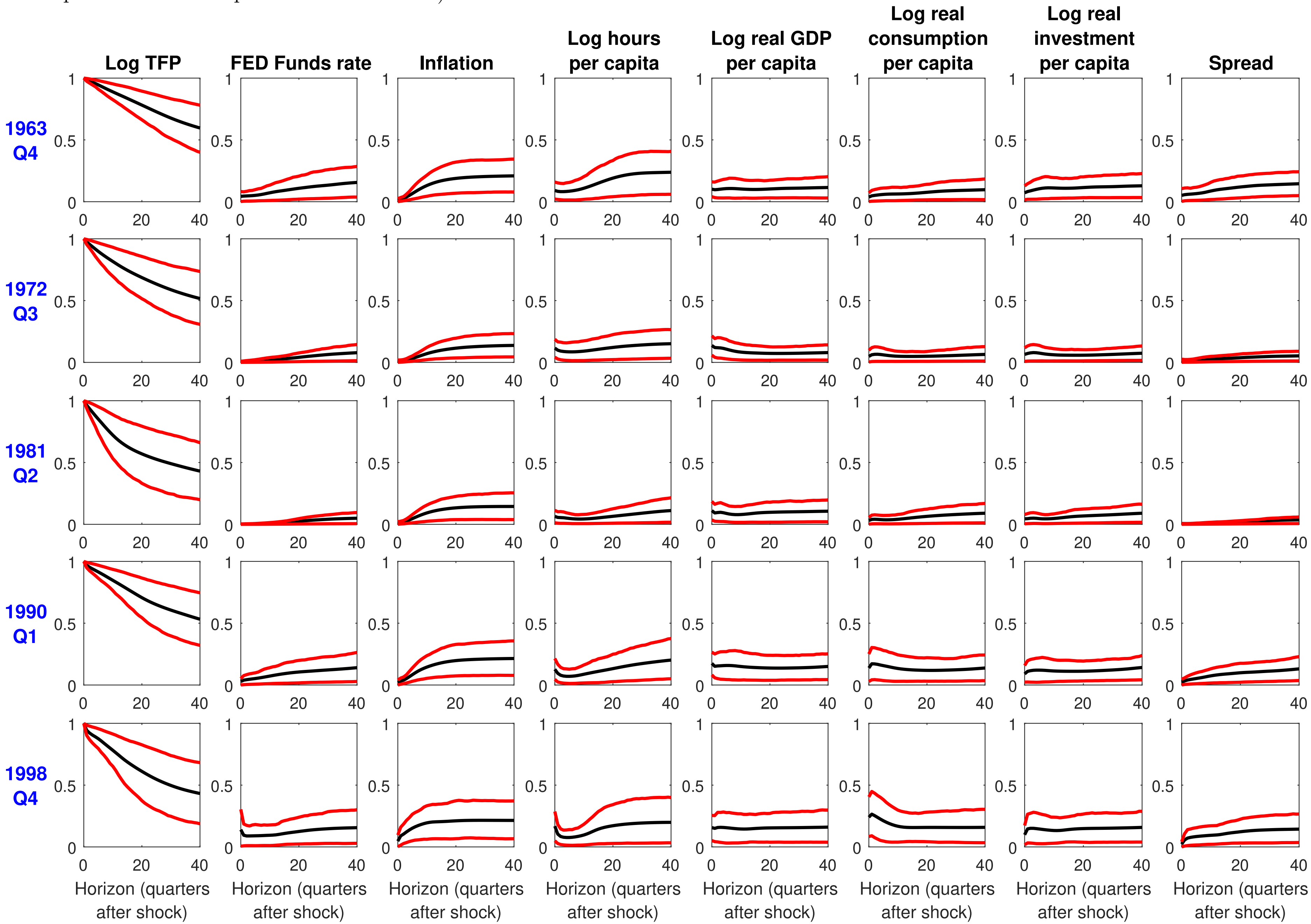

Figure 7: Time-varying fractions of forecast error variance explained by non-news and news shocks on impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

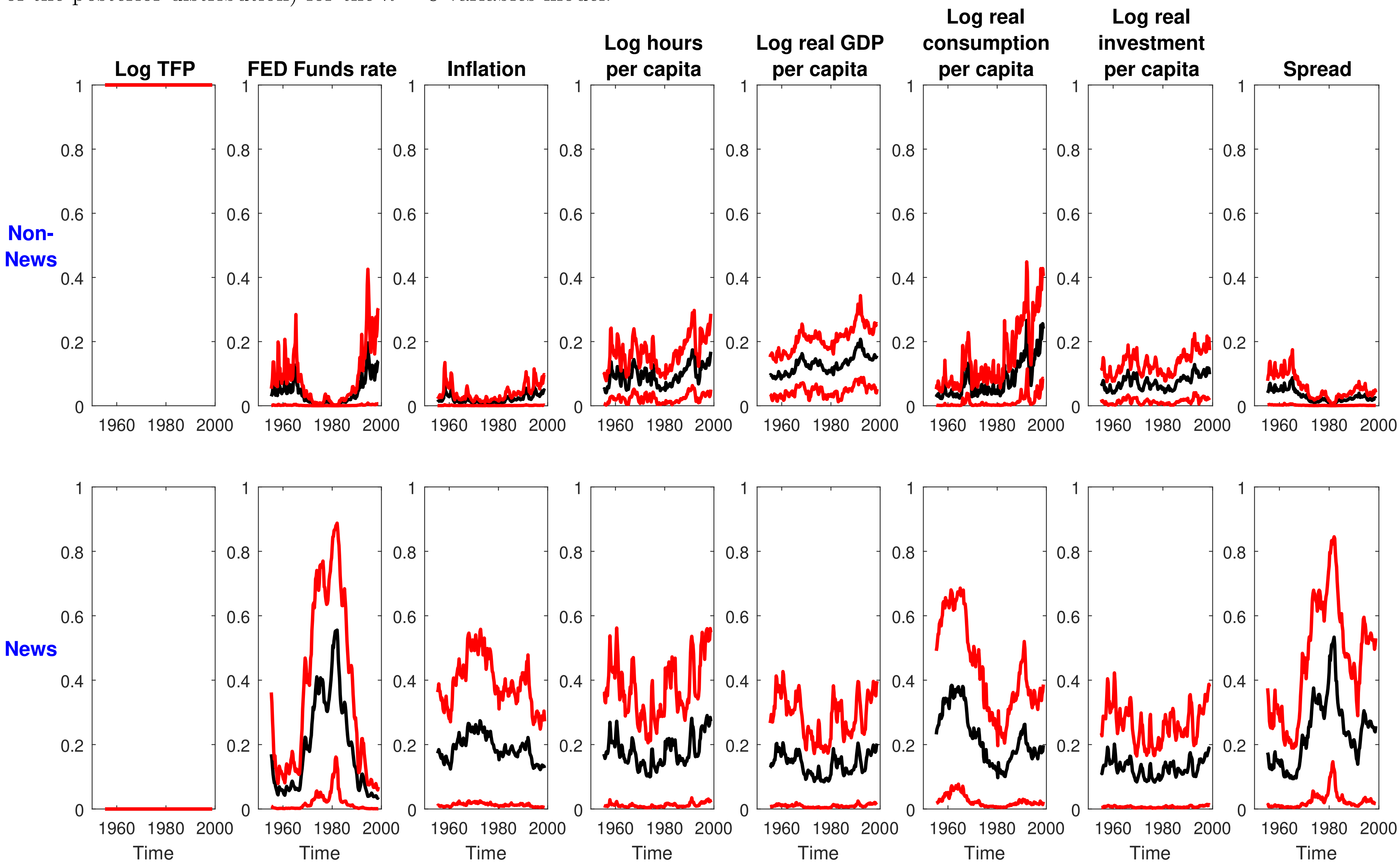

Figure 8: Time-varying fractions of forecast error variance explained by non-news and news shocks at 40 quarters after impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

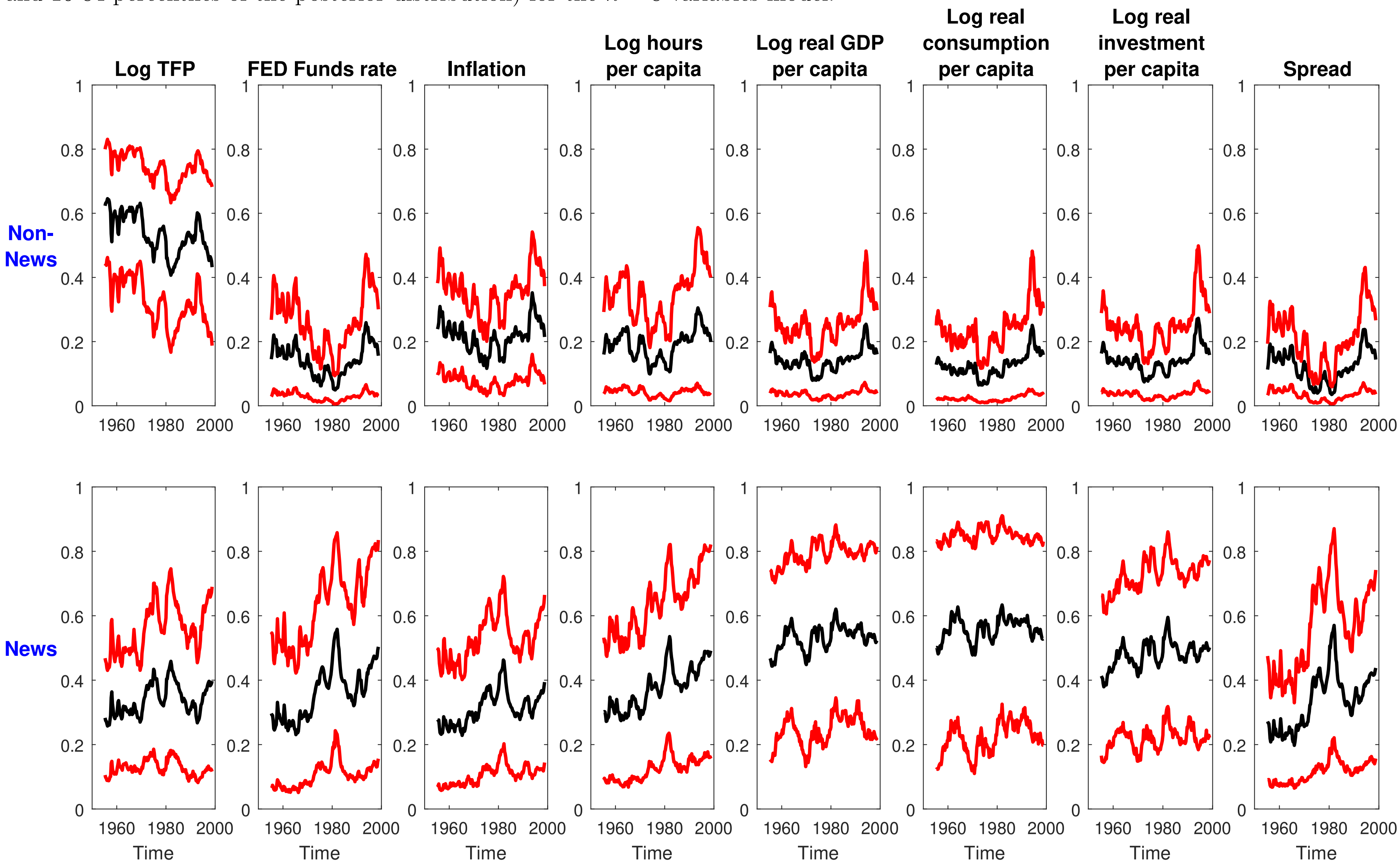

Figure 9: Impulse-response functions to non-news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

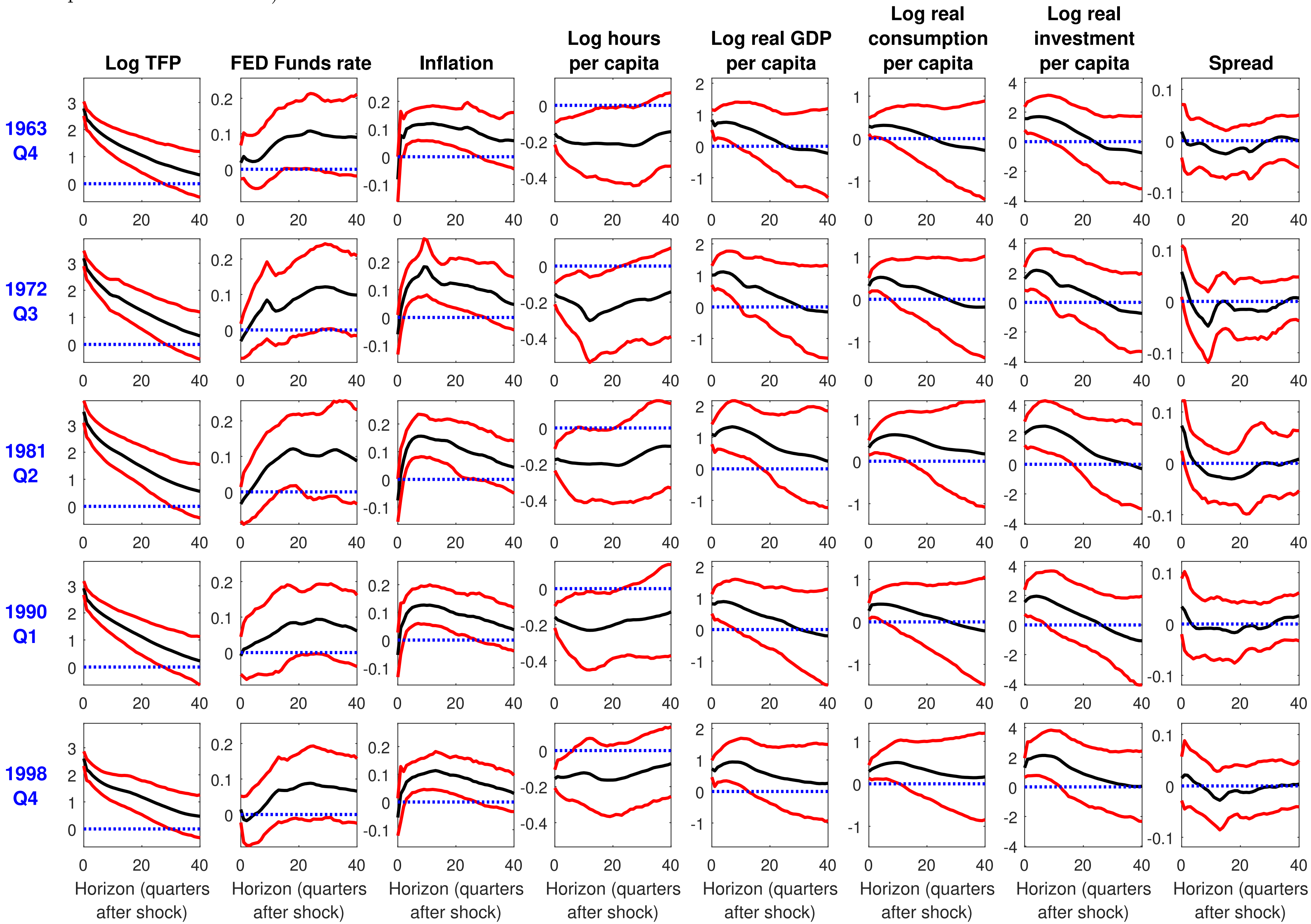

Figure 10: Impulse-response functions to news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

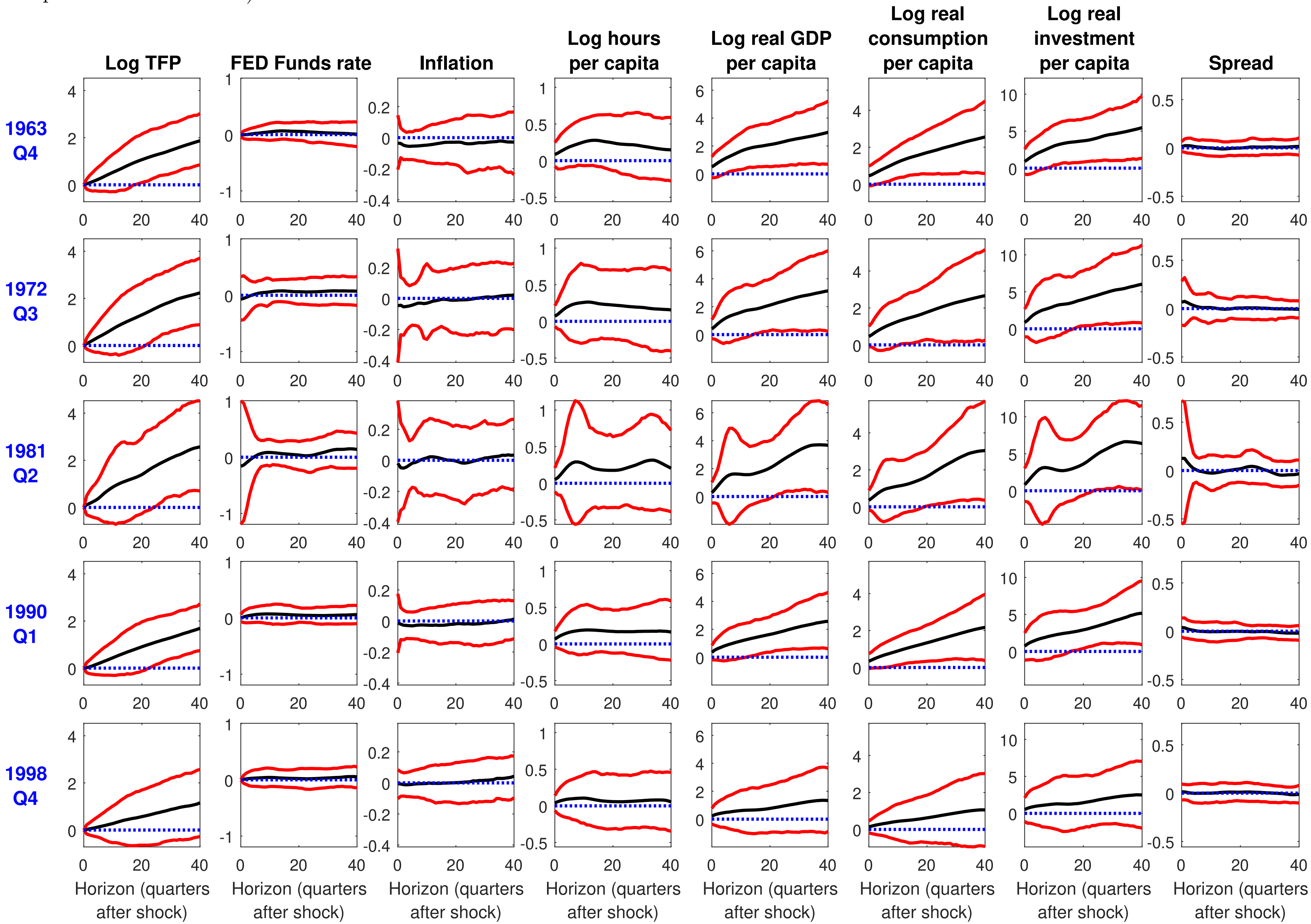

Figure 11: Time-varying responses to non-news and news shocks on impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

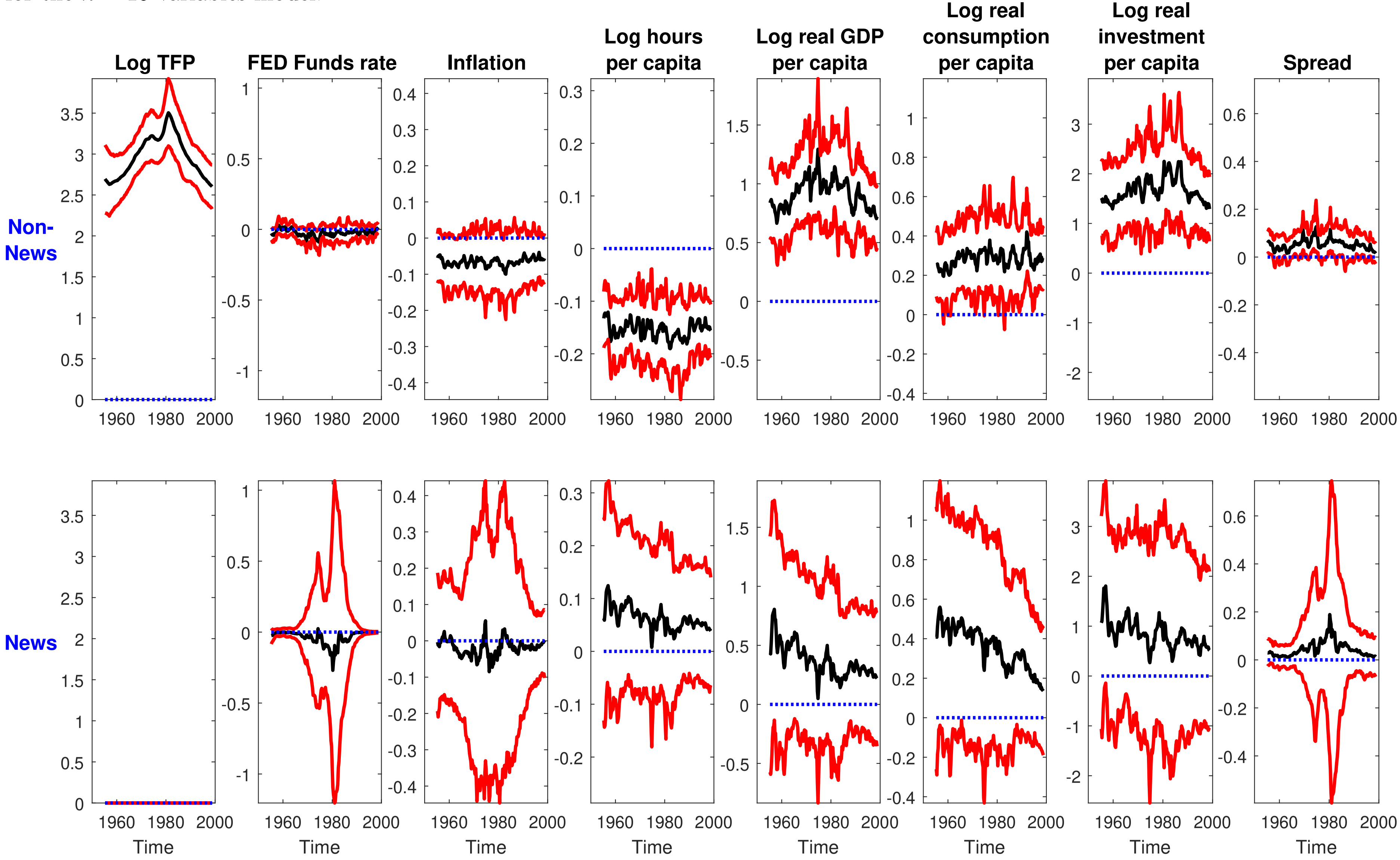

Figure 12: Time-varying responses to non-news and news shocks at 40 quarters after impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

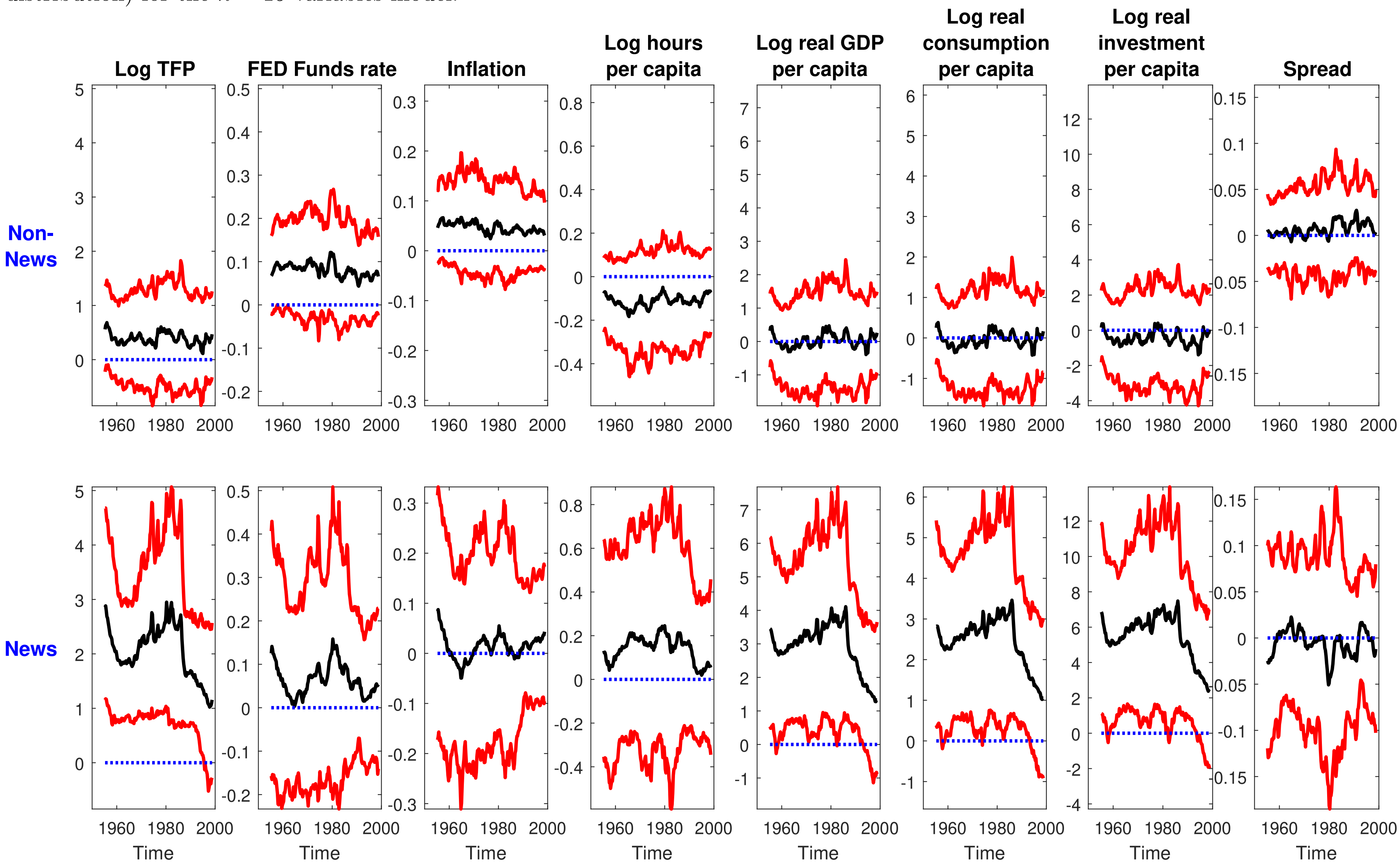

Figure 13: Fractions of forecast error variance explained by non-news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

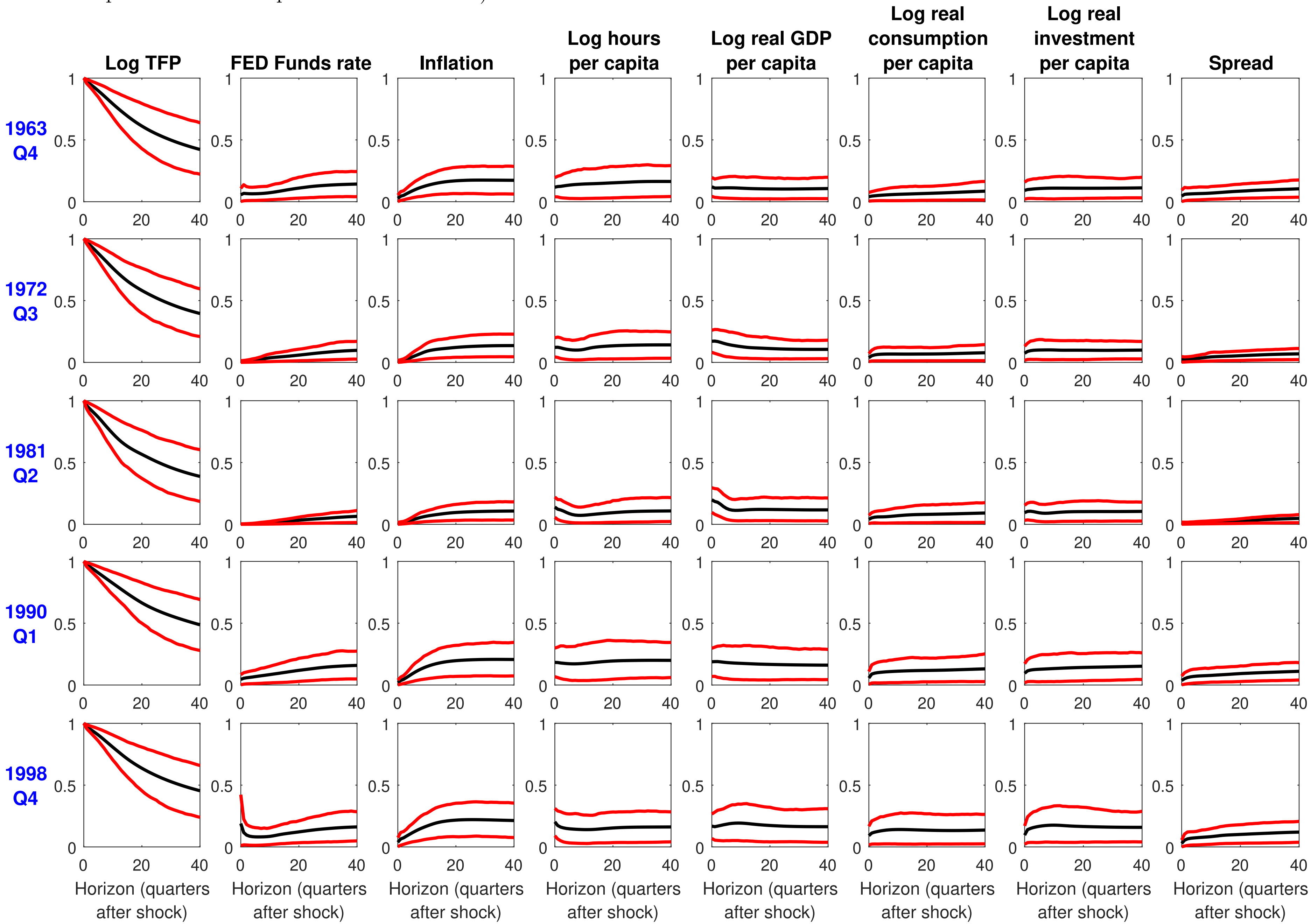

Figure 14: Fractions of forecast error variance explained by news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

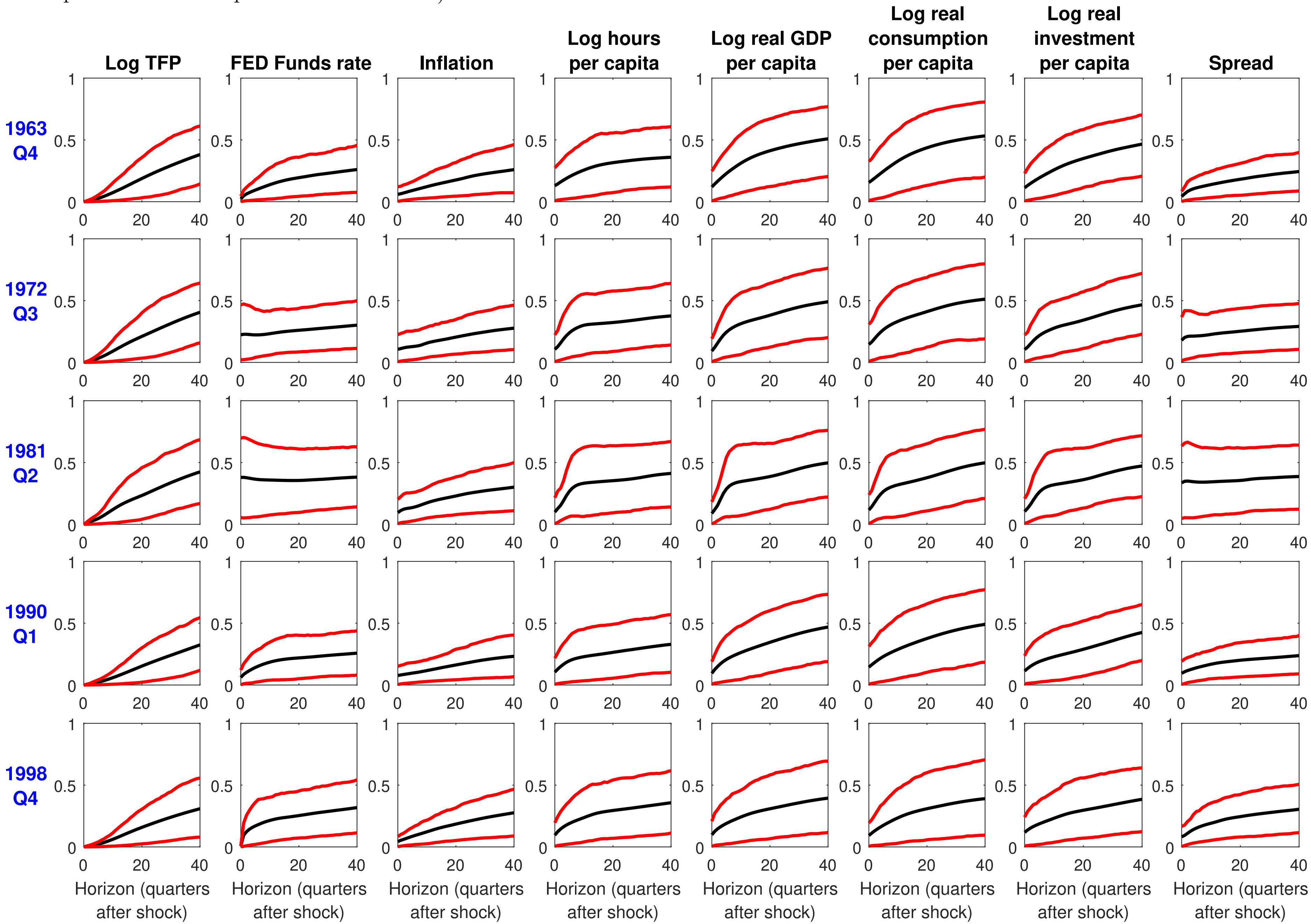

Figure 15: Time-varying fractions of forecast error variance explained by non-news and news shocks on impact (mean, and 16-84 percentiles of the posterior distribution) for the $n=15$ variables model.

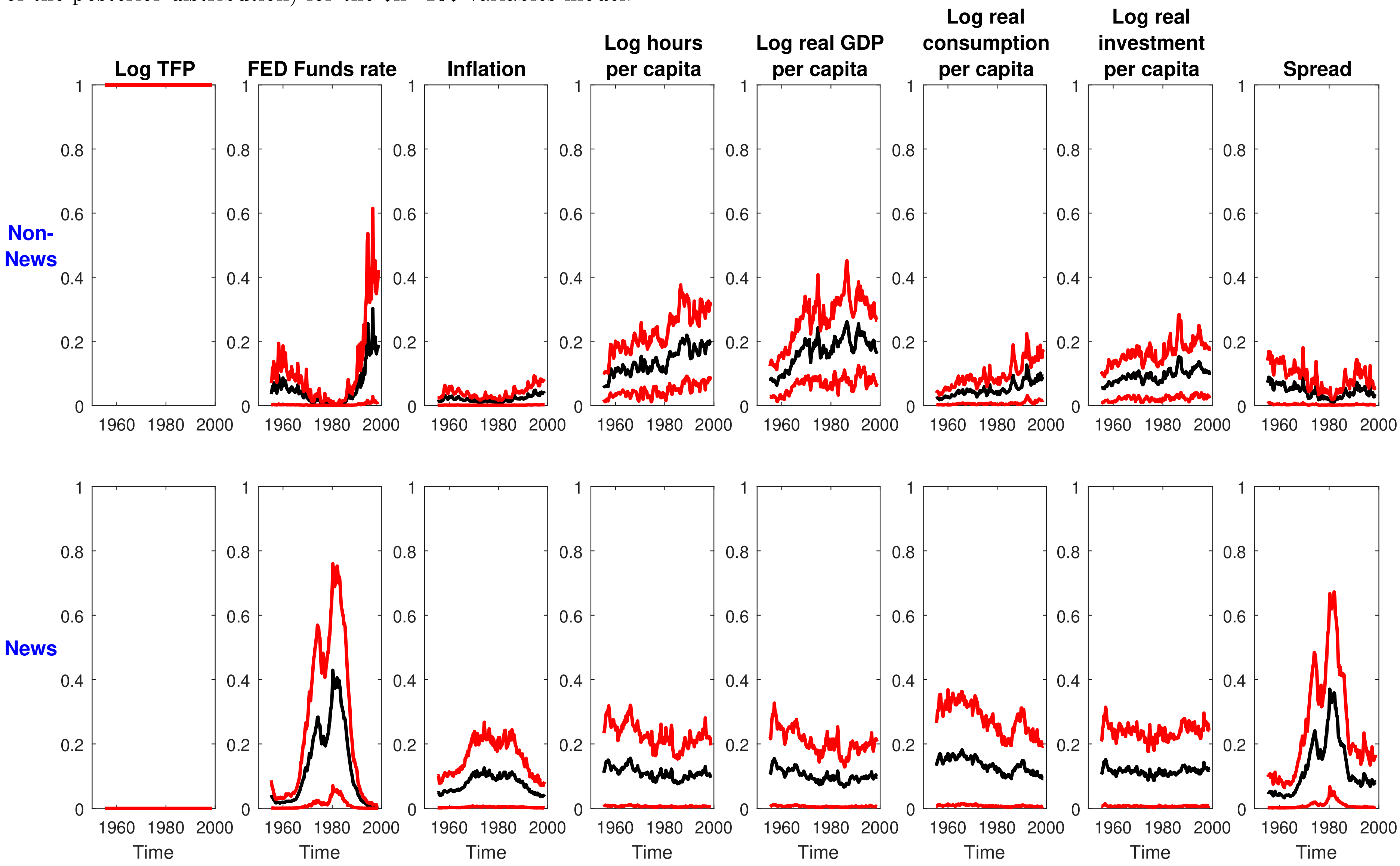

Figure 16: Time-varying fractions of forecast error variance explained by non-news and news shocks at 40 quarters after impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.