UNIVERSITÄT BERN

Faculty of Business, Economics and Social Sciences

Department of Economics

Searching for Hysteresis

Luca Benati, Thomas Lubik

21-07

May, 2021

DISCUSSION PAPERS

Schanzeneckstrasse 1
CH-3012 Bern, Switzerland
http://www.vwi.unibe.ch

Searching for Hysteresis*

Luca Benati
University of Bern†

Thomas Lubik
Federal Reserve Bank of Richmond‡

Abstract

Taking as data-generation process a standard DSGE model, we show via Monte Carlo that reliably detecting hysteresis, defined as the presence of aggregate demand shocks with a permanent impact on output, is a significant challenge, as model-consistent identification schemes (i) spuriously detect it with non-negligible probability when in fact the data-generation process features none, and (ii) have a low power to discriminate between alternative extents of hysteresis. We propose a simple approach to test for the presence of hysteresis, and to estimate its extent, based on the notion of simulating specific statistics (e.g., the fraction of frequency-zero variance of GDP due to hysteresis shocks) conditional on alternative values of hysteresis we impose upon the VAR, and then comparing the resulting Monte Carlo distributions to the corresponding distributions computed based on the actual data via the Kullback-Leibler divergence. Based on two alternative identification schemes, evidence suggests that post-WWII U.S. data are compatible with the notion of no hysteresis, although the most plausible estimate points towards a modest extent, equal to 7 per cent of the frequency-zero variance of GDP.

Keywords: Hysteresis, permanent shocks, long-run restrictions, sign restrictions, Bayesian methods; Kullback-Leibler divergence.
JEL Classification: E2, E3.

---

*We wish to thank Olivier Blanchard for comments; Francesco Furlanetto for sharing with us, at an early stage of this project, his work in progress on hysteresis; Juan Rubio-Ramirez for useful discussions, comments, and suggestions; and Rodney Strachan for extremely helpful suggestions on Bayesian cointegration. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Richmond, or of the Federal Reserve System.

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

‡Research Department, Federal Reserve Bank of Richmond, 701 E Byrd Street, Richmond, VA 23 218, USA. E-mail: Thomas.Lubik@rich.frb.org

1 Introduction

The term hysteresis captures the notion that economic disturbances that are typically regarded as transitory (e.g., monetary policy shocks) can have very long-lived, or even permanent effects. It is most often used within the context of the labor market, where long and deep recessions can lead to long-term unemployment and, through this channel, could raise the natural (or equilibrium) unemployment rate. Similarly, a broad array of demand shocks could have long-lasting effects on productivity and output through channels such as business formation or exit. The experience of the financial crisis and the Great Recession, and that of the ongoing recession caused by the COVID-19 pandemic, have rekindled interest in the possible presence of hysteresis effects in macroeconomic data. Although the theoretical channels for the transmission of hysteresis shocks are well-understood, empirical evidence is few and far between.

In this paper we search for hysteresis, which we define as the presence of aggregate demand shocks with a permanent impact on output, in post-WWII U.S. data. We identify aggregate demand and aggregate supply permanent GDP shocks by imposing a combination of zero restrictions on their long-run impact on GDP, and sign restrictions on both their short- and their long-run impacts on GDP and the price level. We provide three main contributions:

first, taking as data-generation process (DGP) a standard real business cycle (RBC) model featuring the possible presence of hysteresis effects (as discussed below), we show via Monte Carlo that reliably detecting hysteresis in macroeconomic data is a significant challenge. In particular, we show that model-consistent identification schemes (i) spuriously detect hysteresis with non-negligible probability when in fact the DGP features none, and (ii) have a low power to discriminate between alternative extents of hysteresis.

Second, in order to address these problems we propose a straightforward approach to test for the presence of hysteresis, and to estimate its extent, that is broadly conceptually related to Stock and Watson's (1996, 1998) TVP-MUB methodology for exploring the presence of random-walk time-variation in the data. The proposed approach is based on the notion of (1) simulating via Monte Carlo specific statistics (e.g., the fraction of frequency-zero variance of GDP due to hysteresis shocks) conditional on alternative extents of hysteresis that we impose upon estimated VARs, and then (2) estimating the extent of hysteresis in the data by comparing the thus obtained Monte Carlo distributions to the corresponding distributions computed based on the actual data. For example, a natural estimate of the extent of hysteresis—let us call it λ^*—is that particular value such that median of the Monte Carlo distribution of the fraction of frequency-zero variance of GDP due to hysteresis shocks simulated conditional on λ^* is closest to the distribution computed based on the actual data, where 'closeness' is defined in terms of the Kullback–Leibler divergence.

Third, when applying the proposed approach to the post-WWII United States, evidence based on Bayesian VARs (either in levels, or cointegrated), and two alternative identification schemes, suggests that the data are compatible with the notion of no hysteresis, although the most plausible estimate points towards a modest extent, equal to 7 per cent of the frequency-zero variance of GDP. Under this respect the proposed Monte Carlo-based correction makes a material difference to the results, with (e.g.) the simple estimates produced by the cointegrated VARs pointing towards a fraction of the long-horizon forecast error variance (FEV) of GDP explained by hysteresis shocks slightly in excess of 20 per cent.

1.1 Related literature

The classic paper on hysteresis is Blanchard and Summers (1986), which introduced the concept to the economics profession and used it in order to explain the persistence of European unemployment in the 1980s. Although they provided some basic statistical evidence, their main argument was largely based on a theoretical model of the labor market. In a similar vein, Ljungqvist and Sargent (1998) provided a more microfounded theoretical approach to the same issue, and offered corroborating evidence based on labor market data. More recently, theoretical frameworks based on the idea of endogenous TFP have also allowed for a possible role of hysteresis effects (see e.g., Anzoategui et al., 2019; or Jaimovich and Siu, 2020).

Empirical evidence in favor of or against hysteresis is somewhat sparse, arguably because of the difficulty in distinguishing permanent and highly persistent components in aggregate data. In fact, in his recent survey of the literature on the natural rate hypothesis, Blanchard (2018) concluded that the evidence is not sufficiently clear-cut to allow to reach strong conclusions. Cerra and Saxena (2008) showed that in a sample of 190 countries over the period 1960-2001 deep recessions permanently reduced the productive capacity of an economy. Ball (2009) used a simple Phillips curve framework to back out the effects of changes in inflation on unemployment, conditional on having observed large disinflations associated with deep recessions. Although he argued for the presence of hysteresis, in the end he did not distinguish between permanent and highly persistent effects. Galí (2015, 2020) took up Ball's analysis and integrated it within a New Keynesian model featuring an insider-outsider labor market framework as in Blanchard and Summers (1986). Based on a quantitative analysis of the model, he argued in favor of hysteresis as being an important driver of the European unemployment and wage and price inflation experience.

Furlanetto et al. (2021) is the paper that is closest to our work. They also use a structural VAR framework that combines long-run zero restrictions with short- and long-run sign restrictions. They do detect an important role for hysteresis effects in U.S. data. There are several differences between the present work and Furlanetto et al.'s (2021): in particular, they work with VARs in differences, whereas we work with either cointegrated VARs or VARs in levels, and they use a different set of variables. Of note is also the recent contribution by Jordà et al. (2020), who estimate a dynamic panel with local projections and detect large (compared with most contributions in the literature) effects of monetary policy shocks on GDP in the very long run.

The paper is organized as follows. The next section outlines our approach to searching for hysteresis effects on output, whereas section 3 shows Monte Carlo evidence on the performance of two alternative identification schemes. Section 4 discusses our proposed approach to address the problems highlighted by the Monte Carlo exercise. Section 5 discusses the empirical evidence for the post-WWII United States. Section 6 concludes.

2 Identifying Hysteresis Shocks

We assume that the permanent component of real GDP is driven by two, and only two shocks, possessing respectively aggregate supply (AS) and aggregate demand (AD) features. We label the two disturbances as 'BQ' (from Blanchard and Quah, 1989) and 'H' (for 'hysteresis'), respectively. The identifying restrictions we impose in order to disentangle the two disturbances are motivated by the standard AD-AS framework found in many macroeconomic textbooks, but they are also consistent with standard New Keynesian models.

The motivation behind our restrictions can be briefly outlined as follows. We assume that

(I) the AD curve is downward-sloping both in the short and in the long run, whereas

(II) the AS curve is upward-sloping in the short run, and it is vertical in the long run. We further assume that

(III) BQ shocks only affect the AS curve, whereas

(IV) H shocks, which do affect the AD curve, may or may not also affect the AS curve depending on whether there are, or there are not, hysteresis effects. In the case of hysteresis, a negative (positive) H shock has a negative (positive) permanent impact on the long-run AS curve.

The first three assumptions are standard, and are consistent with a wide array of macroeconomic frameworks ranging from the textbook AD-AS model, to simple New Keynesian models, and large-scale macroeconomic models used in policy institutions. The fourth assumption captures the essence of the notion of hysteresis. In such a case, a negative (positive) AD shock has a negative (positive) permanent impact on long-run aggregate supply. Typical examples are the notion that deep recessions permanently scar the economy by reducing its potential productive capacity, either by increasing the equilibrium unemployment rate or by reducing firm entry, and thereby long-run productivity. Alternatively, 'running the economy hot' could permanently attract people into the labor force.

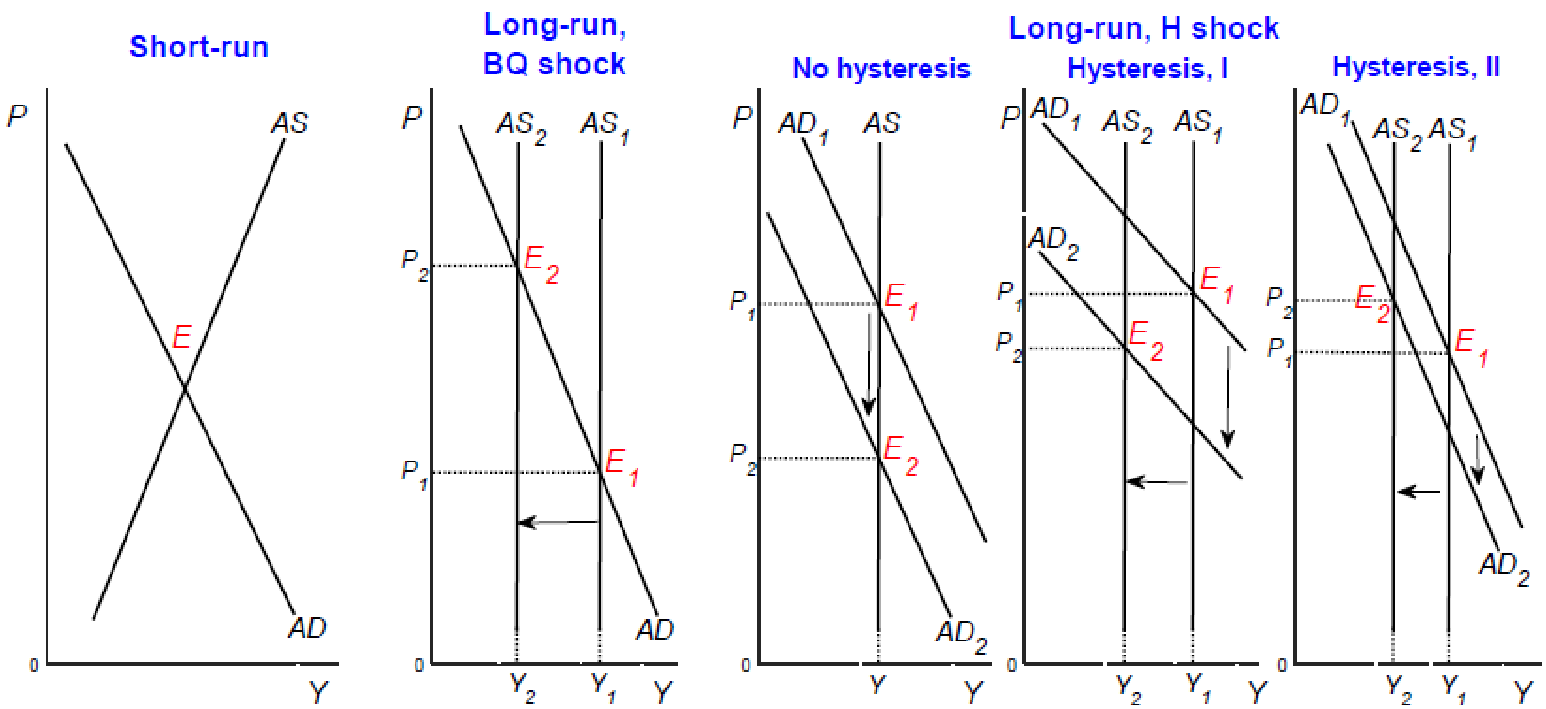

Figure 1 A simple illustration of the identifying assumptions

Figure 1 provides a simple illustration of assumptions (I)-(IV), and it allows for a straightforward discussion of our identifying restrictions.

In the output (Y)-price level (P) space, in the *short run* (first panel) BQ shocks, by shifting the AS curve, and therefore moving the economy along the AD curve, cause output and prices to move in opposite directions. By the same token, H shocks, by moving the economy along the AS curve, cause output and prices to move in the same direction. The first set of restrictions we will impose in order to disentangle the two shocks is therefore that both on impact, and up to subsequent h quarters, BQ shocks generate, for output and prices, impulse responses with *opposite signs*, whereas H shocks produce impulse responses with the *same sign*.[1]

Turning to the *long run*, BQ shocks (second panel), by assumption (III), only impact upon the AS curve, and therefore produce permanent impacts on output and prices with *opposite signs*. As for H shocks there are three possible cases:

(i) *no hysteresis* (third panel): in this case the long-run impact on output is *zero*, whereas the *sign* of the long-run impact on prices is the *same* as the sign of the short-run impact.

(ii) *Hysteresis* (last two panels): here a negative (positive) H shock shifts the AD curve down and to the left (up and to the right), and the long-run AS curve to the left (right). As a result, the long-run impact on output is unambiguously negative (positive), so that its sign is the *same* as the sign of its short run response to the H shock. The long-run impact on prices, on the other hand, is ambiguous, and it crucially depends on (a) the size of the hysteresis effect on output, and (b) the slope of the AD curve. In particular, the flatter the AD curve, and the smaller the hysteresis effect on output, the greater the likelihood that a negative (positive) H shock causes a decrease (increase) in prices. In the limit (third panel), if the hysteresis impact on output tends to zero the long-run impact on prices of a negative (positive) H shock is unambiguously negative (positive). In general, however, this cannot be assumed. In spite of this, since many researchers may find the assumption that a negative (positive) H shock has a negative (positive) long-run impact not only on output, but also on prices, more plausible, and intuitively sensible,[2] in what follows we will consider two alternative sets of results, obtained by imposing, and respectively not imposing such restriction. To anticipate, the two sets of results are qualitatively the same, so that, in practice, imposing or not imposing this restriction does appear to make a material difference.

[1] These restrictions are the same as Canova and Paustian's (2011) DSGE-based 'robust sign restrictions' (see their Table 2, page 348). E.g., Canova and Paustian's (2011) DSGE model features two demand-side shocks, labelled as 'Monetary' and 'Taste', respectively, with the taste shocks being essentially a non-monetary demand-side shock. For both shocks, the impacts on GDP and inflation have the same sign.

[2] E.g., Furlanetto et al. (2021) do impose such restriction.

Table 1 Identifying restrictions

Impact on:	*Shock:*				
	BQ	H	BQ	H	
	Short run:		Long run:		
Output	+	+	+	0	+
Price level	−	+	−	+	?/+

Table 1 summarizes the previously discussed restrictions. We now turn to discussing two alternative empirical strategies for identifying BQ and H shocks in the data, and to Monte Carlo evidence on their comparative effectiveness.

3 Two Alternative Empirical Strategies

Working with Bayesian VARs, either cointegrated or in levels, we identify BQ and H shocks by imposing the restrictions summarized in Table 1 based on two alternative empirical strategies, which we label, respectively, as Scheme I and Scheme II.

Scheme I jointly identifies *both* a BQ *and* an H shock. So, when working with cointegrated VARs identified via a combination of zero long-run restrictions on GDP, and both short- and long-run sign restrictions on GDP and the price level, this scheme identifies *two, and only two* permanent GDP shocks, one being BQ and the other being H. Although this would appear as the natural and logical way of imposing the restrictions in Table 1, as we show below via Monte Carlo, taking as DGP a standard RBC model featuring the possible presence of hysteresis, Scheme I suffers from the shortcoming that it spuriously detects hysteresis with a sizeable probability when in fact the DGP features none. The reason for this is straightforward: since this scheme jointly identifies both a BQ *and* an H shock, in fact it *imposes* upon the data the very existence of hysteresis by 'brute force'. This motivates an alternative approach to imposing the identifying restrictions in Table 1.

Whereas Scheme I identifies both a BQ and an H shock, *Scheme II* identifies a *single shock*, which ought to be *either* BQ *or* H. So, focusing again on cointegrated VARs, this scheme identifies a single permanent GDP shock, which ought to satisfy the restrictions in Table 1 pertaining to either of the two disturbances. There are two crucial differences with Scheme I. First, whereas Scheme I *imposes* the existence of hysteresis upon the data, Scheme II only *allows* it: in fact, in principle all of the models obtained via Scheme II could feature BQ shocks. Because of this, as we show via Monte Carlo, this scheme exhibits a significantly smaller tendency to spuriously identifying hysteresis when the DGP features none. Second, whereas the posterior distribution produced by Scheme I pertains to a specific model, the distribution produced by Scheme II pertains to two different types of models.

As we will in Section 5, in spite of such stark difference between the two schemes, when applied to the data via the Monte Carlo-based approach discussed in Section 4 they produce near-identical results.

3.1 Monte Carlo evidence on the performance of the two strategies

3.1.1 The data-generation process

The DGP we use in order to assess the reliability of the two alternative strategies is described in detail in Online Appendix A. In essence, the model is the one described in Galí's (2015) Chapter 2, augmented with habit-formation in consumption and a unit root in technology (and therefore in the natural level of output), and featuring the possible presence of hysteresis effects. Households solve the following problem

$$U_0^* \equiv \underset{C_t, N_t}{\text{Max}}\ E_0 \sum_{t=0}^{\infty} \beta^t \left[\ln(C_t - bC_{t-1}) - \frac{N_t^{1+\phi}}{1+\phi} \right] \tag{1}$$

subject to

$$P_t C_t + Q_t B_t = B_{t-1} + W_t N_t - T_t \tag{2}$$

where C_t and N_t are real consumption and hours worked, respectively; $0<b<1$ is the habit-formation parameter; P_t is the price of consumption goods; B_t is the stock of nominal bonds; Q_t is the price, at time t, of a nominal bond paying 1 dollar at time t+1; W_t is the nominal wage; T_t are nominal lump-sum taxes; and the rest of the notation is obvious. As for the firms we exactly follow Galí (2015), with the only difference pertaining to the process for technology. Firms produce output (Y_t) via the production function $Y_t = A_t N_t^{1-\alpha}$, with $0<\alpha<1$, A_t being technology, and the capital stock being constant and normalized to 1, and maximize profits, $P_t Y_t - W_t N_t = P_t A_t N_t^{1-\alpha} - W_t N_t$ with respect to N_t. As for the logarithm of technology, $a_t = \ln(A_t)$, we postulate that it evolves according to

$$a_t = a_{t-1} + \epsilon_t^a + \delta \tilde{y}_{t-1} \tag{3}$$

with $\epsilon_t^a \sim N(0, \sigma_a^2)$ being the technology shock, $\tilde{y}_t$ being the transitory component of output, and $\delta \geq 0$ capturing the possible presence of hysteresis effects. If $\delta = 0$ there is no hysteresis, whereas if $\delta > 0$ positive (negative) transitory output fluctuations cause subsequent permanent increases (decreases) in a_t. This specification, which is conceptually the same as the one used by Jordà et al. (2020), is designed to capture the notion that positive (negative) deviations of GDP from potential (here, deviations of output from its stochastic trend A_t) may have a positive (negative) impact on potential GDP itself. Finally, monetary policy is described by a standard Taylor rule with smoothing,

$$R_t = \rho R_{t-1} + (1-\rho)\phi_\pi \pi_t \tag{4}$$

where the notation is obvious.

Online Appendix A reports the log-linearized equations of the stationarized model (both output and the real wage inherit indeed the unit root in a_t, and therefore ought to be stationarized for standard solutions methods for linear rational expectations models, such as Sims' (2002), to be applicable). This is accomplished by defining the stationarized variables $\tilde{Y}_t \equiv Y_t/A_t$ and $\tilde{\Omega}_t \equiv (W_t/P_t)/A_t$, with $\tilde{y}_t$ and $\tilde{\omega}_t$ being their respective log-deviations from the steady-state, so that $\tilde{y}_t$ is the component of output that is driven by the transitory disturbances.

We calibrate the structural parameters as follows: β=0.99, α=1/3, b=0.8, ϕ=1, ρ=0.9, ϕ_π=1.5, and σ_a=0.007.[3] We augment the model with three transitory AR(1) disturbances, two of them (v_t and u_t) supply-side, and one (e_t) demand-side, and an additive disturbance to the log of the production function, $z_t \sim N(0, \sigma_z^2)$, with σ_z=0.005. The AR parameters are calibrated to ρ_v=ρ_u=ρ_e=0.75, whereas the standard deviations of the disturbances' innovations ϵ_t^v, ϵ_t^u, and ϵ_t^e (all zero-mean, and normally distributed) are set to σ_u=0.001, σ_v=0.005 and σ_e=0.045. Based on this calibration the permanent technology shock (ϵ_t^a) explains exactly 1/3 of the forecast error variance (FEV) of log GDP on impact, and with δ=0 it explains slightly more than 96 per cent of GDP's FEV 15 years ahead. These figures are broadly in line with the evidence produced by the structural VAR literature.[4] Further, based on these values of the structural parameters the demand-side disturbance ϵ_t^e is very close to being the only driver of the transitory component of output, so that the identifying restrictions in Table 1 are, for practical purposes, correct. Finally, as for δ in the Monte Carlo exercise we consider a grid of values, from δ=0 (no hysteresis) to δ=0.1256, for which the demand-side shock ϵ_t^e explains exactly 20 per cent of the frequency-zero variance of output, as at long horizons the other three shocks explain vitually *nil* of output's variation.

Figure 2 shows, for output and prices, the IRFs in response to ϵ_t^a and ϵ_t^e, together with the fractions of FEV of output and inflation explained by the two shocks, for both δ=0 and δ=0.1256.[5] The evidence in the figure suggests that the RBC model is an appropriate DGP for assessing the ability of the identifying restrictions in Table 1 to effectively recover the authentic extent of hysteresis in the data (if any). First, for either of the two values of δ (the same holds for any value in between) the two shocks explain sizeable, or even dominant fractions of the FEV of either variable at least over some non-negligible horizon. For example, even for δ=0 ϵ_t^e is the dominant driver of output at short horizons, and of prices at medium-to-long horizons. Second, from the IRFs it is apparent how ϵ_t^a and ϵ_t^e satisfy the restrictions for BQ and H shocks, respectively, reported in Table 1. In particular, with δ=0.1256—and in fact for any δ>0[6]—the long-run impact of a positive ϵ_t^e (i.e. H) shock on the price level is positive, instead of being ambiguous as in the last two panels of Figure 1. Accordingly, in the

[3]The value for σ_a is based on Watson's (1986, p. 60) estimate of the standard deviation of shocks to the stochastic trend of log real GDP. The rationale is that, within the present context, a_t is the random-walk component of log real GDP.

[4]For example, as for the fraction of the FEV of GDP explained by permanent shocks on impact see Table I.2 of Cochrane (1994).

[5]Figures A.1 and A.2 in the Online Appendix report the full set of IRFs and fractions of FEV for output, the price level, hours, real wage, and technology.

[6]This evidence is available upon request.

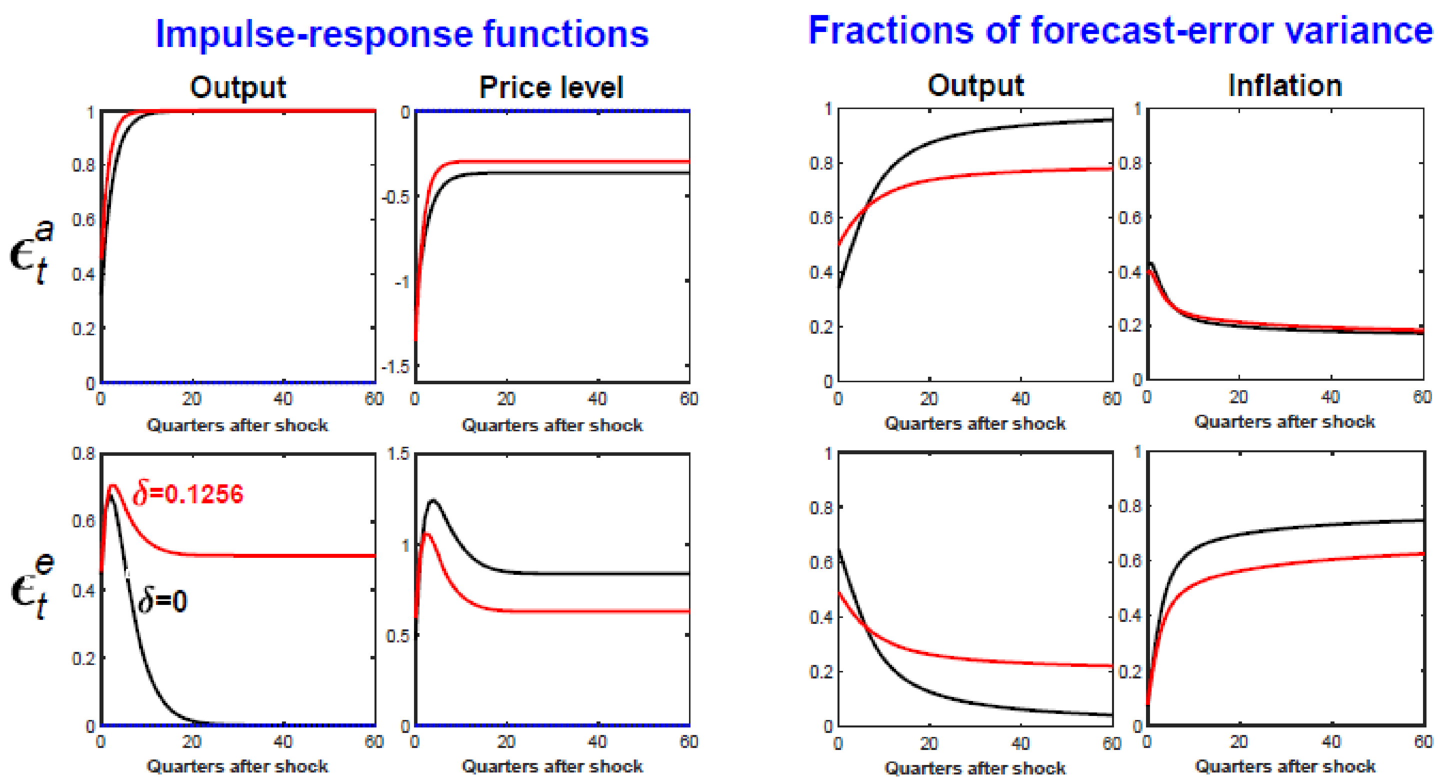

Figure 2 Theoretical impulse-response functions and fractions of forecast error variance for the RBC model for output and prices, for ϵ_t^a (BQ) and ϵ_t^e (H) shocks

Monte Carlo exercises below we will impose this restriction in the identification of H shocks.

Although we could have considered more complex models (in particular, New Keynesian models featuring a capital stock and a rich array of real frictions), and more complex formalizations of the notion of hysteresis, we have chosen to adopt the present setup for one main reason. Our objective here is simply to illustrate, within a simple and transparent setting, the two problems already mentioned in the Introduction, i.e. that model-consistent identification schemes spuriously detect hysteresis with a sizeable probability when the DGP features none, and they have a low power to discriminate between alternative extents of hysteresis. For such purpose the present DGP is perfectly appropriate.[7]

3.1.2 Evidence

We simulate the model for samples of length equal to that of the actual sample we will work with in Section 5, i.e. 261 observations.[8] Based on each artificial sample we then estimate Bayesian VARs for output (y_t), the real wage (ω_t), hours (n_t), inflation (π_t), and technology (a_t). Whereas n_t and π_t are I(0), y_t and ω_t, as mentioned, inherit the unit root in a_t, and they are both cointegrated with log technology with cointegration vector $[1, -1]'$. In what follows we therefore work with either $Y_{1,t} = [y_t, \omega_t, a_t, \pi_t, n_t]'$ or $Y_{2,t} = [\Delta y_t, (y_t\text{-}a_t), (\omega_t\text{-}a_t), \pi_t, n_t]'$. In the former case, by the arguments in Sims, Stock and Watson (1990) the VAR in levels effectively captures, in principle, cointegration between a_t and either y_t and ω_t. In the latter case, on the other hand, cointegration is imposed directly into the estimated system.[9] For each Monte Carlo simulation, and either $Y_{1,t}$ or $Y_{2,t}$, we choose the lag order as the maximum between the lag orders chosen by the Schwartz and Hannan-Quinn criteria, based on VARs estimated via OLS. For simplicity, in what follows we will refer to the VAR for $Y_{1,t}$ as the 'VAR in levels', and to that for $Y_{2,t}$, with a slight abuse of notation, as the 'VAR in differences'.

We estimate the Bayesian VARs as in Uhlig (1998, 2005). Specifically, we exactly follow Uhlig's approach in terms of both distributional assumptions (the distributions for the VAR's coefficients and its covariance matrix are postulated to belong to the Normal-Wishart family) and of priors. For estimation details the reader is therefore referred to either the Appendix of Uhlig (1998), or to Appendix B of Uhlig (2005).

For each value of δ and either VAR specification we perform 10,000 Monte Carlo simulations, and based on each of them we estimate the Bayesian VARs by taking

[7]In fact, we checked the robustness of the results reported in the next sub-section based on a standard New Keynesian model with backward- and forward-looking components in both the IS and the Phillips curve, and technology still evolving according to (3). The results (available upon request) are qualitatively the same.

[8]For each simulation we run a pre-sample of 100 observations that we then discard.

[9]See the discussion in (e.g.) Cochrane (1994) about alternative ways of estimating cointegrated systems.

10,000 draws from the posterior distribution.

Imposing the correct identifying restrictions with no hysteresis We start by checking that, conditional on the model featuring *no hysteresis* (i.e., with δ=0) the VARs correctly recover the main features of the DGP—in terms of impulse-response functions (IRFs), and fractions of FEV)—when we impose the correct identifying restriction, i.e. that the DGP features a single shock (ϵ_t^a) with a permanent impact on output. This is shown in Figure A.3 in the Online Appendix based on VARs in levels (the corresponding evidence based on VARs in differences is qualitatively the same, and it is available upon request). When working in levels we identify ϵ_t^a as in Uhlig (2003, 2004), i.e. as the most powerful shock for y_t at a 'long' horizon, which we set to 15 years ahead,[10] whereas when working in differences we identify it as in Blanchard and Quah (1989), i.e. as the only shock having a permanent impact on y_t. In brief, the evidence in Figure A.1 (as well as the corresponding evidence for the VARs in differences) shows that either VAR recovers with great precision the main features of the DGP, in terms of both IRFs and fractions of FEV. This is the case, in particular, for the main series of interest within the present context, i.e. output.

Imposing Schemes I and II with no hysteresis Conditional on the DGP featuring no hysteresis, we then explore to which extent Schemes I and II tend to spuriously detect it. Based on the VAR in differences we consider both schemes, whereas based on the VAR in levels we only consider Scheme II.[11] Based on the VARs in differences we implement Scheme I by jointly imposing the zero long-run restrictions, and the short- and long-run sign restrictions, based on the by now standard methodology proposed by Arias et al. (2018). We impose the short-run sign restrictions both on impact and for the subsequent eight quarters. As for Scheme II we identify, for each draw from the posterior distribution, the only shock that has a permanent impact on GDP, and we check whether it satisfies both the short- and the long-run restrictions characterizing either BQ or H shocks. If the shock is identified as either BQ or H we keep the draw (i.e. the model), otherwise we discard it. Finally, based on the VARs in levels we implement Scheme II by identifying, for each draw from the posterior, the most powerful shock for y_t at the 15 years horizon. If the shock is identified as either BQ or H we keep the draw, otherwise we discard it.[12]

[10] We experimented with alternative horizons, from 10 to 25 years ahead, and they tend to produce near-identical results (this evidence is available upon request).

[11] The reason is that, to the best of our knowledge, no algorithm has been developed for combining sign restrictions with Uhlig (2003, 2004)-style identification of the most powerful shock(s) at a specific 'long' horizon.

[12] One concern with working with VARs in levels and Uhlig (2003, 2004)-style identification of the most powerful shock at a specific long-horizon is that, in principle, the two sets of models featuring BQ and H shocks could explain systematically different fractions of the long-horizon FEV of output. (This is obviously not a concern when working with VARs in differences: since both shocks are identified as the only permanent shock for output, by construction they both explain 100 per cent

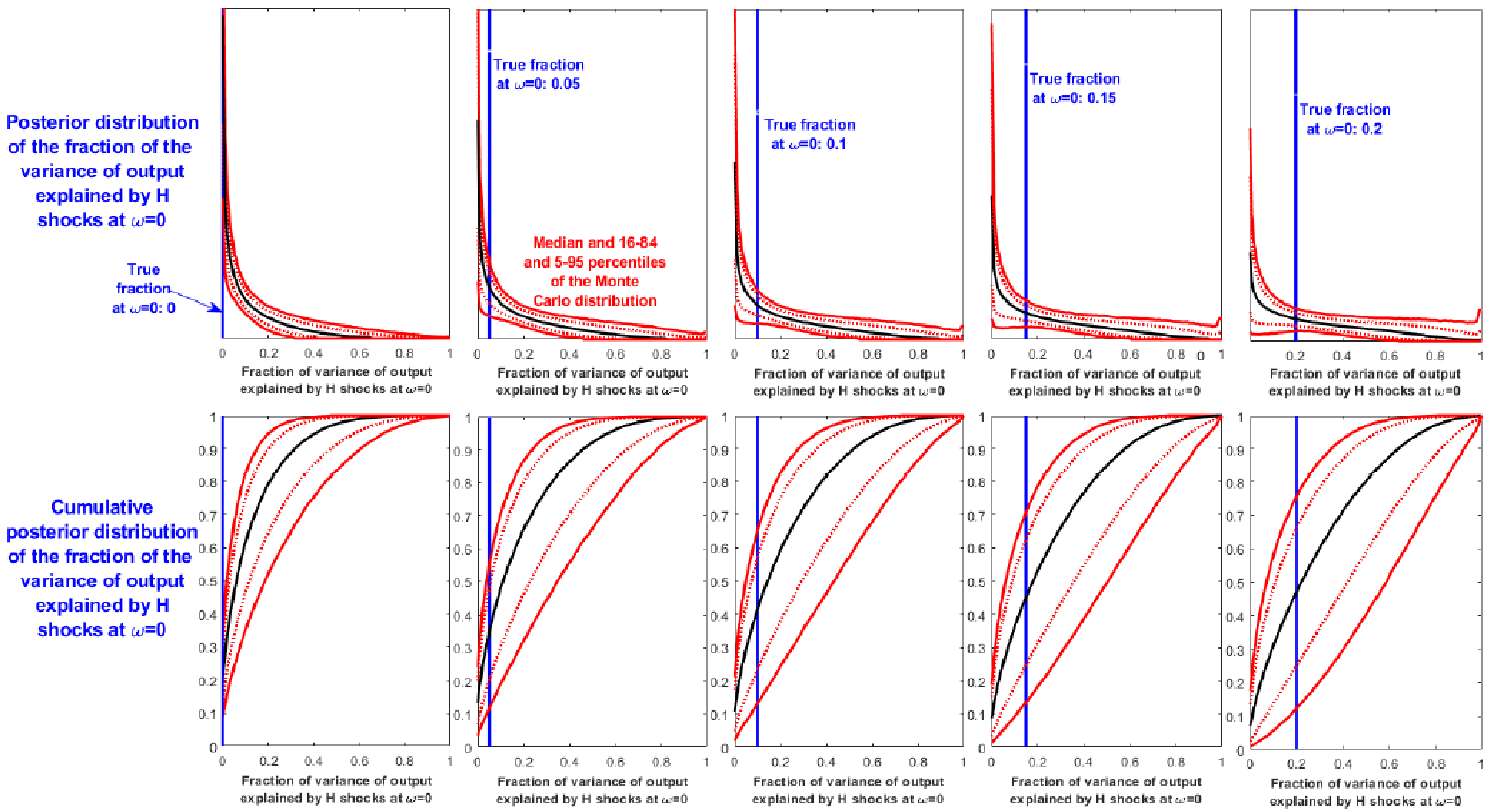

Figure 3 True fractions of the variance of output explained by ϵ_t^e (H) shocks at ω=0 in the RBC model, and medians and 16-84 and 5-95 percentiles of the Monte Carlo distributions of the estimated fractions, based on VARs in differences and Scheme I

The first column of Figure 3 shows the results for Scheme I applied to VARs in differences. The two panels show the means, across the Monte Carlo distribution, of the median and the 16-84 and 5-95 percentiles of the posterior distribution (top panel), and posterior cumulative distribution (bottom panel), of the fractions of the variance of output explained by identified H shocks at ω=0, together with the true fractions of the variance of output at ω=0 explained by ϵ_t^e in the RBC model.[13] The evidence in the two panels is very clear: Scheme I tends to spuriously identify hysteresis with non-negligible, and in fact quite sizeable probability, when there is none in the DGP. A likely (although not the only) reason for this is that, as discussed in Section 3, by its very logic this scheme imposes upon the data the very existence of both BQ *and* H shocks.

Table 2 Fraction (Φ) of the variance of output explained by ϵ_t^e (H) shocks at ω=0, and Monte Carlo fraction of identified models featuring H shocks, based on Scheme II

	Based on VARs in levels:		*Based on VARs in differences:*	
Φ	Mean	Median and 5-95 per cent credible set	Mean	Median and 5-95 per cent credible set
0.000	0.002	0.001 [0.000 0.006]	0.002	0.000 [0.000 0.006]
0.050	0.011	0.000 [0.000 0.036]	0.014	0.000 [0.000 0.056]
0.100	0.024	0.001 [0.000 0.084]	0.026	0.001 [0.000 0.126]
0.150	0.034	0.001 [0.000 0.172]	0.042	0.002 [0.000 0.226]
0.200	0.049	0.002 [0.000 0.303]	0.068	0.004 [0.000 0.481]

It is therefore of interest to compare these results with the corresponding set of results based on Scheme II, which are reported in the first row of Table 2. Interestingly, based on VARs in either levels or differences, when the DGP features no hysteresis Scheme II identifies a fraction of models featuring H shocks that is essentially equal to zero. This suggests that this scheme might provide a more reliable strategy for searching for hysteresis in macroeconomic data. In order to ascertain whether this is in fact the case, we now turn to DGPs that do feature hysteresis.

of the frequency-zero variance of output within their own model.) For example, models featuring BQ shocks could explain, on average, 95 per cent of the long-horizon FEV of output, whereas those featuring H shocks could explain, on average, 50 per cent. Under these circumstances simply focusing on the fraction of models featuring either BQ or H shocks would be misleading. In fact, this is not the case. In each single application, based on either Monte Carlo simulations, or the actual data, the posterior distributions of the long-horizon FEV of output explained by the models featuring BQ and H shocks are consistently very close, and often near-identical. This suggests that, for all practical purposes, Scheme II applied to VARs in levels identifies two sets of models, each of them featuring a single permanent shock.

[13] So, to be clear, for each Monte Carlo simulation the Bayesian SVAR produces a posterior distribution of the fraction of the variance of output explained by H shocks at ω=0. For each simulation we extract the median and the 16-84 and 5-95 percentiles of this distribution, which we store thus building up the Monte Carlo distribution of these objects. What Figure 3 shows are the means of these objects taken across the Monte Carlo distributions.

Imposing Schemes I and II in the presence of hysteresis Columns 2 to 4 of Figure 3, and rows 2 to 4 of Table 2 report the corresponding sets of results for four different values of the true fraction of the variance of output explained by ϵ_t^e at ω=0 in the RBC model, Φ. The main finding in Figure 3 is that, conceptually in line with the previously discussed evidence for Φ=0, Scheme I has a low power to discriminate between alternative extents of hysteresis. This is especially apparent from the evidence in the top row, as the posterior distributions of the fraction of the variance of output explained by H shocks are consistently spread out, with materially different values of Φ being associated with quite similar probabilities. For example, when Φ is equal to 10 per cent, which one might regard, *ex ante*, as a plausible value, the mode of the posterior in panel (1,3) still has a clear peak at Φ=0, but values in excess of 20, 30, or even 40 per cent are associated with non-negligible probabilities.

Turning to the corresponding evidence for Scheme II, two things ought to be pointed out of the results in Table 2. First, and least importantly, the mean (i.e., the expected value of the) fraction of identified models featuring H shocks does not reliably capture the true value of Φ in the DGP, it is rather systematically lower, and it exhibits a non-linear relationship with Φ. For our purposes, however, this bears no negative implication, since there should be no presumption that the estimated fraction of models featuring H shocks should be a reliable estimator of Φ. The crucial point is rather that, as Φ increases in the DGP, the mean fraction of models featuring H shocks increases monotonically: this suggests that, in principle, based on a Monte Carlo-based approach conceptually akin to Stock and Watson's (1996, 1998) TVP-MUB methodology, it should be possible to back out a reliable estimate of Φ from the Monte Carlo distributions generated conditional on a grid of values of Φ.

4 A Monte Carlo-Based Approach

Let us start from Scheme II, for which the Monte Carlo-based approach we propose is more straightforward.

4.1 Scheme II

When applied to Bayesian VARs Scheme II produces a certain fraction θ of models (i.e. of SVARs) for which the permanent output shock is H, with $0 \leq \theta \leq 1$. Let θ^* be the fraction of such models obtained based a specific sample of actual data. The proposed strategy, which (as discussed below) is conceptually related to Stock and Watson's (1996, 1998) TVP-MUB methodology for testing for the presence of random-walk time-variation, and estimating its extent, can be described as follows.

We start by simulating estimated VARs upon which (as described below) we impose specific values of hysteresis, i.e. of the fraction (Φ) of the frequency-zero variance of output explained by H shocks. We consider a grid of K values for Φ from 0 (no hysteresis) to Φ_{Max}=0.5 (for which H shocks explain half of the frequency-zero variance of output). We label the individual values of Φ in the grid as Φ_k, with $k = 1, 2, ..., K$. Based on each simulated, artificial sample j, with $j = 1, 2, ..., J$, conditional on Φ_k we then estimate the same Bayesian VAR we previously estimated based on the actual data, and we implement Scheme II. This produces a fraction θ_j^k of identified models for which the permanent output shock is H. In this way, for each Φ_k we build up the Monte Carlo distribution of the fraction of models with H shocks. The Monte Carlo evidence in Section 3.1.2 suggests that the larger Φ_k the greater the probability of identifying models featuring H shocks, so that as Φ_k increases from 0 to Φ_{Max}, the resulting Monte Carlo distributions of the θ_j^k's shift more and more to the right. Let $\bar{\theta}^k$ be the median of the Monte Carlo distribution of the θ_j^k's, computed conditional on Φ_k. A natural estimate of the fraction of the frequency-zero variance of output explained by H shocks (i.e., the extent of hysteresis) is the value of Φ_k such that $\bar{\theta}^k = \theta^*$, the fraction of models featuring H shocks obtained based on the actual data. Finally, by comparing θ^* with the Monte Carlo distribution of the θ_j^1's (i.e., the distribution for Φ_1=0, with no hysteresis) it is possible to compute a 'pseudo *p*-value' for the null hypothesis of no hysteresis.[14]

This approach bears a close conceptual similarity with Stock and Watson's (1996, 1998) TVP-MUB methodology. Working within a Classical context, they perform a test for a joint break in the model's coefficients, and then build up the Monte Carlo distribution of the test statistic conditional on a grid of values for the extent of random-walk time-variation in the coefficients. Once again, the intuition is that the larger the extent of random-walk time variation, the more the Monte Carlo distribution of the break test statistics gets shifted to the right. The TVP-MUB estimate of the extent of random-walk time-variation is that specific value conditional upon which the median of the resulting Monte Carlo distribution is equal to the value of the break test based on the actual data.

Let us now turn to Scheme I.

4.2 Scheme I

Based on each sample, this scheme produces a posterior distribution for Φ. Let $F^*(\Phi)$ be the posterior distribution obtained by applying Scheme I to a Bayesian cointegrated VAR estimated based on the actual data. Exactly as for Scheme II, for each value of Φ_k in the grid from 0 to Φ_{Max} we simulate the VAR J times. Based on each artificial sample j we then estimate the same VAR we previously estimated based on the actual data, and we implement Scheme I, thus obtaining a posterior distribution $f_j^k(\Phi)$ of the fraction of the frequency-zero variance of output explained by H shocks. In this way, for each Φ_k we build up the Monte Carlo distribution of the $f_j^k(\Phi)$'s. The Monte Carlo evidence in Figure 2 suggests that the larger Φ_k the more the Monte Carlo distributions of the $f_j^k(\Phi)$'s will shift to the right. Let $\bar{f}^k(\Phi)$

[14]We label it 'pseudo' because we are working within a Bayesian context.

be the median of the Monte Carlo distribution of the $f_j^k(\Phi)$'s computed conditional on Φ_k. A natural estimate of Φ is the value of Φ_k such that $\bar{f}^k(\Phi)$ is closest to the posterior distribution obtained based on the actual data, $F^*(\Phi)$, where 'closeness' is defined in terms of the Kullback-Leibler divergence.

4.3 Imposing specific values of hysteresis upon estimated VARs

Finally, we impose specific values of Φ upon the estimated VARs that we then simulate as follows. Based on the actual data sample we analyze (discussed in Section 5), we start by estimating a cointegrated VAR via Johansen's maximum likelihood estimator as described in Hamilton (1994), imposing in estimation the number of cointegration vectors identified based on Johansen's trace and maximum eigenvalue tests (these tests and their results are discussed in detail in Appendix B). When identifying a single permanent real GDP shock via long-run zero restrictions, the shock satisfies all of the restrictions characterizing BQ shocks in Table 1. In the Monte Carlo simulations conditional on Φ_1=0 (i.e., no hysteresis) we therefore generate the artificial samples by simply bootstrapping this SVAR, randomly drawing with replacement from the identified structural shocks. As for the Φ_k's for k>1 we proceed as follows. Based on the point estimate of the reduced-form cointegrated VAR we identify both a BQ and an H shock by imposing the restrictions in Table 1, together with the long-run zero restriction that the two shocks are the only ones allowed to have a permanent impact on real GDP. We implement this by drawing 100,000 random rotation matrices jointly imposing the long-run zero restrictions, and the short- and long-run sign restrictions, via Arias et al.'s (2018) algorithm. This produces a distribution of the impact matrix of the structural shocks, A_0, in the cointegrated SVAR $\Delta Y_t = B_0 + B_1 \Delta Y_{t-1} + ... + B_p \Delta Y_{t-p} + \alpha\beta' Y_{t-1} + A_0 \epsilon_t$, where β is the matrix of the cointegration vectors, α is the matrix of the loadings, $E[\epsilon_t' \epsilon_t] = I_N$, with I_N being the $N \times N$ identity matrix, and the rest of the notation is obvious. Let $\bar{A}_0$ be the median of this distribution. From the distribution we select the A_0 matrix that minimizes the sum of the squared deviations of its individual elements from the corresponding elements of $\bar{A}_0$, i.e. the sum across all of the N^2 elements of $(A_0\text{-}\bar{A}_0)^2$. Finally, we impose specific values of Φ_k>0, by appropriately rescaling the column of A_0 associated with H shocks, and we generate the artificial samples for the Monte Carlo exercises by bootstrapping the resulting SVAR exactly as we do for Φ_1=0.

5 Evidence for the Post-WWII United States

Our baseline system features the logarithms of real GDP, real consumption, real investment, and total hours worked; inflation, computed as the log-difference of the GDP deflator; the 10-year government bond yield; and Wu and Xia's (2016) 'shadow rate'. The sample period is 1954Q3-2019Q4, and the sources of the data are described in detail in Appendix A.

We start by discussing the evidence produced by Scheme II applied to Bayesian VARs in levels, which based on the Monte Carlo evidence in Section 3.1.2 should be expected to produce, for plausible value of hysteresis, more precise results, and we then turn to discussing that produced by either Scheme I or Scheme II applied to Bayesian cointegrated VARs. As already mentioned, the two sets of results are very similar, and suggest that H shocks likely explain about 6-7 per cent of the frequency-zero variance of GDP, although strictly speaking evidence is compatible with the notion of no hysteresis.

5.1 Evidence based on VARs in levels and Scheme II

The approach we use is the same as in the Monte Carlo exercise of Section 3.1.2, i.e. we estimate the reduced-form Bayesian VAR as in Uhlig (1998, 2005), taking 100,000 draws from the posterior distribution. For each draw we then identify, as in Uhlig (2003, 2004), the most powerful shock for real GDP at the 15 years horizon, and we check whether it satisfies the restrictions in Table 1 characterizing either BQ or H shocks. If the draw satisfies either of the two sets of restrictions we keep it, otherwise we discard it. As for the long-run impact of H shocks on the GDP deflator, we either leave it unrestricted, or we impose the restriction that its sign is the same as the sign of its short-run impact.

Starting from the quantitative importance of hysteresis, in terms of the fraction of the variance of GDP at ω=0 explained by H shocks (i.e., Φ), the first column of Figure 4 shows the fraction of models featuring H shocks estimated in the data, together with the medians of the Monte Carlo distributions generated (as described in Section 4) conditional on alternative values of Φ. The fractions of models featuring H shocks estimated in the data are equal to 6.7 when no restriction on their long-run impact on the GDP deflator is imposed, and to 6.1 per cent when it is instead imposed. As expected, the medians of the Monte Carlo distributions increase monotonically with Φ. The Monte Carlo-based estimate of Φ, equal to the value for which the median of the corresponding Monte Carlo distribution is equal to the fraction of models featuring H shocks estimated in the data, is equal to 7 and 6.5 per cent when restrictions on the long-run impact of H shocks on the GDP deflator are not, and are imposed respectively, thus showing that, in fact, imposing or not imposing such restriction makes little difference.

The second column of Figure 4 shows the deconvoluted density of Φ, where the deconvolution has been performed based on exactly the same methodology used by Benati (2007) to deconvolute the density of λ, the extent of random-walk drift in the data estimated via Stock and Watson's (1996, 1998) TVP-MUB methodology. Once again, imposing or not imposing restrictions on the long-run impact of H shocks on the GDP deflator makes virtually no difference. The evidence in the two panels

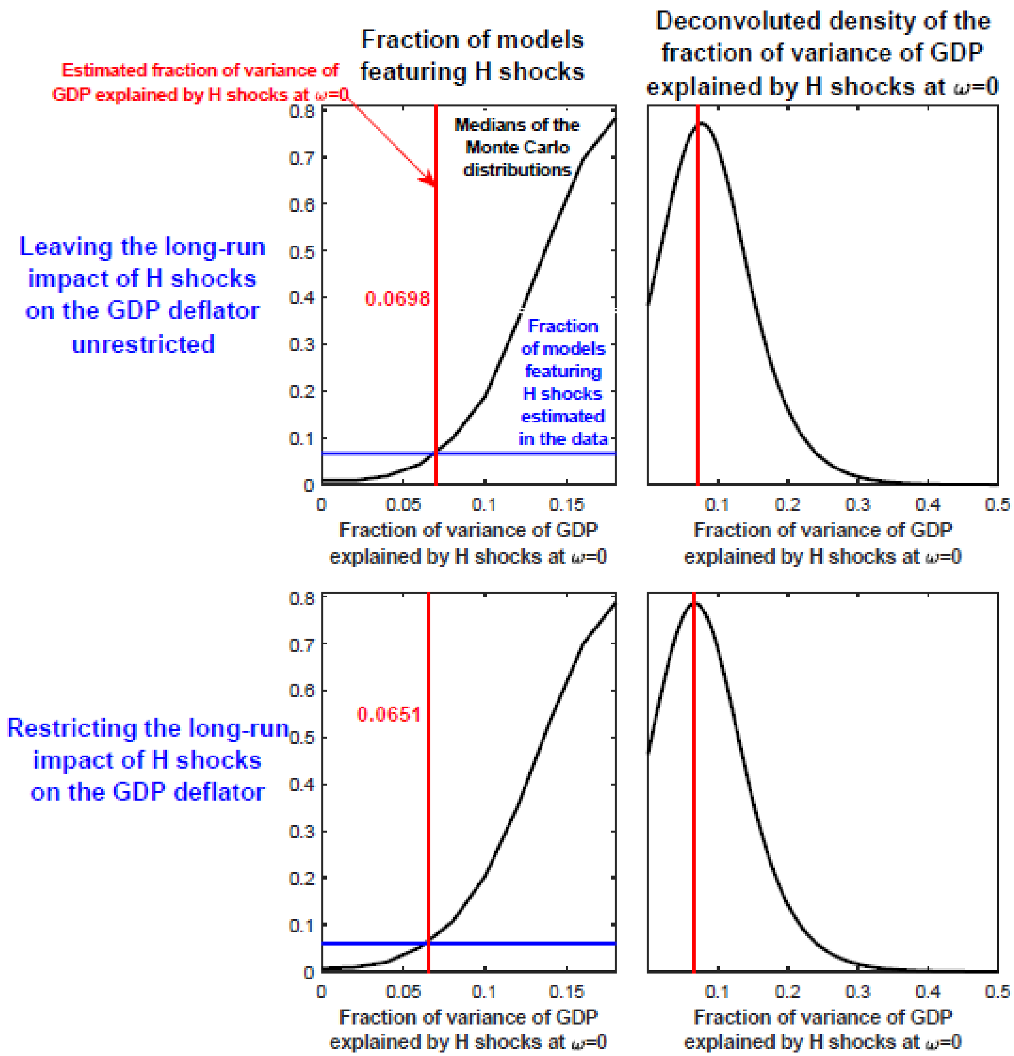

Figure 4 Evidence based on VARs in levels and Scheme II

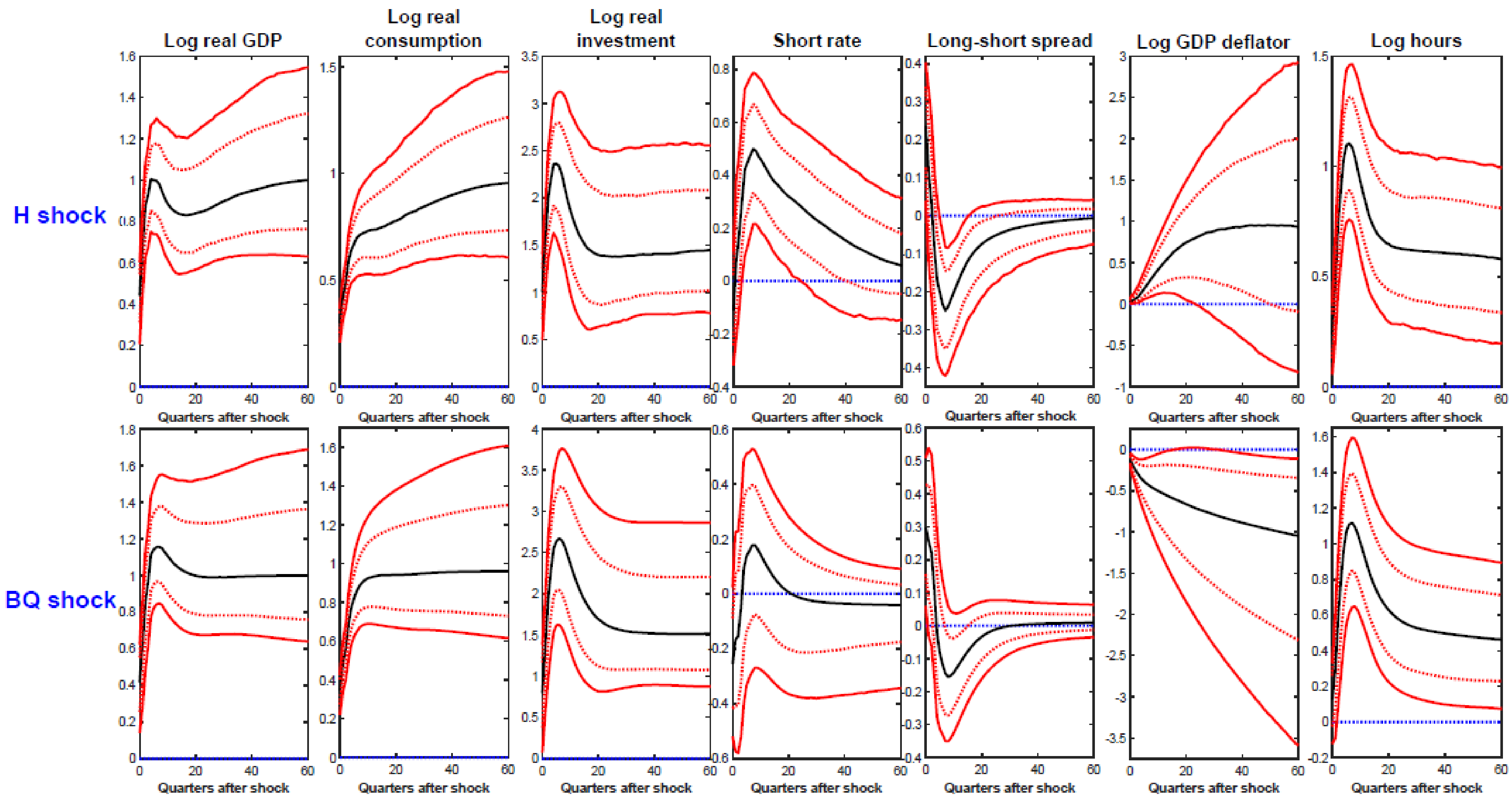

Figure 5 Impulse-response functions based on VARs in levels and Scheme II

suggests that although the just-mentioned estimates of Φ, around 6-7 per cent, are quite small, in fact the only values that can be ruled out with very high confidence are those in excess of 30 per cent, and even quite large values such as 15-20 per cent are still associated with non-negligible probabilities. One obvious reason for such a large uncertainty about estimates of Φ is that the extent of hysteresis is a feature of the data pertaining to the infinite long run, and therefore, as discussed by Faust and Leeper (1997), it should be difficult to pin down with precision in finite samples. In line with Uhlig (2003, 2004), one possibility to obtain more precise estimates might therefore be to focus upon a long, but finite horizon. In fact, this is not the case, as the results obtained by focusing upon the 15 years ahead horizon, as opposed to the frequency ω=0, and numerically very close (these results are available upon request).

Figure 5 shows the IRFs produced by the two types of models, with no restriction being imposed upon the long-run impact of H shocks on the GDP deflator, whereas Figure A.4 in the Online Appendix shows the fractions of FEV explained by BQ and H shocks based on either model. IRFs have been normalized so that the median impact of either shock upon real GDP at the 15 years horizon is equal to one. The most notable result in Figure 5 is that a positive H shock causes a strongly statistically significant increase in hours worked at all horizons, which appears to broadly stabilize about 5 years after impact, thus suggesting that the long-run impact of the shock is in fact permanent. This suggests that one channel through which hysteresis effects manifest themselves is via changes in hours worked. Based on the median estimates, the long-horizon impact on hours is about half, in percentage points, as that on GDP. As we will see in Section 5.2.2, this result is confirmed by the evidence produced by cointegrated SVARs. A second interesting result in Figure 5 is that a non-negligible fraction of the models featuring H shocks is associated with a *negative* impact upon the price level (i.e., with the case depicted in the very last panel of Figure 1). This finding, however, should not be over-emphasized since, the bulk of the mass of the posterior distribution is in fact associated with a positive impact.

The evidence in the first column of Figure A.4 suggests that the fractions of the FEV of GDP explained by BQ and H shocks within their respective models is essentially the same. In particular, at the 15 years horizon, the median and the 90 per cent credible set of the fraction of FEV is equal to 0.884 [0.762 0.955] for BQ shocks, and to 0.900 [0.798 0.961] for H shocks.[15] Either shock also explains large fractions of the FEV of hours and (as expected) consumption and investment, whereas the role played for the remaining variables is significantly smaller.

We now turn to the corresponding evidence based on cointegrated VARs.

[15]The corresponding fractions obtained by imposing restrictions on the long-run impact of H shocks on the GDP deflator are virtually identical.

5.2 Evidence based on Bayesian cointegrated VARs

Our approach to Bayesian cointegration is based on methods introduced by Strachan and Inder (2004) and Koop et al. (2010), which we discuss in the following subsections.

5.2.1 Methodology

Identification of the cointegration vectors Let the cointegrated VECM representation be

$$\Delta Y_t = B_0 + B_1 \Delta Y_{t-1} + ... + B_p \Delta Y_{t-p} + \alpha \beta' Y_{t-1} + u_t, \tag{5}$$

where β is the matrix of the cointegration vectors, α is the matrix of the loadings, $E[u_t' u_t] = \Sigma$, and the rest of the notation is standard.

For the r cointegration vectors to be uniquely identified, each of the r columns of the matrix β ought to feature at least r restrictions (see Kleibergen and van Dijk, 1994, or Bauwens and Lubrano, 1996). Within a Classical context, the standard solution proposed by Johansen (1988, 1991) is to rely on the identification method for reduced-rank regression models introduced by Anderson (1951). Within a Bayesian setting, several methods have been proposed. Following Koop et al. (2010), we adopt the approach proposed by Strachan and Inder (2004), which is based on imposing the normalization that β is semi-orthogonal:

$$\beta' \beta = I_r, \tag{6}$$

where I_r is the $r \times r$ identity matrix.[16] This restriction is imposed by defining a semi-orthogonal matrix $H = H_g (H_g' H_g)^{-1/2}$, which is used to center the prior for the cointegration space around the value that a researcher considers the most plausible. Since H and H_g span the same space, this is obtained by appropriately selecting the columns of H_g based on a priori information, for instance derived from economic theory.

As mentioned, in our baseline specification $Y_t = [y_t, c_t, i_t, R_t, r_t, \pi_t, h_t]'$, where y_t, c_t, i_t, and h_t are, respectively, the logarithms of GDP, consumption, investment, and total hours worked; $\pi_t = p_t$-p_{t-1} is inflation; and R_t and r_t are the short- and the long-term nominal interest rates. As discussed in Appendix B, based on Elliot et al.'s (1996) tests the null hypothesis of a unit root cannot be rejected for any of the series (whereas it is strongly rejected for their differences). Further, Johansen's tests identify three cointegration vectors for the 7-variables system, whereas the corresponding tests for the bivariate systems featuring GDP and consumption, consumption and investment, and the short- and the long-term nominal interest rates detect cointegration for all

[16] We also considered the "linear normalization" approach proposed by Geweke (1996). Although the results are qualitatively similar to those based on Strachan and Inder (2004), we focus on the latter because Geweke's (1996) approach is less general.

of the three systems. In the light of this evidence, the natural choice is to center the prior for β at the three cointegration relationships identified in the bivariate systems,[17] which is obtained by setting H_g to

$$H_g' = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}. \tag{7}$$

Our calibration for the priors for the second elements of the normalized cointegration vectors for GDP and consumption, consumption and investment, and the short and the long rate (corresponding to the first, second and third rows of H_g', respectively) reflects the fact that in the long run GDP and consumption, and short- and long-term nominal interest rates, respectively, tend to move essentially one-for-one in all countries, and for all historical periods we are aware of,[18] whereas investment typically tends to increase somewhat faster than GDP. For example, in the United States since 1947Q1 real GDP has increased, on a log scale, by 1.97, whereas the corresponding increase for real investment has been equal to 2.75, so that the increase for the former series has been equal to 71.6 per cent of the increase for the latter one. It is important to stress, however, that given the flatness of the prior for β (discussed below, and in Appendix C), the centering of its prior encoded in this choice for H_g imposes, in fact, extremely weak restrictions.

Estimation We estimate the cointegrated VAR via the Gibbs-sampling algorithm proposed by Koop et al. (2010). The algorithm cycles through four steps associated with (i) the loadings matrix α in equation (5), (ii) the matrix of the cointegration vectors β, (iii) the VAR coefficient matrices B_j, and (iv) the covariance matrix Σ, which jointly describe a single pass of the Gibbs sampler. We run a burn-in pre-sample of 100,000 draws, which we then discard. We then generate 1,000,000 draws, which we "thin" by sampling every 100 draws in order to reduce their autocorrelation. This leaves us with 10,000 draws from the ergodic distribution, which we use for inference.[19]

Appendix C discusses in detail both the priors, which we take from either Koop et al. (2010) or Geweke (1996), and the conditional distributions we use in order

[17]Standard neoclassical growth theory suggests that, to the extent that the real interest rate is I(1), it should be cointegrated with both the consumption/GDP and the investment/GDP ratios. As the evidence in Table A.1 in the Online Appendix shows, the null hypothesis of a unit root for the *ex post* real rate (computed as the difference between the short rate and inflation) cannot indeed be rejected. As discussed next, however, our priors for the cointegration vectors are extremely weak, so that the three cointegration relationships in the bivariate systems are only used to center the priors very loosely, and in fact they allow the short rate and inflation to enter the cointegration relationship between GDP and either consumption or investment.

[18]We can provide extensive evidence on this based on long-run data that, for several countries, stretches back in time to the aftermath of the Napoleonic Wars.

[19]The online appendix reports evidence on the convergence of the Markov chain by showing the first autocorrelation of the draws for each individual parameter. For all parameters the autocorrelations are extremely low, typically between -0.1 and 0.1.

to draw from the posterior in steps (i)-(iv), which are taken from either Koop et al. (2010) or Geweke (1996). All of the priors are extremely weak. In particular, the one for the matrix of the cointegration vectors β is flat.

Imposing the identifying restrictions As for Identification Scheme I, we jointly impose the zero long-run restrictions, and the short- and long-run sign restrictions, based on the methodology proposed by Arias et al. (2018). We impose the short-run sign restrictions both on impact and for the subsequent eight quarters.

As for Identification Scheme II we identify, for each draw from the ergodic distribution, the only shock that has a permanent impact on GDP, and we check whether it satisfies both the short- and the long-run restrictions characterizing either BQ or H shocks. If the shock is identified as either BQ or H we keep the draw, otherwise we discard it.

5.2.2 Empirical evidence

Scheme I Figure 6 shows, for a grid of values of the fraction of the variance of GDP explained by H shocks at ω=0 (i.e., Φ), the Kullback-Leibler (KL) divergence statistic between (1) the posterior distribution of Φ estimated in the data based on Scheme I applied to the Bayesian cointegrated VAR, and (2) the median of the Monte Carlo distribution of Φ computed (as discussed in Section 4.2) conditional on each value in the grid. Conceptually in line with the corresponding evidence based on Scheme II applied to VARs in levels discussed in Section 5.1 the goal is, once again, to determine which, among the values of Φ in the grid, is most likely to have produced the posterior distribution of Φ estimated in the data. The evidence in the figure is quite clear, with the KL statistic exhibiting a hump-shaped pattern with a minimum corresponding to Φ=0.07, exactly in line with the results from Scheme II applied to VARs in levels.

Figures 7 and 8 report the median and the 16-84 and 5-95 percentiles of the posterior distributions of the IRFs to BQ and H shocks, and of the fractions of the FEV of the series explained by the two disturbances. The IRFs are qualitatively the same as those produced by Scheme I applied to VARs in levels in Figure 5, with the only difference that here the impact of H shocks on the short rate is estimated to be permanent, rather than transitory. As for the fractions of FEV in Figure 8, the most notable result is that whereas, as mentioned, our Monte Carlo-based approach suggests that Φ is equal to 7 per cent, the median fraction of the long-horizon FEV of real GDP explained by H shocks is slightly greater than 20 per cent. This shows that the Monte Carlo-based methodology we propose does make a material difference to the estimates, pointing towards a smaller extent of hysteresis than suggested by the simple estimates produced by the cointegrated VAR.

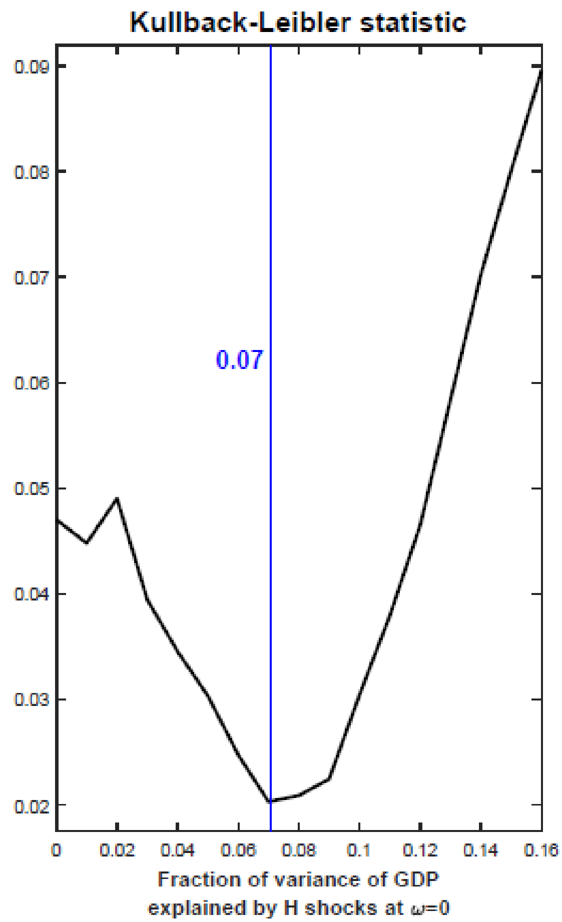

Figure 6 The Kullback-Leibler divergence for alternative values of the fraction of the variance of GDP explained by H shocks at ω=0, based on cointegrated VARs and Scheme I

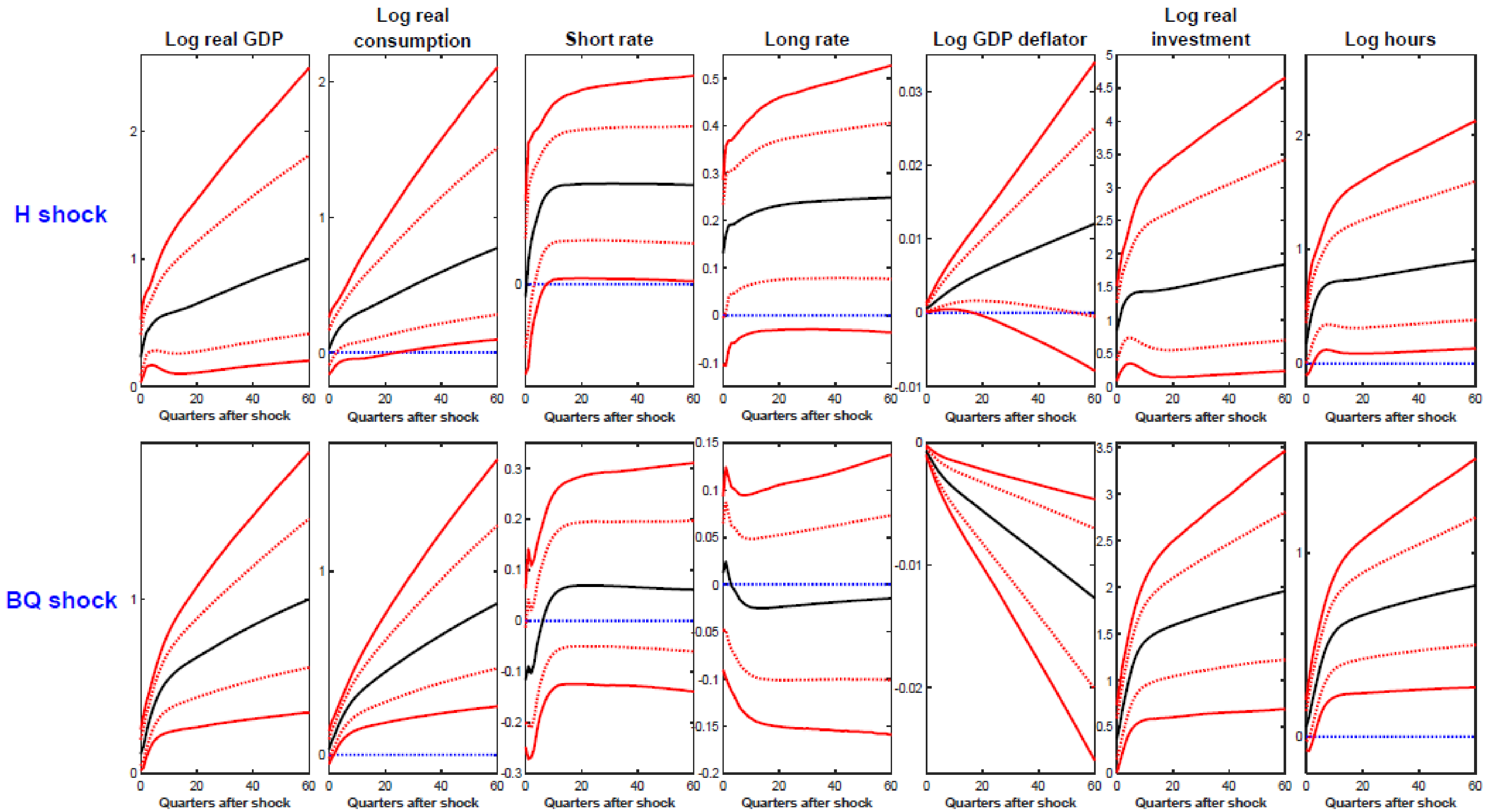

Figure 7 Impulse-response functions based on cointegrated VARs and Scheme I

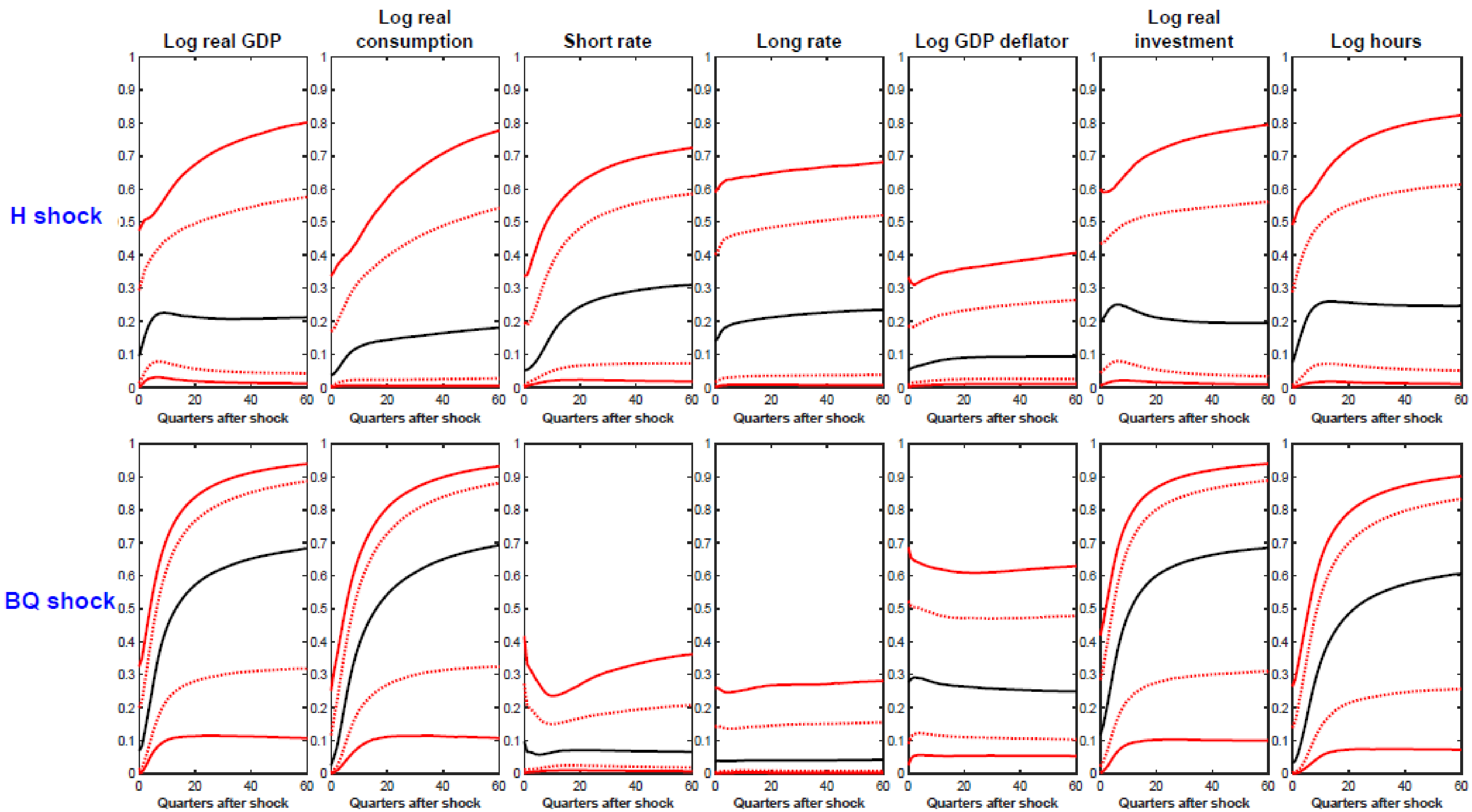

Figure 8 Fractions of forecast error variance based on cointegrated VARs and Scheme I

Scheme II Turning finally to the results produced by Scheme II, they are uniformly in line with the corresponding set of results based on VARs in levels discussed in Section 5.1, and they are therefore not reported here, but they are available upon request.

6 Conclusions

Taking as data-generation process a standard DSGE model, we have shown via Monte Carlo that reliably detecting hysteresis, defined as the presence of aggregate demand shocks with a permanent impact on output, is a significant challenge, as model-consistent identification schemes spuriously detect it with non-negligible probability when in fact the data-generation process features none, and have a low power to discriminate between alternative extents of hysteresis. We propose a simple approach to test for the presence of hysteresis, and to estimate its extent, based on the notion of simulating specific statistics (e.g., the fraction of frequency-zero variance of GDP due to hysteresis shocks) conditional on alternative values of hysteresis we impose upon the VAR, and then comparing the resulting Monte Carlo distributions to the corresponding distributions computed based on the actual data via the Kullback-Leibler divergence. Based on two alternative identification schemes, evidence suggests that post-WWII U.S. data are compatible with the notion of no hysteresis, although the most plausible estimate points towards a modest extent, equal to about 7 per cent of the frequency-zero variance of GDP.

7 References

Anderson, T.W. (1951), "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions", *Annals of Mathematical Statistics*, 22, 327-351.

Anzoategui, D., Comin, D., Gertler, M., and Martinez, J. (2019): "Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence". *American Economic Journal: Macroeconomics*, 11(3), pp. 67-110

Arias, J.E., Rubio-Ramírez, J.F., and Waggoner, D.F. (2018*a*): "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications", *Econometrica*, 685-720.

Arias, J.E., Caldara, D., and Rubio-Ramírez, J.F. (2018*b*): "The systematic component of monetary policy in SVARs: An agnostic identification procedure", *Journal of Monetary Economics*, 1-13.

Ball, L.M. (2009): "Hysteresis in Unemployment: Old and New Evidence". In: Jeff Fuhrer et al. (eds.): *Understanding Inflation and the Implications for Monetary Policy*. MIT Press, Cambridge, MA.

Bauwens, L., and M. Lubrano (1996): "Identification Restrictions and Posterior Densities in Cointegrated Gaussian VAR Systems", in Advances in Econometrics 11, Part B (JAI Press, Greenwich), 3-28

Benati, L. (2007): "Drift and Breaks in Labor Productivity", *Journal of Economic Dynamics and Control*, 31, 2847-2877

Benati, L. (2015): "The Long-Run Phillips Curve: A Structural VAR Investigation", *Journal of Monetary Economics*, 76(November), 15-28.

Benati, L., R. E. Lucas Jr., J.P. Nicolini, and W. Weber (2019): "International Evidence on Long-Run Money Demand", Minneapolis FED Working Paper 587, June 2019

Benati, L., Lucas Jr., R.E., Nicolini, J.P., and Weber, W. (2021): "International Evidence on Long-Run Money Demand". *Journal of Monetary Economics*, 117(1), 43-63

Blanchard, O.J., and Quah, D.T. (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances". *American Economic Review*, 79(14), pp. 655-673.

Blanchard, O.J., and Summers, L. (1986): "Hysteresis and the European Unemployment Problem". NBER Macroeconomics Annual, 1, pp. 15-7.

Blanchard, O.J. (2018): "Should We Reject the Natural Rate Hypothesis?", *Journal of Economic Perspectives*, 32(1, Winter 2018), 97-120

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

Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2012): "Bootstrap Determination of the Cointegration Rank in Vector Autoregressive Models", *Econometrica*, 80(4), 1721-1740.

Cerra, V., and Saxena, S.C. (2008): "Growth Dynamics: The Myth of Economic Recovery". *American Economic Review*, 98(1), pp. 439-57.

Cochrane, J.H. (1988): "How Big Is the Random Walk in GNP?", *Journal of Political Economy*, 96(5), 893-920

Cochrane, J.H. (1994): "Permanent and Transitory Components of GNP and Stock Prices", *Quarterly Journal of Economics*, 109(1), 241-265

Eggertson, G.B. (2010): "What Fiscal Policy Is Effective at Zero Interest Rates?", in NBER Macroeconomics Annuals 2010, pp. 59-112

Elliot, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64(4), 813-836.

Engle, R. F., and C. W. Granger (1987): "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, 55(2), 251-276.

Erceg, C.J., and Levin, A.T. (2014): "Labor Force Participation and Monetary Policy in the Wake of the Great Recession". *Journal of Money, Credit, and Banking*, 46(S2), pp. 3-49.

Forni, M., and L. Gambetti (2014): "Testing for Sufficient Information in Structural VARs", *Journal of Monetary Economics*, 66(September), 124-136.

Furlanetto, F., Lepetit, A., Robstad, Ø., Rubio-Ramirez, J., and Ulvedal, P. (2021): "Estimating Hysteresis Effects". Norges Bank Working Paper 2020-13.

Galí, J. (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?", *American Economic Review*, 89(1), 249-71

Galí, J. (2015*a*): *Monetary Policy, Inflation and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*, Second Edition, Princeton University Press.

Galí, J. (2015*b*): "Hysteresis and the European Unemployment Problem Revisited". In: *Inflation and Unemployment in Europe*, Proceedings of the ECB Forum on Central Banking, European Central Bank, Frankfurt am Main, pp. 53-79.

Galí, Jordi (2020): "Insider-Outsider Labor Markets, Hysteresis and Monetary Policy". Forthcoming, *Journal of Money, Credit and Banking*.

Geweke, J. (1992): "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments", in J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics*, Oxford University Press, Oxford, pages 169-193.

Jaimovich, N., and Siu, H. (2020): "Job Polarization and Jobless Recoveries". *Review of Economics and Statistics*, 102(1), pp. 129-147.

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, 69, 111-132.

Jordà, Ò., Singh, S.R., and Taylor, A.M. (2020): "The Long-Run Effects of Monetary Policy". *NBER Working Paper* No. 26666.

Kleibergen, F. and H.K. van Dijk (1994): "On the Shape of the Likelihood/Posterior in Cointegration Models", *Econometric Theory*, 10, 514-551.

Koop, G., Strachan, R., van Dijk, H., and Villani,M. (2006): "Bayesian Approaches to Cointegration", in K. Patterson and T. Mills, editors, *The Palgrave Handbook of Theoretical Econometrics*, Palgrave MacMillan

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

Luetkepohl, H. (1991): *Introduction to Multiple Time Series Analysis*, 2nd edition. Springer-Verlag.

Ohanian, L.E., and Raffo, A. (2011): "Aggregate Hours Worked in OECD Countries: New Measurement and Implications for Business Cycles", *Journal of Monetary Economics*, 59(1), 40-56

Sims, C.A., Stock, J.H., and Watson, M.W. (1990), "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 58(1), 113-144

Sims, C.A. and Zha, T. (2006): "Were There Regime Switches in U.S. Monetary Policy?", *American Economic Review*, 96(1), 54-81

Sims, C.A. (2002): "Solving Linear Rational Expectations Models", *Computational Economics*, 20, 1-20

Strachan, R. and Inder, B. (2004): "Bayesian Analysis of the Error Correction Model", *Journal of Econometrics* 123, 307-325.

Yagan, D. (2019): "Employment Hysteresis from the Great Recession", *Journal of Political Economy*, 127(5), pp. 2505-2558.

Uhlig, H. (2003): "What Drives GNP?", Unpublished manuscript, Euro Area Business Cycle Network

Uhlig, H. (2004): "Do Technology Shocks Lead to a Fall in Total Hours Worked?", *Journal of the European Economic Association*, 2(2-3), 361-371

Wu, J.C., and F.D. Xia (2016): "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound". *Journal of Money, Credit, and Banking*, 48(2-3), pp. 253-291

Wu, J.C., and F.D. Xia (2017): "Time Varying Lower Bound of Interest Rates in Europe". Manuscript.

A Sources of the Data

Quarterly seasonally adjusted series for real GDP ('GDPC1: Real Gross Domestic Product, Billions of Chained 2012 Dollars, Quarterly, Seasonally Adjusted Annual Rate') and the GDP deflator ('GDPCTPI: Gross Domestic Product: Chain-type Price Index') are from the U.S. Department of Commerce, Bureau of Economic Analysis. Inflation has been computed as the log-difference of the GDP deflator. Quarterly seasonally adjusted series for real chain-weighted investment and real chain-weighted consumption 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 United States National Income and Product Accounts. Whereas consumption pertains to non-durables and services, investment has 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. A quarterly seasonally adjusted series for total hours worked by all persons in the nonfarm business sector ('HOANBS: Nonfarm Business Sector: Hours of All Persons, Index 2009=100') is from the U.S. Department of Labor, Bureau of Labor Statistics. A monthly series for the 10-year government bond yield ('GS10: 10-Year Treasury Constant Maturity Rate') is from the Board of Governors of the Federal Reserve System. A monthly series for Wu and Xia's 'shadow rate' is from Cynthia Wu's website, at: https://sites.google.com/view/jingcynthiawu. Both the Wu-Xia 'shadow rate', and the 10-year government bond yield, have been converted to the quarterly frequency by taking averages within the quarter. A quarterly seasonally adjusted series for average hours worked by all persons in the nonfarm business sector has been computed by dividing the previously mentioned series for total hours worked by total employment in the nonfarm business sector (the FRED II acronym is PRS85006013). Quarterly seasonally adjusted series for the civilian unemployment rate, the level of unemployment (in thousands of persons) and the labor force overall participation rate have been computed by taking averages within the quarter of the monthly series produced by the U.S. Department of Labor, Bureau of Labor Statistics (the FRED II acronyms are UNRATE, UNEMPLOY, and CIVPART).

B Unit Root and Cointegration Properties of the Data

Table A.1 in the Online Appendix reports results from Elliot *et al.* (1996) unit root tests for either the (log) levels or the (log) differences of the series. In short, the null of a unit root cannot be rejected for any of the series, whereas it can be rejected for their (log) differences. Turning to the cointegration properties of the data, basic economic theory suggests that, within the present context, we should expect at least three cointegration relationships: one between the short- and the long-term nominal interest rates, and two between GDP, consumption, and investment. Table 1 provides evidence that this is indeed the case. The table reports bootstrapped p-values[20] for Johansen's trace and maximum eigenvalue tests for three bivariate systems featuring GDP and consumption, consumption and investment, and the short and the long rate. For all systems evidence of cointegration is very strong, with the largest p-value across the three systems being equal to just 0.0033. Turning to the 7-variables system, Table 2 reports bootstrapped p-values for the corresponding tests.[21] Whereas the trace tests uniformly reject the null of no cointegration against any of up to five cointegration vectors, results from maximum eigenvalues tests point towards three cointegration vectors, with the test of three versus four vectors producing a p-value of 0.6376. In the main text we therefore work under the assumption that the 7-variables system features the three cointegration vectors suggested by economic theory.

Table A.1 Bootstrapped *p*-values[a] for Johansen's tests for the bivariate systems

Test	Log GDP and log consumption	Log consumption and log investment	Short rate and long rate
Trace	2.0e-4	3.e-4	0.0029
Maximum eigenvalue	5.0e-4	0.0015	0.0033

[a] Based on 10,000 bootstrap replications of Cavaliere *et al.*'s (2012) procedure.

Table A.2 Bootstrapped *p*-values for Johansen's tests[a] for the seven-variables system

Trace tests of the null of no cointegration against the alternative of h or more cointegration vectors:				
h = 1	*h = 2*	*h = 3*	*h = 4*	*h = 5*
0.0050	0.0025	0.0100	0.0746	0.0865
Maximum eigenvalue tests of h versus h+1 cointegration vectors:				
0 versus 1	*1 versus 2*	*2 versus 3*	*3 versus 4*	*4 versus 5*
0.0854	0.0072	0.0875	0.6376	–

[a] Based on 10,000 bootstrap replications of Cavaliere *et al.*'s (2012) procedure.

[20] Bootstrapping has been implemented as in Cavaliere *et al.* (2012). Cavaliere *et al.* (2012), Benati (2015), and especially the Online Appendix of Benati *et al.* (2021) provide extensive Monte Carlo evidence on the excellent performance of this bootstrapping procedure.

[21] Again, bootstrapping has been implemented based on Cavaliere *et al.*'s (2012) methodology.

C Choice of the Priors, and Drawing from the Posterior, for the Bayesian Cointegrated VAR of Section 5.1

C.1 Choice of the priors

We follow Koop *et al.* (2010) in the choice of the priors for α, β, and Σ. The prior for α is a shrinkage prior with zero mean:

$$\alpha \mid \beta, B_0, B_1, ..., B_p, \Sigma, \tau, \nu \sim MN\left(0, \nu(\beta' P_{1/\tau}\beta)^{-1} \otimes G\right), \tag{8}$$

where MN denotes the matricvariate-normal distribution; $P_{1/\tau} = HH' + \tau^{-1}H_\perp H_\perp'$; and G is a square matrix that, in line with Strachan and Inder (2004), we set to $G = \Sigma$. In addition, τ is a scalar between 0 and 1, and ν is a scalar that controls the extent of shrinkage. In what follows we set $\tau = 1$, which corresponds to a flat, non-informative prior on β, and $\nu = 1000$, corresponding to a very weakly informative prior for α (the standard non-informative prior would be obtained by setting $1/\nu = 0$).

The prior for β is based on a matrix angular central Gaussian distribution:

$$p(\beta) \propto |P_\tau|^{-r/2}|\beta'(P_\tau)^{-1}\beta|^{-N/2}, \tag{9}$$

where $P_\tau = HH' + \tau H_\perp H_\perp'$, and N is the number of variables in the system. The prior for the covariance matrix Σ is the non-informative prior:

$$p(\Sigma) \propto |\Sigma|^{-(N+1)/2}. \tag{10}$$

As for the coefficient matrices B_j, we follow Geweke (1996) and choose a multivariate normal prior, which implies that the posterior is also multivariate normal.[22]

C.2 Drawing from the posterior

The conditional posterior distribution for $G = \Sigma$ is inverse-Wishart:

$$\Sigma \mid \alpha, \beta, B_0, B_1, ..., B_p, Y_t \sim IW\left(u_t' u_t + \nu\alpha(\beta' P_{1/\tau}\beta)^{-1}\alpha', T + r\right), \tag{11}$$

where T is the sample length. The conditional posterior for the coefficient matrices B_j is given by (see Geweke, 1996)

$$vec(A) \mid \alpha, \beta, \Sigma, \tau, \nu, Y_t \sim N\left\{(\Sigma^{-1} \otimes Z'Z + \lambda^2 I)^{-1}(\Sigma^{-1} \otimes Z'Z)vec(A'), (\Sigma^{-1} \otimes Z'Z + \lambda^2 I)^{-1}\right\}, \tag{12}$$

with $A' \equiv [B_0\ B_1\ B_2\ ...\ B_p]$ and $\hat{A} = (Z'Z)^{-1}Z'(\Delta Y - \tilde{Y}\beta\alpha')$. ΔY and $\tilde{Y}$ are $T \times N$ matrices, whose t-th rows are equal to $\Delta Y_t'$ and Y_{t-1}', respectively. Z is a $T \times 1 + Np$

[22] We thank Rodney Strachan for confirming to us that taking step (*iii*) of the Gibbs-sampling algorithm from Geweke (1996) and the remaining steps from Koop et al. (2010) is indeed appropriate.

matrix, whose t-th row is $Z_t \equiv [1\ \Delta Y'_{t-1}\ \Delta Y'_{t-2}\ ...\ \Delta Y'_{t-p}]$. Following Geweke (1996) we set λ^2=10.

We derive the posterior distributions for α and β as in Koop et al. (2010). Using the transformation $\beta\alpha' = (\beta\kappa)(\alpha\kappa^{-1})' = \left[\beta(\alpha'\alpha)^{1/2}\right]\left[\alpha(\alpha'\alpha)^{-1/2}\right]' \equiv ba'$, with $\kappa = (\alpha'\alpha)^{1/2}$, $a = \alpha\kappa^{-1}$, $b = \beta(\alpha'\alpha)^{1/2}$, $\beta = b(b'b)^{-1/2}$, and $b'b = \alpha'\alpha$, the priors for α and β imply the following priors for a and b:

$$b \mid a, B_0, B_1, ..., B_p, \Sigma, \tau, \nu \sim MN\left(0, (a'G^{-1}a)^{-1} \otimes \nu P_\tau\right), \tag{13}$$

$$p(a) \propto |G|^{-r/2}|a'G^{-1}a|^{-N/2}. \tag{14}$$

The conditional draws for α and β, that is, steps (i) and (ii) of the Gibbs-sampling algorithm, can then be obtained via the following two steps:

(i') draw $\alpha^{(*)}$ from $p(\alpha|\beta, B_0, B_1, ..., B_p, \Sigma, \tau, \nu, Y_t)$ and transform it to obtain a draw $a^{(*)} = \alpha^{(*)}(\alpha^{(*)\prime}\alpha^{(*)})^{-1/2}$; and

(ii') draw b from $p(b|a^{(*)}, B_0, B_1, ..., B_p, \Sigma, \tau, \nu, Y_t)$ and transform it to obtain draws $\beta = b(b'b)^{-1/2}$ and $\alpha = a^{(*)}(b'b)^{1/2}$.

In the Online Appendix we report evidence on the convergence properties of the Koop et al. (2010) Gibbs-sampling algorithm.

Online Appendix for: Searching for Hysteresis

Luca Benati
University of Bern*

Thomas A. Lubik
Federal Reserve Bank of Richmond†

A The Real Business Cycle Model Used in the Monte Carlo Exercise of Section 3

Here follows a detailed description of the real business cycle (RBC) model we use as data-generation process (DGP) in the Monte Carlo exercise of Section 3. In essence, the model is that described in Galí's (2015) Chapter 2, augmented with habit-formation in consumption and a unit root in technology (and therefore in the natural level of output), and featuring the possible presence of hysteresis effects (as discussed below). Whenever possible we follow Galí's own notation. Households solve the following problem

$$U_0^* \equiv \underset{C_t, N_t}{\text{Max}} \; E_0 \sum_{t=0}^{\infty} \beta^t \left[\ln(C_t - bC_{t-1}) - \frac{N_t^{1+\phi}}{1+\phi} \right] \tag{A.1}$$

subject to

$$P_t C_t + Q_t B_t = B_{t-1} + W_t N_t - T_t \tag{A.2}$$

where C_t and N_t are real consumption and hours worked, respectively; $0<b<1$ is the habit-formation parameter; P_t is the price of consumption goods; B_t is the stock of nominal bonds; Q_t is the price, at time t, of a nominal bond paying 1 dollar at time $t+1$; W_t is the nominal wage; T_t are nominal lump-sum taxes; and the rest of the notation is obvious. The two first-order conditions (FOCs) are

$$N_t^{\phi} = \frac{W_t}{P_t} E_t \left[\frac{1}{C_t - bC_{t-1}} - \beta \frac{b}{C_{t+1} - bC_t} \right] \tag{A.3}$$

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

†Research Department, P.O. Box 27622, Richmond, VA 23261. Tel.: +1-804-697-8246. Email: thomas.lubik@rich.frb.org.

and

$$E_t\left[\frac{Q_t}{C_t - bC_{t-1}}\frac{P_{t+1}}{P_t}\right] = \beta E_t\left[\frac{1+bQ_t}{C_{t+1} - bC_t} - \beta\frac{b}{C_{t+2} - bC_{t+1}}\right] \tag{A.4}$$

With no habit-formation (i.e., with b=0) the two FOCs collapse to those found in Galí (2015), i.e.

$$N_t^{\phi} = \frac{W_t}{P_t}\frac{1}{C_t} \quad \text{and} \quad 1 = \beta E_t\left[\frac{P_t C_t}{Q_t P_{t+1} C_{t+1}}\right] \tag{A.5}$$

As for the firms we exactly follow Galí (2015), with the only difference pertaining to the process for technology. Firms produce output (Y_t) via the production function $Y_t = A_t N_t^{1-\alpha}$, with $0<\alpha<1$ being the Cobb-Douglas parameter, A_t being technology, and the capital stock being constant and normalized to 1. Maximization of profits, $P_t Y_t - W_t N_t$, produces the FOC

$$\frac{W_t}{P_t} = A_t(1-\alpha)N_t^{1-\alpha} \tag{A.6}$$

As for the logarithm of technology, $a_t = \ln(A_t)$, we postulate that it evolves according to

$$a_t = a_{t-1} + \epsilon_t^a + \delta\tilde{y}_{t-1} \tag{A.7}$$

with $\epsilon_t^a \sim N(0, \sigma_a^2)$, $\tilde{y}_t$ being the transitory component of output (to be discussed below), and $\delta \geq 0$ capturing the possible presence of hysteresis effects. If $\delta = 0$ there is no hysteresis, whereas if $\delta > 0$ positive (negative) transitory fluctuations of output cause subsequent permanent increases (decreases) in the level of technology. This specification, which is conceptually the same as the one used by Jordà et al. (2020), is designed to capture, in a very simple and stripped-down fashion, the notion that positive (negative) deviations of GDP from potential (here, deviations of output from its stochastic trend A_t) may have a positive (negative) impact on potential GDP itself. Although we could have considered more complex formalizations of the notion of hysteresis, we chose to use (A.7) because for our own purposes (i.e., performing Monte Carlo simulations) it is perfectly appropriate. Finally, since there is no investment, no government expenditure, and no foreign sector, $C_t = Y_t$.

Since a_t is I(1), log-linearization of the FOCs and of the production function requires the preliminary stationarizazion of the relevant variables. We define the stationarized variables

$$\tilde{Y}_t \equiv \frac{Y_t}{A_t} \quad \text{and} \quad \tilde{\Omega}_t \equiv \frac{\left(\frac{W_t}{P_t}\right)}{A_t}, \tag{A.8}$$

with $\tilde{y}_t$ and $\tilde{\omega}_t$ being the log-deviations of $\tilde{Y}_t$ and $\tilde{\Omega}_t$ from the steady-state, so that $\tilde{y}_t$ is the component of output that is driven by the transitory disturbances (discussed below). With $\tilde{Y}_t$ and $\tilde{\Omega}_t$ defined as in (A.8), the production function and the three FOCs can be trivially stationarized. Then, log-linearization of the stationarized production function and of the three stationarized FOCs produces the following four log-linear relationships characterizing the dynamics of the economy in a neighborhood of the steady-state,

$$\tilde{y}_t - (1-\alpha)n_t = 0 \tag{A.9}$$

$$\tilde{\omega}_t + \alpha n_t - v_t = 0 \tag{A.10}$$

$$-R_t + \pi_{t+1|t} + u_t + \phi n_t - \tilde{\omega}_t - \phi n_{t+1|t} + \tilde{\omega}_{t+1|t} = 0 \tag{A.11}$$

$$\phi n_t - \tilde{\omega}_t - e_t + \frac{1}{(1\text{-}b\beta)(1\text{-}b)}\left[(1+\beta b^2)\tilde{y}_t - b\beta\tilde{y}_{t+1|t} - b\tilde{y}_{t-1} + b\Delta a_t\right] = 0 \tag{A.12}$$

where $n_t \sim I(0)$ is the log-deviation of N_t from the steady-state, $\pi_t \equiv p_t\text{-}p_{t-1} = \ln(P_t)\text{-}\ln(P_{t-1})$ is inflation, e_t is a transitory demand disturbance, and v_t and u_t are two transitory supply disturbances, with

$$v_t = \rho_v v_{t-1} + \epsilon_t^v \tag{A.13}$$

$$u_t = \rho_u u_{t-1} + \epsilon_t^u \tag{A.14}$$

$$e_t = \rho_e e_{t-1} + \epsilon_t^e \tag{A.15}$$

with $\epsilon_t^v \sim N(0,\sigma_v^2)$, $\epsilon_t^u \sim N(0,\sigma_u^2)$, $\epsilon_t^e \sim N(0,\sigma_e^2)$, $|\rho_v,\rho_u,\rho_e| < 1$, and where the rest of the notation is obvious. The logarithm of technology, a_t, evolves according to (A.7). Finally, monetary policy is characterized by a standard Taylor rule with smoothing,

$$R_t = \rho R_{t-1} + (1-\rho)\phi_\pi \pi_t \tag{A.16}$$

where the notation is obvious. By defining the state vector as

$$\xi_t = [R_t, \pi_t, \tilde{y}_t, n_t, \tilde{\omega}_t, \Delta a_t, v_t, u_t, e_t, \pi_{t+1|t}, \tilde{y}_{t+1|t}, n_{t+1|t}, \tilde{\omega}_{t+1|t}]' \tag{A.17}$$

and augmenting the system with the definition of the four rational expectations forecast errors

$$\pi_t = \pi_{t|t-1} + \eta_t^\pi \tag{A.18}$$

$$\tilde{y}_t = \tilde{y}_{t|t-1} + \eta_t^{\tilde{y}} \tag{A.19}$$

$$n_t = n_{t|t-1} + \eta_t^n \tag{A.20}$$

$$\tilde{\omega}_t = \tilde{\omega}_{t|t-1} + \eta_t^{\tilde{\omega}} \tag{A.21}$$

the system can be put into the 'Sims canonical form' (see Sims, 2002) and solved.

We calibrate the structural parameters as follows: β=0.99, α=1/3, b=0.8, ϕ=1, ρ=0.9, ϕ_π=1.5, and σ_a=0.007.[1] We augment the model with three transitory AR(1) disturbances, two of them (v_t and u_t) supply-side, and one (e_t) demand-side, and

[1] The value for σ_a is close to Watson's (1986, p. 60) estimate of the standard deviation of shocks to the stochastic trend of log real GDP (0.0057). The rationale is that, within the present context, a_t is the random-walk component of log real GDP.

an additive disturbance to the log of the production function, $z_t \sim N(0, \sigma_z^2)$, with σ_z=0.005. The AR parameters are calibrated to ρ_v=ρ_u=ρ_e=0.75, whereas the standard deviations of the disturbances' innovations ϵ_t^v, ϵ_t^u, and ϵ_t^e (all zero-mean, and normally distributed) are set to σ_u=0.001, σ_v=0.005 and σ_e=0.045. Based on this calibration the permanent technology shock (ϵ_t^a) explains exactly 1/3 of the forecast error variance (FEV) of log GDP on impact, and with δ=0 it explains slightly more than 96 per cent of GDP's FEV 15 years ahead. These figures are broadly in line with the evidence produced by the structural VAR literature: for example, as for the fraction of the FEV of GDP explained by permanent shocks on impact see Table I.2 of Cochrane (1994). Further, based on these values of the structural parameters the demand-side disturbance ϵ_t^e is quite close to being the only driver of the transitory component of output, so that the identifying restrictions in Table 1 are, for practical purposes, correct. Finally, as for δ in the Monte Carlo exercise we consider a grid of values, from δ=0 (no hysteresis) to δ=0.348, for which the technology shock ϵ_t^a and the demand-side shock ϵ_t^e both explain virtually half of the frequency-zero variance of output.

B Additional Tables and Figures

Table A.1*a* Bootstrapped *p*-values for Elliot, Rothenberg, and Stock unit root tests[a] for the (log) levels of the series

	In levels				*In differences*			
Series	*p*=2	*p*=4	*p*=6	*p*=8	*p*=2	*p*=4	*p*=6	*p*=8
Log real GDP	0.5780	0.8241	0.7787	0.8105	0.0000	0.0000	0.0000	0.0002
Log real consumption	0.7368	0.8979	0.9060	0.9204	0.0000	0.0000	0.0000	0.0002
Log real investment	0.3562	0.3501	0.2409	0.3119	0.0000	0.0000	0.0000	0.0001
Log total hours	0.5413	0.7210	0.7594	0.9082	0.0000	0.0000	0.0000	0.0000
Long rate	0.6642	0.6366	0.7697	0.8215	0.0000	0.0000	0.0000	0.0000
Short rate	0.2491	0.1753	0.1433	0.3722	0.0000	0.0000	0.0000	0.0000
Inflation	0.1083	0.2288	0.1881	0.2336	0.0000	0.0000	0.0000	0.0000
Ex post real short rate[b]	0.0697	0.1709	0.1436	0.2034	0.0000	0.0000	0.0000	0.0000

[a] Based on 10,000 bootstrap replications. For details, see Appendix A. [b] Short rate minus inflation.

C References

Cochrane, J.H. (1994): "Permanent and Transitory Components of GNP and Stock Prices", *Quarterly Journal of Economics*, 109(1), 241-265.

Galí, J. (2015): *Monetary Policy, Inflation and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*, Second Edition, Princeton University Press.

Sims, C.A. (2002): "Solving Linear Rational Expectations Models", *Computational Economics*, 20, 1-20.

Watson, M. (1986): "Univariate Detrending Methods with Stochastic Trends", *Journal of Monetary Economics*, 18, 49-75.

Figures for the Online Appendix

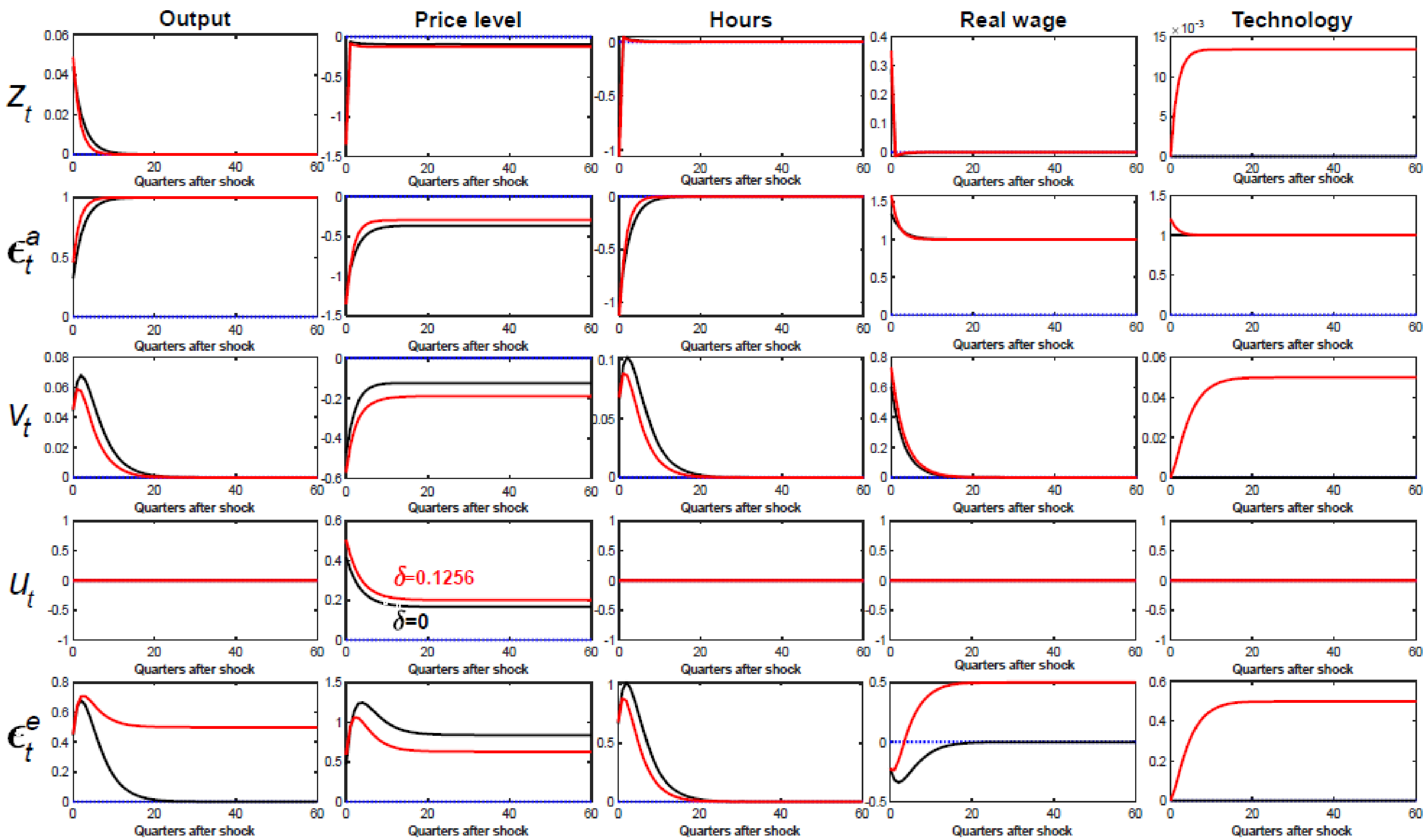

Figure A.1 Theoretical impulse-response functions for the RBC model

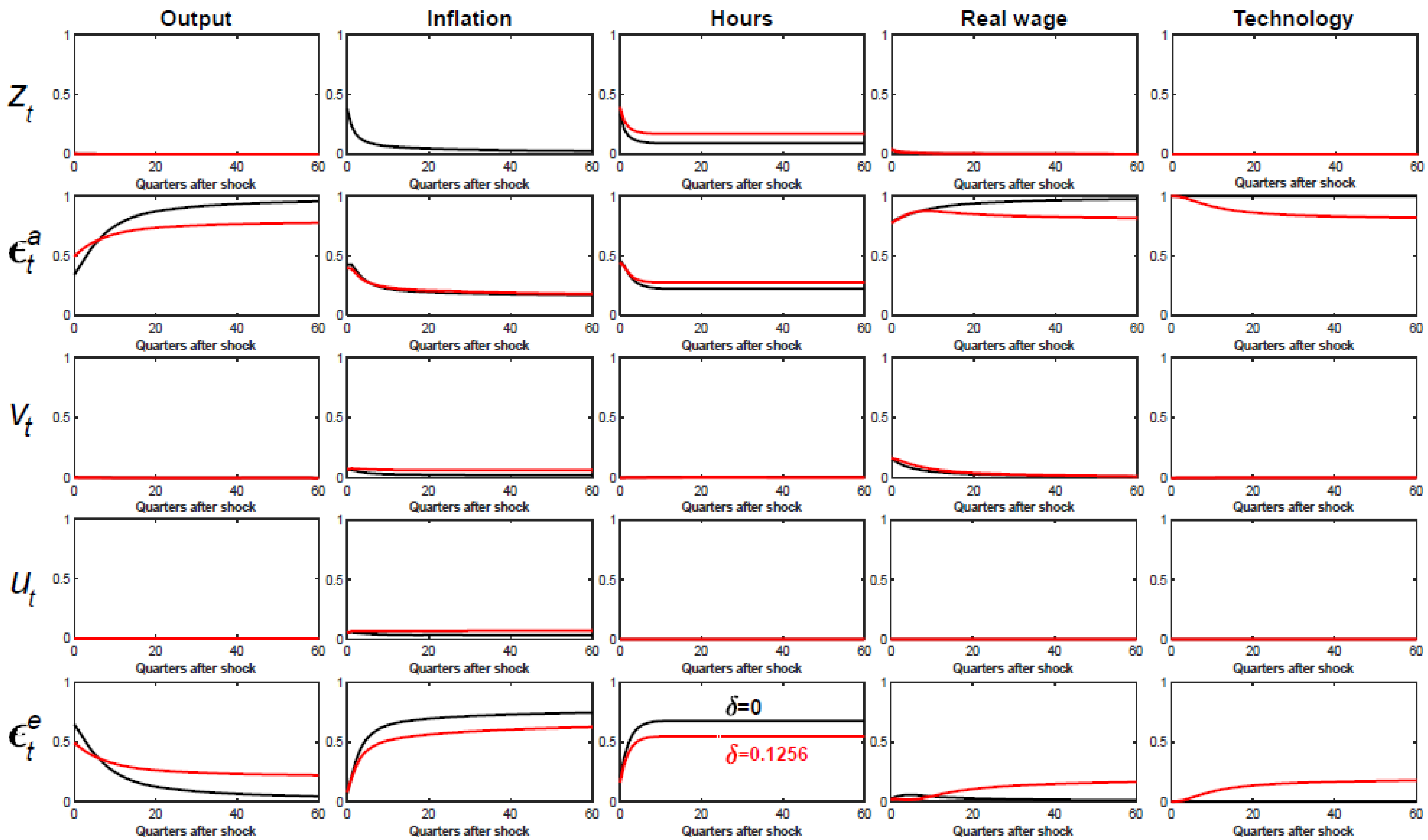

Figure A.2 Theoretical fractions of forecast error variance for the RBC model

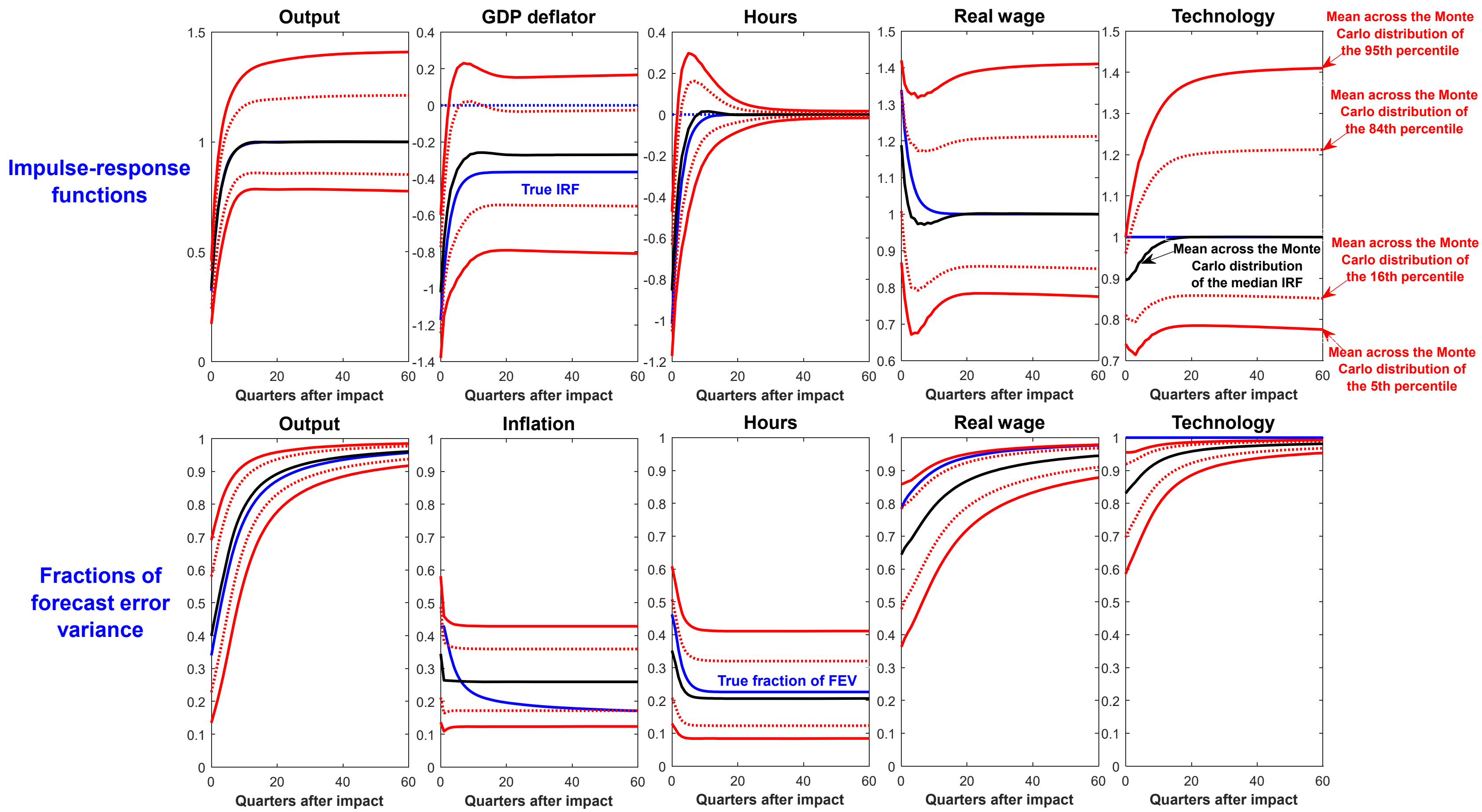

Figure A.3 Monte Carlo evidence on recovering the RBC model's true IRFs and fractions of forecast error variance with no hysteresis, conditional on the correct identification scheme

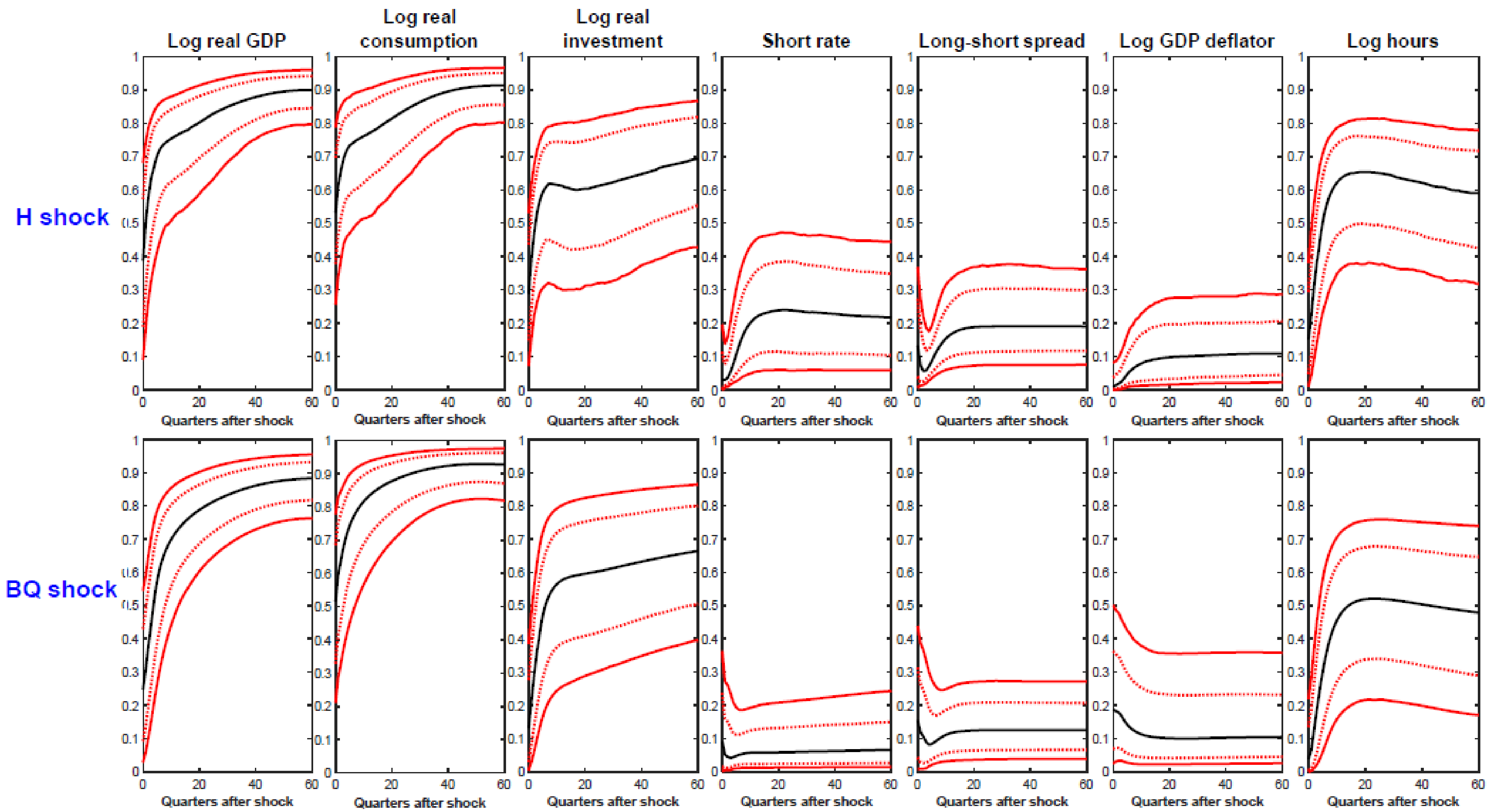

Figure A.4 Fractions of forecast error variance based on VARs in levels and Scheme II

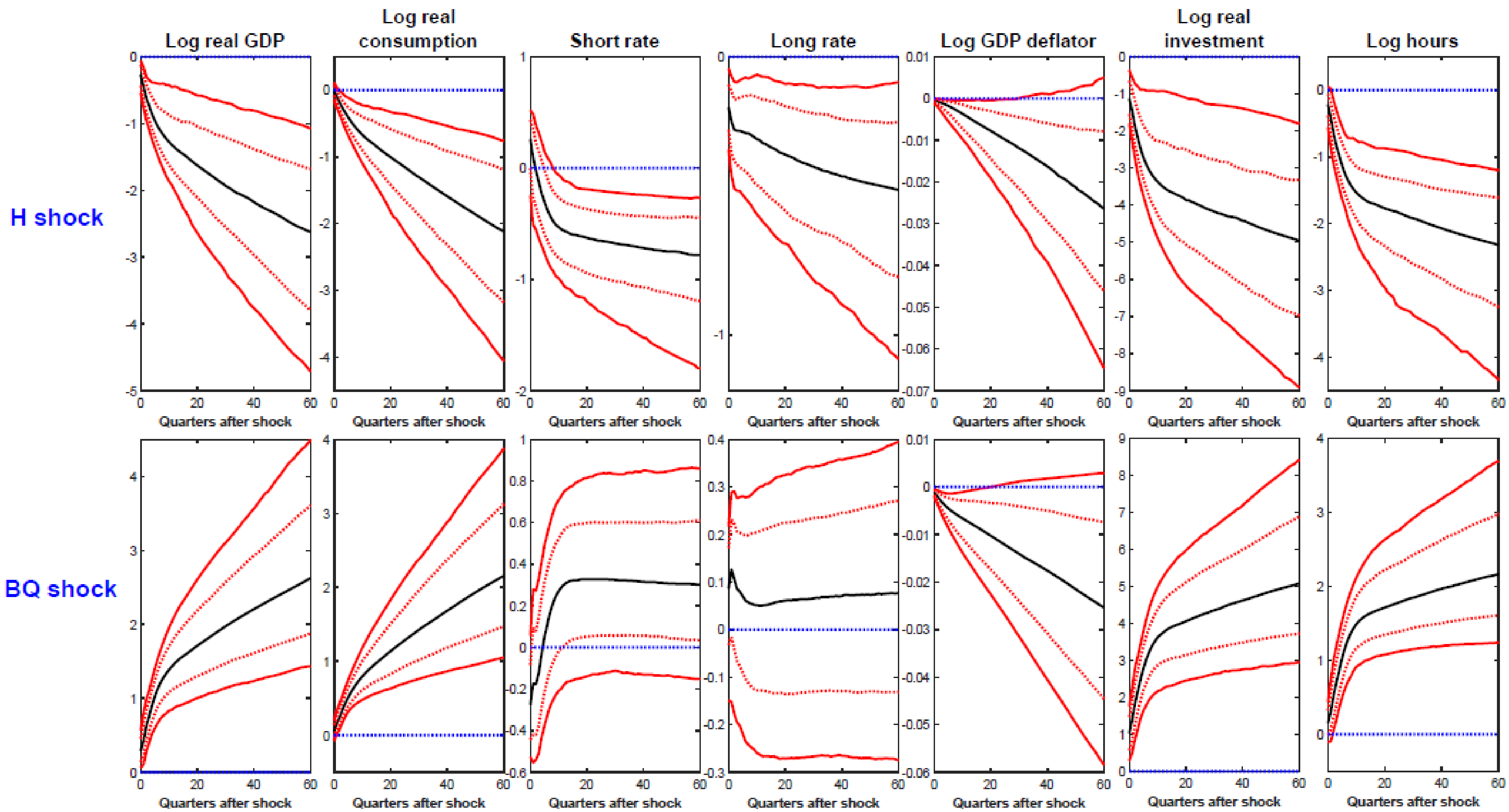

Figure A.5 Impulse-response functions based on cointegrated VARs and Scheme II (without imposing restrictions on the long-run impact of H shocks on the GDP deflator)

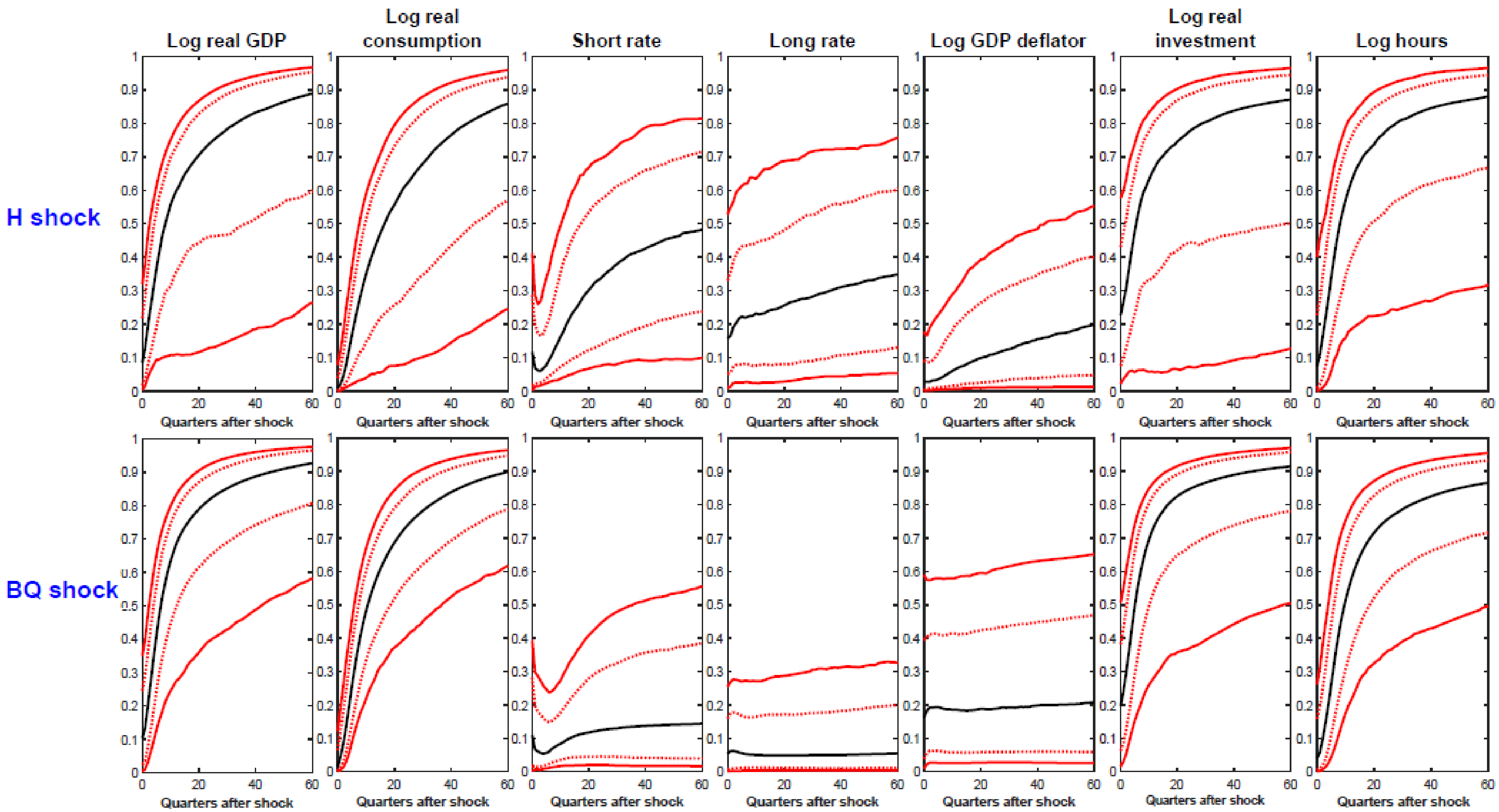

Figure A.6 Fractions of forecast error variance based on cointegrated VARs and Scheme II (without imposing restrictions on the long-run impact of H shocks on the GDP deflator)