Searching for Hysteresis*

Luca Benati
University of Bern†

Thomas A. Lubik
Federal Reserve Bank of Richmond‡

December 12, 2022

Abstract

We search for hysteresis in aggregate US data. Defining hysteresis as the presence of aggregate demand shocks with a permanent impact on real GDP, we identify such shocks in Bayesian VARs with a combination of zero long-run restrictions and both short- and long-run sign restrictions. Our findings suggest that hysteresis effects have been virtually absent for the sample excluding the financial crisis and the Great Recession. They only appear when we include the period following the collapse of Lehman Brothers. Even then, these effects only account for at most 10 percent of the long-run variance of GDP. We demonstrate via a DSGE-model based simulation analysis that there is a high probability of detecting hysteresis effects even when the data-generating process features none by construction. We account for this misspecification with a Monte Carlo-based correction to the 'simple' Bayesian estimates.

Keywords: Hysteresis; Bayesian methods; unit roots; permanent shocks.
JEL Classification: E2, E3.

*We are grateful to our discussant Thomas Drechsel and Tamim Bayoumi, Olivier Blanchard and Francesco Furlanetto for comments; Juan Rubio-Ramirez for useful discussions, comments, and suggestions; and Rodney Strachan for extremely helpful suggestions on Bayesian cointegration. We also wish to thank seminar participants at Norges Bank, King's College, the University of Lausanne, the College of Europe, the 2022 NBER meeting on 'Methods and Applications for DSGE Models' and the 2022 NBER-NSF Time Series Conference for comments and suggestions. 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, P.O. Box 27622, Richmond, VA 23261. Tel.: +1-804-697-8246. Email: thomas.lubik@rich.frb.org.

1 Introduction

The term hysteresis captures the notion that economic disturbances that are generally regarded as transitory can have very long-lived or even permanent effects. A typical example is a monetary policy shock, which is usually associated with short-run economic stabilization efforts by means of affecting aggregate demand. Policy shocks as well as other demand shocks could have long-lasting effects on productivity and output through channels such as business formation or exit. Similarly, hysteresis effects can arise in the labor market, where long and deep recessions can lead to long-term unemployment and could raise the natural (or equilibrium) unemployment rate by permanently changing job-finding prospects.

The experience of the Great Recession and of the recession caused by the COVID-19 pandemic have rekindled interest in the possible presence of hysteresis in macroeconomic data. While there is a substantial body of work that studies channels for the transmission of hysteresis shocks theoretically, conclusive empirical evidence is scarce. The fundamental challenge is to identify a second driver of the long-run component of GDP and other macroeconomic aggregates besides the trend in total factor productivity. We interpret this second driver as hysteresis, which we define as the presence of aggregate demand shocks with a permanent effect on output.

In this paper, we search for hysteresis by developing a statistical framework that allows us to distinguish between these permanent components. We use a Bayesian cointegrated VAR and apply it to post-WW2 US data. We identify permanent aggregate demand and aggregate supply shocks by imposing a combination of zero restrictions on their long-run impact on GDP and sign restrictions on their short- and their long-run impacts on GDP and the price level. The identifying restrictions are derived from a standard macroeconomic model of aggregate demand and supply and are consistent with a wide range of theoretical frameworks, such as the DSGE models that are used in central banks.

Our paper makes three main contributions. First, we find a non-negligible, but modest extent of hysteresis in the data. Hysteresis shocks explain roughly 20 percent of the frequency zero (i.e., long-run) variance of real GDP and similar magnitudes of other macroeconomic aggregates. When we exclude the Great Recession and its aftermath from the sample, the degree of hysteresis drops by half, suggesting a role for long and deep recessions in driving hysteresis effects.

However, we show via Monte-Carlo analysis of a theoretical DSGE model with a hysteresis mechanism that reliably detecting hysteresis in macroeconomic data is a significant challenge. In particular, we demonstrate that model-consistent identification schemes spuriously detect hysteresis with non-negligible probability when there is none in the data-generating process. In addition, such identification schemes tend to have low power to discriminate between alternative extents of hysteresis.

Our second contribution lies in developing a Monte-Carlo approach to test for the presence and extent of hysteresis, which also corrects for potential misclassification bias. For a given estimated Bayesian VAR we construct a data-generating process (DGP), from which we simulate specific statistics, such as the fraction of the long-run output variance that is due to hysteresis shocks. Conditional on alternative degrees of hysteresis imposed on the DGP we obtain Monte-Carlo distributions of the statistics of interest, which we can compare to the corresponding distributions based on the actual data using the Kolmogorov-Smirnov (KS) statistic. Our bias-corrected estimate of hysteresis is thus the value of the target statistic that minimizes KS. When applied to our dataset we find an even more modest extent of hysteresis, equal to about 10 percent of the long-run variance of GDP. When we exclude the Great Recession period from the sample, any evidence of hysteresis disappears.

The third contribution is methodological. To the best of our knowledge, our paper is the first that implements short-run and long-run sign restrictions jointly. Our time-series model is a Bayesian cointegrated VAR based on Strachan and Inder (2004) and Koop et al. (2010). We choose a cointegration framework instead of a more conventional VAR setting since we want to utilize the information embedded in long-run comovement between consumption, investment and GDP for inference on two possible long-run components. Shock identification is based on the long-run restrictions of Blanchard and Quah (1989) and the sign-restriction approach of Arias et al. (2018). Our innovation is to apply the latter to the long-run, too, since it allows us to distinguish between the effects of a permanent supply and a permanent demand shock.

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. 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) argues that the evidence is not sufficiently clear-cut to reach strong conclusions. Cerra et al. (2022) give an overview over the state of the empirical and theoretical literature.

Among empirical studies, Cerra and Saxena (2008) use growth regressions and show that in a panel of 190 countries over the period 1960-2001 deep recessions caused by financial crises permanently reduced the productive capacity of an economy. In a similar vein and more recently, Aikman et al. (2022) show that deep contractions in and of themselves lead to permanent scarring.

Ball (2009) applies 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 argues for the presence of hysteresis, in the end he does not distinguish between permanent and highly persistent effects. Galí (2015, 2020) takes up Ball's analysis and integrates 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 argues in favor of hysteresis as being an important driver of the European unemployment and wage/price inflation experience.

Modeling hysteresis in theoretical frameworks goes back to at least the contribution of Ljungqvist and Sargent (1998). In contrast with the other papers, they provide a more microfounded theoretical approach using a search-and-matching framework in which previous employment experiences inform a worker's reservation wage. When laid off during a long and deep recession, workers are reluctant to accept job offers at lower wages during a recovery, holding out for pre-lay off wages. Since skills and attachment to the labor force deteriorate over time, a subset of these workers will drop out, leading to permanently lower employment. Offering corroborating evidence based on labor market data, Ljungqvist and Sargent (1998) confirm this mechanism in European data. It is such a mechanism that we have in mind in driving employment-based hysteresis in our empirical findings.

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; Jaimovich and Siu, 2020; or Sim, 2022). The general idea in these models is that firm entry is a key determinant of TFP in that young firms develop new products that expand production and improve productivity. When such firms cannot find financing or a market to sell into because of an economic downturn, they may not come into existence and output would be permanently lower. In our DSGE-based simulation analysis, we utilize such a mechanism for inducing hysteresis effects.

On the empirical side, Furlanetto et al. (2021) is the paper that is closest to our work. They also use a structural VAR framework that combines zero long-run restrictions with short-run sign restrictions, whereas we also impose long-run sign restrictions in a cointegration setting. In contrast with our results, they do detect an important role for hysteresis effects in U.S. data. As we discuss in the next section, one of the differences between Furlanetto et al. (2021) and the present work is that they use a weaker set of restrictions in order to identify hysteresis shocks. 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 discusses our identifying restrictions for hysteresis shocks. In section 3 we outline our empirical methodology and the VAR specification in detail. We present the empirical findings from a baseline specification and some robustness checks in section 4. Section 5 illustrates the difficulty of robust identification of hysteresis effects in macroeconomic data even when based on model-consistent identification schemes. To this end we perform a Monte-Carlo analysis of a DSGE model with embedded hysteresis. Section 6 discusses our proposed solution to this problem, based on a Monte Carlo-based correction to the 'simple' estimates produced by the Bayesian procedure. We also show and discuss estimates produced by the proposed Monte Carlo-based correction. Section 7 concludes. An extensive Online Appendix provides additional details on the methodology and on several aspects of our empirical work.

2 Identifying Hysteresis Shocks

It is often assumed in theoretical and empirical macroeconomics that permanent changes in GDP originate on the supply side of the economy, for instance, in form of permanent shocks to total factor productivity (TFP). The key idea underlying the notion of hysteresis is that some permanent shocks to output could instead originate on the demand side. The challenge for empirical researchers is therefore to separate permanent supply from permanent demand shocks and also from persistent, but transitory changes in macroeconomic aggregates. We now discuss our identification assumptions on how to accomplish this.

We assume that the permanent component of real GDP is driven by two shocks that we associate with aggregate supply (AS) and aggregate demand (AD) disturbances. 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. They are also consistent with a wide range of theoretical frameworks, such as medium- or large-scale DSGE models that are used in policy institutions. Specifically, we make the following assumptions:

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

(II) the AS curve is upward-sloping in the short run, but vertical in the long run;

(III) BQ shocks only affect the AS curve;

(IV) H shocks affect the AD curve and can affect the AS curve in the case of hysteresis, when 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 range of macroeconomic frameworks such as the textbook AD-AS model, simple New Keynesian models, and the large-scale macroeconomic models. The fourth assumption captures the essence of the notion of hysteresis: A negative (positive) AD shock has a negative (positive) permanent impact on long-run aggregate supply. As an example, deep recessions can 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, thereby increasing potential output. Figure 1 provides an illustration of our identification assumptions in output (Y)-price level (P) space.

In the first panel, BQ shocks in the short run shift the AS curve and therefore move the economy along the AD curve. This causes output and prices to move in opposite directions. Similarly, H shocks move the economy along the AS curve with output and prices moving in the same direction. Consequently, the first set of restrictions we impose in order to disentangle the two shocks is that on impact and up to h quarters, BQ shocks generate impulse responses for output and prices with *opposite signs*. In contrast, H shocks produce impulse responses with the *same sign*.[1] The second panel of Figure 1 illustrates assumption (III), namely that in the long run BQ shocks only affect the AS curve. This results in permanent effects on output and prices with *opposite signs*.

As for H shocks there are three possible cases. First, the case of no hysteresis in the third panel shows that the long-run impact on output is *zero*, while the *sign* of the long-run impact on prices is the *same* as the sign of the short-run impact. In the case of hysteresis

[1] These restrictions are the same as Canova and Paustian's (2011) DSGE-based 'robust sign restrictions' (see their Table 2, page 348). Their model features two demand shocks, namely a monetary and a preference shock. For both shocks, the impact on output and inflation are of the same sign.

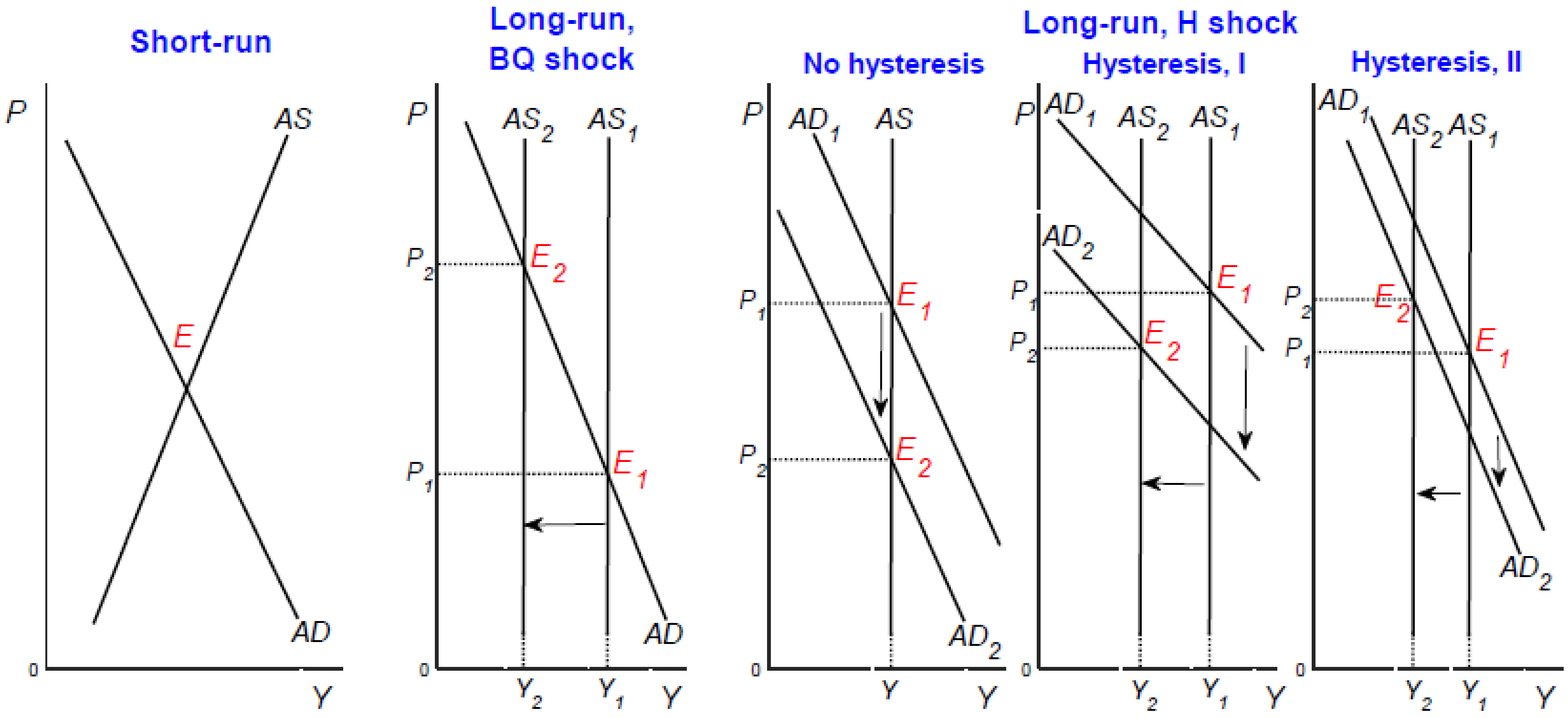

Figure 1 A simple illustration of the identifying assumptions

(fourth and fifth panels) 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). Consequently, the long-run effect on output is unambiguously negative (positive), so that its sign is the *same* as that of its short run response to the H shock. The long-run impact on prices, on the other hand, is ambiguous. It depends on the size of the hysteresis effect on output and 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 as in the fourth and fifth panels. In the limit case of no hysteresis on output in the long run (third panel), the long-run impact on prices of a negative (positive) H shock is unambiguously negative (positive).

In general, however, this pattern cannot be assumed. Some 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. A key difference between our work and Furlanetto et al. (2021) is that they do not impose sign restrictions on the long-run impact of BQ and H shocks on output and the price level. Rather, they only impose the restriction that BQ and H shocks are the only two disturbances that have a permanent impact on output and employment (see their Table 1). In what follows, we therefore consider two sets of results, where we alternatively impose and not impose this restriction on prices. To anticipate the results, both assumptions lead to qualitatively the same outcomes, so that in practice imposing or not imposing this restriction does not appear to make a material difference. We summarize the previously discussed restrictions in Table 1.

Table 1: Identifying Restrictions

	Shock:				
Impact on:	BQ	H	BQ	H	
	Short run:		Long run:		
Output	$+$	$+$	$+$	0	$+$
Price level	$-$	$+$	$-$	$+$	$+$

3 Empirical Approach

We now discuss our empirical specification, how we estimate the model via Bayesian methods, our choice of priors, and how we implement the identifying restrictions outlined in Table 1. Our baseline empirical specification includes the logarithms of real GDP, real consumption, real investment, and total hours worked per capita; a long- and a short-term nominal interest rate; and the logarithm of the PCE deflator. We collect these variables in the vector $Y_t = [y_t, c_t, i_t, R_t, r_t, p_t, h_t]'$, where y_t, c_t, i_t, and h_t are, respectively, the logarithms of GDP, consumption, investment, and total hours worked per capita; p_t is the logarithm of the PCE deflator; and R_t and r_t are a short- and a long-term nominal interest rates. For the short-term rate we use the 'shadow rate' from Wu and Xia (2016) to capture the effective policy stance during the zero-lower bound period. The long-term rate is the 5-year government bond yield.

The data and their sources are described in detail in Online Appendix A. We consider two sample periods, 1954Q3-2008Q4 and 1954Q3-2019Q4, whereby the latter sample includes the Zero Lower Bound (ZLB) period. As discussed in Online Appendix B, Elliot et al.'s (1996) tests of the null hypothesis of a unit root suggest that all of the series in Y_t are I(1), whereas a unit root is rejected for inflation.

We proceed with the empirical analysis as follows. Our baseline specification is a Bayesian cointegrated VAR given the findings from the unit root tests. Specifically, this VAR includes the logarithm of the price level. We also consider a second specification in terms of stationary variables. This is motivated by the argument that interest rates and hours per capita are likely not I(1) on conceptual grounds. Results from this specification are included in the Online Appendix, but they paint a similarly consistent picture as the baseline model.

Our approach to Bayesian cointegration is based on methods introduced by Strachan and Inder (2004) and Koop et al. (2010). We specify a VAR in form of a vector error correction model (VECM):

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

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

For the r cointegration vectors to be uniquely identified, the columns of the matrix β need to feature at least r restrictions (see Kleibergen and van Dijk, 1994, or Bauwens and Lubrano, 1996). The standard classical 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:

$$\beta'\beta = I_r, \tag{2}$$

where I_r is the $r \times r$ identity matrix.[2] This restriction is imposed by defining a semi-

[2] We also consider the "linear normalization" approach proposed by Geweke (1996). The results are

orthogonal matrix $H = H_g(H_g'H_g)^{-1/2}$, which centers 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, it can obtained by appropriately selecting the columns of H_g based on a priori information, for instance derived from economic theory.

As discussed in Online Appendix B, Johansen's tests strongly reject the null hypothesis of no cointegration for any of the three bivariate systems featuring GDP and consumption, GDP and investment, and a short- and a long-term nominal interest rate. Similarly, the same tests identify three cointegration vectors for the baseline 7-variables system in $Y_t = [y_t, c_t, i_t,\ R_t, r_t, p_t, h_t]'$ for the period 1954Q3-2019Q4 and three for the period 1954Q3-2008Q4. In the following, we therefore proceed under the assumption that the baseline system features three cointegration vectors for either sample.

In light of this evidence, we center the prior for β at the three cointegration relationships consistent with three bivariate systems. Specifically, this is obtained by setting H_g to:

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

The rows of H_g' correspond to the normalized cointegration vectors for GDP and consumption, GDP and investment, and the short and the long rate, respectively. This choice 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. On the other hand, investment typically tends to increase somewhat faster than GDP. For example, in the U.S. over the period 1947Q1-2019Q4 real GDP has increased by 1.97 on a log scale, whereas the corresponding increase for real investment has been 2.75. The increase for the former series is 71.6 per cent of the increase for the latter. We want to stress, however, that given the uninformativeness of the prior for β (discussed in Appendix A), the centering encoded in this choice for H_g imposes, in fact, very weak restrictions.

We estimate the cointegrated VAR using the Gibbs-sampling algorithm proposed by Koop et al. (2010). The algorithm iterates on four steps: (i) the loading matrix α in equation (1), (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 discard. We then generate 1,000,000 draws that we thin by sampling every 100 draws in order to reduce their autocorrelation.

qualitatively similar, but we use Strachan and Inder (2004) because it is more general.

This leaves us with 10,000 draws from the ergodic distribution, which we use for inference.

Our prior choice is informed by Geweke (1996) and Koop et al. (2010), from which we also take the conditional distributions to draw from the posterior described in steps (i)-(iv) above. The priors are generally not very informative. In particular we impose a flat prior on the matrix of the cointegration vectors β. We discuss the priors in more detail in Appendix A. Figure A.20 in the Online Appendix reports evidence on the convergence of the Markov chain, by showing Geweke's (1992) inefficiency factors of the draws for each individual parameter. For all parameters the factors are equal to at most 8, well below the values of 20-25 that are typically taken to indicate problems in the convergence of the Markov chain.

We jointly identify a BQ and an H shock using a combination of zero long-run restrictions on GDP and both short- and long-run sign restrictions on GDP and the price level; that is, we impose the restrictions summarized in Table 1 using the methodology proposed by Arias et al. (2018) and adapted for our Bayesian cointegrated VAR. Further details on the procedure are relegated to the Online Appendix. We impose the short-run sign restrictions both on impact and for the subsequent either four or eight quarters. Since the two sets of results are qualitatively very similar, we focus on those based on imposing the restrictions for eight quarters. For each draw from the posterior distribution of the reduced-form VAR we consider 100 random rotation matrices that we draw based on Arias et al.'s (2018) algorithm. We set the number of Gibbs-sampling iterations in the algorithm to 10.

4 Empirical Evidence on Hysteresis

We report three sets of empirical results. Figure 2 shows the impulse response functions (IRFs) of our baseline model to BQ and H shocks, while Figure 3 contains fractions of forecast error variance (FEV) of the variables explained by the shocks over the same horizon. Table 2 reports the posterior distributions of the fractions of the long-run variance of GDP and consumption explained by H shocks, which is a key metric in assessing the extent of hysteresis. As discussed above, the baseline specification is a cointegrated VAR estimated for the sample including the ZLB period and without imposing restrictions on the long-run impact of H shocks on the price level. Figures A.4 and A.5 in the Online Appendix show the corresponding results for the sample excluding the ZLB period. Figures A.6-A.9 report for either sample IRFs and fractions of FEV obtained by imposing the restriction that an H shock causing a permanent increase in GDP, consumption, and investment also causes a permanent increase in the price level.

4.1 Impulse Response Functions

Figure 2 shows the median and the 16-84 and 5-95 percentiles of the posterior distributions of the IRFs to BQ and H shocks. The responses are essentially as expected and are, in fact, the result of the restrictions imposed on GDP, consumption, investment and the price level. Specifically, a BQ shock permanently increases the quantity variables and lowers the price level as would be consistent with a permanent TFP shock. A similar pattern applies to the IRFs to an H shock, with the exception of the price level. In contrast, the responses of the interest rates to either shock is insignificant at long horizons although the median responses are positive. The response of hours is estimated to be permanent for either shock, so that a positive BQ or H shock causes a long-run increase in hours worked.[3]

Turning to the price level, the IRFs for the two samples obtained without imposing restrictions on the long-run impact of H shocks reveal an interesting observation. In the full sample including the ZLB (Figure 2) a small, but non-negligible mass of the posterior distribution of the IRF to H shocks lies below zero. Moreover, the long-run impact of H shocks on the price level is estimated to be negative for 18.9 percent of the draws. This suggests that the data provide some support to the notion discussed in Section 2 (see the panel labelled as 'Hysteresis II' in Figure 1) that a positive (negative) H shock may counterintuitively have a negative (positive) long-run impact on the price level.

In the sample that excludes the ZLB period (reported in Figure A.4 in the Online Appendix) 50 percent of the draws are associated with a negative impact of H shocks on prices at the 15-year horizon. In the long run, the corresponding fraction is equal to 65.6 per cent. Taken at face value these results imply that possibility illustrated in the panel labelled as 'Hysteresis II' in Figure 1 is significantly more likely than in the full sample. At the same time, the fallout from the financial crisis was a persistently high unemployment and a historically high long-term unemployment rate, which can be seen as a prime example of hysteresis. Including this time period in the estimation thus appears to produce prima facie evidence for the presence of hysteresis shocks, evidenced by an increase in the price level.

However, there is a second possible interpretation. It may very well be that in this sample H shocks are virtually absent. The result is then simply an artifact of the identifying assumptions, namely that our model imposed the existence of an H shock on the data by brute force. Suppose that the true data-generating process (DGP) only features BQ shocks.

[3] The IRFs produced by stationary VARs are closely in line with the evidence from the cointegration specification. The main difference is that by construction the responses of hours mean-revert to zero.

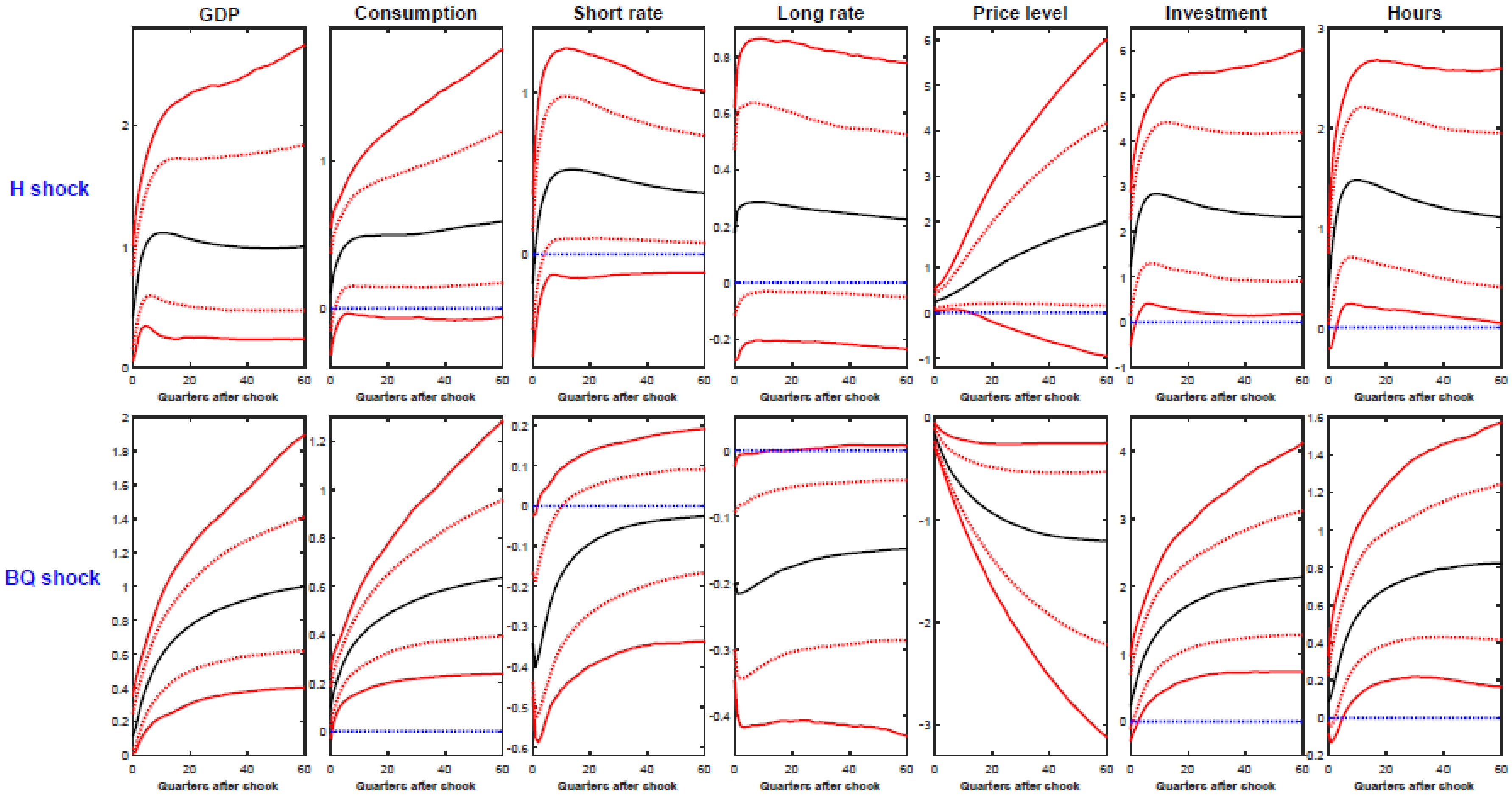

Figure 2 Impulse-response functions to H and BQ shocks, based on the 7-variables cointegrated VAR (including the Zero Lower Bound)

Since we impose only short-run restrictions on the price level's response to H shocks the set of identified H shocks is obtained by retaining all of the models randomly generated by Arias et al.'s (2018) algorithm, for which one of the two permanent GDP shocks exhibits aggregate-demand features in the short run.

At the same time, we do not impose any restriction on the long-run impact of H shocks on prices; moreover, all permanent GDP shocks cause a permanent decrease in the price level. Consequently, the response of prices to H shocks will tend to be negative in the long run. This insight offers a suggestion that there is the possibility of spuriously identifying hysteresis when the DGP features none. We investigate this issue in the next section.

4.2 Forecast Error Variance

Figure 3 shows the fractions of the forecast error variance that is explained by H and BQ shocks for our baseline specification. The two shocks jointly explain very large fractions of the FEV of GDP at long horizons, namely about 80 percent. A similar picture emerges from the other specifications (reported in the Online Appendix) for either sample period and any specification, whether cointegrated or stationary VARs, or either restricting or not restricting the long-run impact of H shocks on the price level. We also find that H shocks explain between 10 and 20 percent of the FEV for the other variables at the 15-year horizon, with hours worked at the upper end of this range.

It is noteworthy that the fraction of the consumption FEV is overall fairly small, ranging between about 4 and 10 percent, whereas the corresponding fraction explained by BQ shocks is between 60 and 65 percent. This is an important finding since the stochastic trend in GDP carries over closely to consumption (see Cochrane, 1994). It should therefore be informative about issues pertaining to the unit root component of output. The fact that H shocks are estimated to play such a small role for consumption suggests that if hysteresis effects are present in the data they are likely small. In addition, in the baseline specification over the full sample and without restricting the long-run impact of H shocks on prices, H shocks explain 14 percent of the overall fraction of the FEV of GDP at the 15-year horizon. The corresponding value for the sample excluding the ZLB period is 6 percent.

We now dig deeper into the contribution of H shocks by computing the frequency-zero, or long-run, variance of GDP and consumption. Table 2 reports the median and the 16-84 and 5-95 percentiles of the respective posterior distributions attributable to H shocks. We find that the estimates for consumption are consistently smaller than those for GDP, in line with the previous discussion. While the difference is small for the full sample, when we

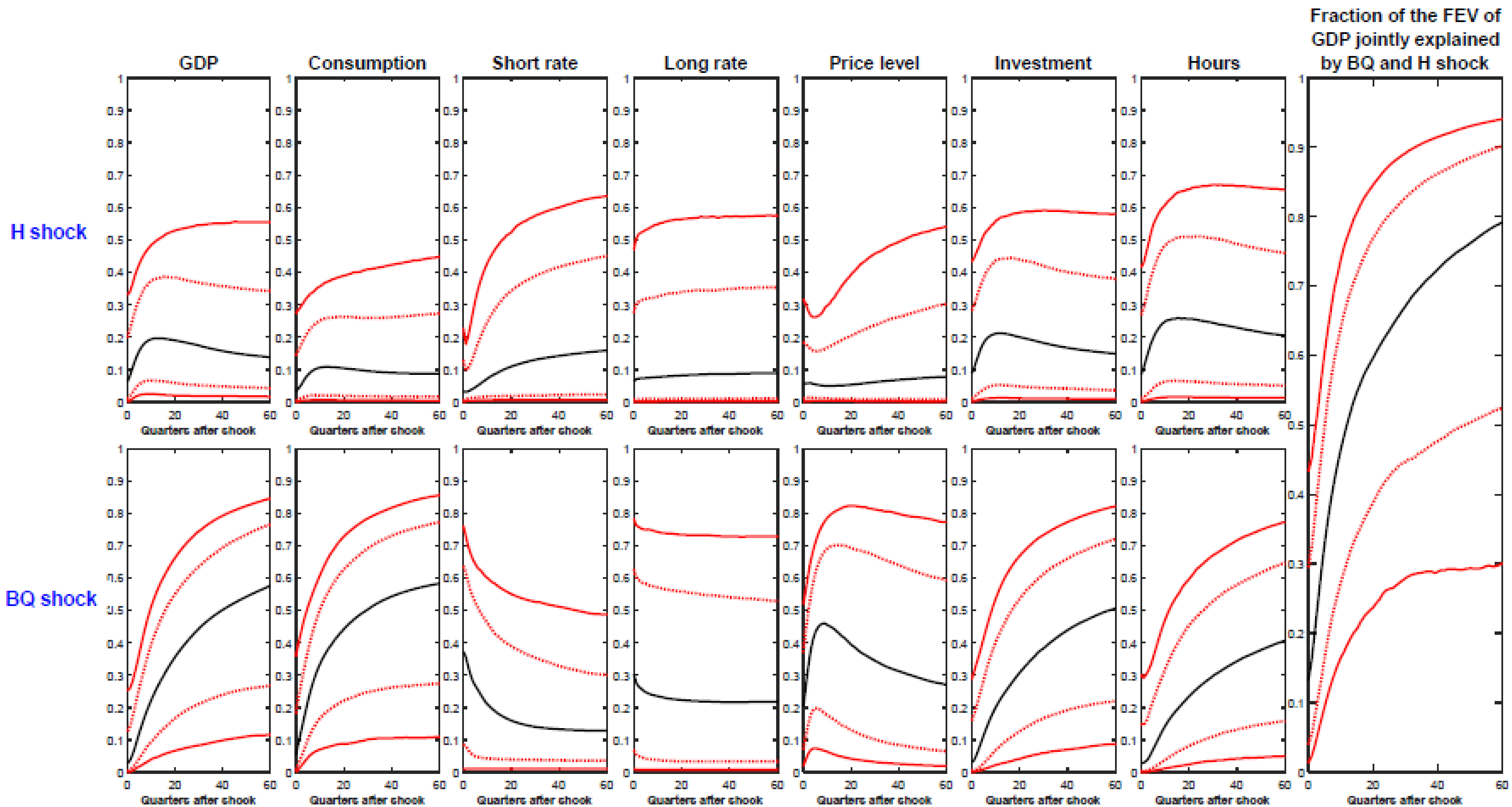

Figure 3 Fractions of forecast error variance explained by H and BQ shocks, based on the 7-variables cointegrated VAR (including the Zero Lower Bound)

exclude the ZLB period it is non-negligible. Based on our preferred model, namely without restricting the long-run impact of H shocks on the price level, the median estimates are 22.1 percent for GDP and 11.9 percent for consumption.

Table 2 Fractions of frequency-zero variance of GDP and consumption explained by H shocks (median and 16-84 and 5-95 percentiles)

	Restricting the long-run impact of H shocks on the price level:	
	No	Yes
1954Q3-2019Q4		
GDP	0.198 [0.033 0.840] [0.008 0.981]	0.221 [0.030 0.853] [0.007 0.983]
Consumption	0.175 [0.018 0.804] [0.002 0.962]	0.218 [0.020 0.819] [0.003 0.964]
1954Q3-2008Q4		
GDP	0.221 [0.045 0.689] [0.012 0.929]	0.117 [0.022 0.582] [0.004 0.913]
Consumption	0.119 [0.013 0.544] [0.002 0.843]	0.056 [0.005 0.403] [0.001 0.798]

When we focus on consumption, the evidence suggests that estimates for the sample excluding the ZLB are materially lower than those for the full sample. Specifically, median estimates for the samples including and excluding the ZLB period, respectively, are 0.175 and 0.119 when not imposing restrictions on the long-run impact of H shocks on the price level. When imposing such restrictions, the corresponding values are 0.218 and 0.056. This shows that for the sample excluding the ZLB period hysteresis effects are very small to negligible and that they only appear when also considering the years including the financial crisis and the Great Recession.

In a sense, this is not surprising as it is that period that likely has seen hysteresis effects at play because of historically high long-term unemployment. But what also plays a role is the assumption on the price level response to H shocks. In particular, when we restrict the long-run impact of H shocks on prices we typically find smaller estimates of the fraction of the unit root of either GDP or consumption explained by H shocks. This is especially apparent in the case of stationary VARs.

Overall, the evidence from Bayesian VARs points towards a modest, but non-negligible extent of hysteresis. In our baseline specification with a cointegrated VAR this amounts to 20 percent of the long-run variance of GDP for the sample including the ZLB period. When we exclude the ZLB period, the evidence is more mixed, but overall points towards a smaller role played by H shocks in driving long-horizon variation in GDP and consumption. The results are thus consistent with the idea that hysteresis effects, to the extent that they are statistically detectable, largely operate in the wake of long and deep recessions. As discussed above, potential economic mechanism include long-term unemployment and firm entry and exit, both of which played a prominent role during and after the Great Recession.

This finding is confirmed by the evidence presented in Figure 4, which shows the estimated H and BQ shocks for the full sample from 1990 on. We report the posterior medians and the 16-84 and 5-95 percentiles. What clearly stands out is the large negative hysteresis shock in 2008Q4 after the collapse of Lehman Brothers, which was the signal event for the financial crisis and the ensuing Great Recession. Moreover, zero lies significantly outside of the posterior coverage intervals, more so than for any other estimated shock. The permanent downward shift in the path of real GDP after 2008 is therefore not driven by a decline in TFP but hysteresis. Similarly, this also rationalizes our finding that before the Great Recession evidence for hysteresis is weak to non-existent.

There are two caveats to this result. First, the degree of uncertainty is fairly high as is evidence by the wide spread of the coverage regions in Figures 1 and 2. This is largely due to the nature of the problem econometricians are facing when attempting to extract information pertaining to either long horizons or in the limit the long run from finite data samples (see Faust and Leeper, 1997). A second caveat is that our approach imposes on the data the presence of H shocks by construction. This raises the obvious question of whether the small, but non-negligible role of H shocks we have identified is a genuine result, or whether it is an artifact of our identification strategy.

We tackle this issue in the following section. We show that the problem is indeed potentially serious by conducting a simulation exercise in a simple DSGE model featuring the possible presence of hysteresis. In particular, we show that model-consistent identification schemes detect a non-negligible, and possibly sizeable role of H shocks even when the DGP features no hysteresis by construction. We then propose a simple Monte Carlo-based correction designed to remedy this issue.

5 Pitfalls in Identifying Hysteresis Shocks

In our baseline implementation we pursued the strategy of jointly identifying a BQ and an H shock, that is, two permanent shocks that affect GDP, the former through what may be deemed supply-side effects, and the latter via an aggregate demand channel. Arguably, this is the most natural way of imposing the restrictions in Table 1. However, this implementation suffers from the potential shortcoming that it may spuriously detects hysteresis when there is none in the underlying DGP. The reason for this is straightforward. Since this strategy jointly identifies both a BQ and H shock, it, in fact, imposes on the data the very existence of hysteresis by assumption. Ultimately, it is an empirical issue how problematic this is in an assessment of hysteresis. To that end, we conduct a simulation study that

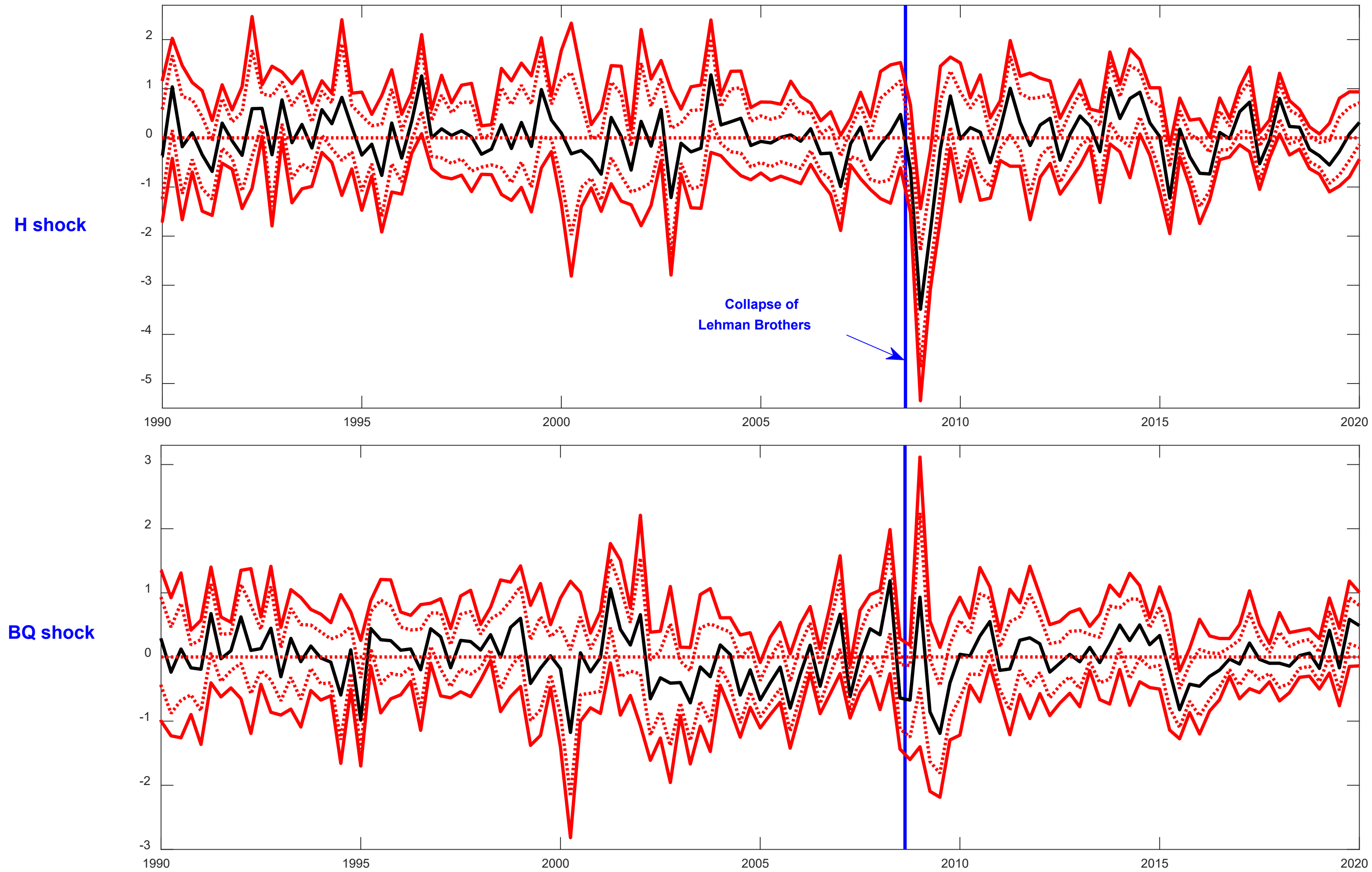

Figure 4 Medians and 16-84 and 5-95 percentiles of the posterior distributions of the H and BQ shocks

explores and quantifies the potential for misidentification.

5.1 The Data-Generating Process

The theoretical model that serves as the DGP for the Monte-Carlo study is a version of the model described in Galí (2015) augmented with habit-formation in consumption and a unit root in technology (and therefore in the natural level of output). Hysteresis is introduced in a somewhat reduced-form manner, designed to capture the essence of the mechanism without taking a precise stand on its origin. The model is described in detail in Online Appendix C. In the following we therefore only provide an outline of its structure.

Households solve the optimization problem:

$$\underset{C_t, N_t}{Max} E_0 \sum_{t=0}^{\infty} \beta^t \left[\ln(C_t - bC_{t-1}) - \frac{N_t^{1+\phi}}{1+\phi} \right], \tag{4}$$

subject to:

$$P_t C_t + Q_t B_t = B_{t-1} + W_t N_t - T_t, \tag{5}$$

where C_t and N_t are real consumption and hours worked; $0 < b < 1$ 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 time-t price of a nominal bond paying one 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 standard.

We follow Galí (2015) in modeling the firms, the only difference being 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. They maximize profits with respect to N_t. We assume that the logarithm of technology, $a_t = \ln(A_t)$ evolves according to:

$$a_t = a_{t-1} + \epsilon_t^a + \delta \tilde{y}_{t-1}, \tag{6}$$

where $\epsilon_t^a \sim N(0, \sigma_a^2)$ is the technology shock, $\tilde{y}_t$ the transitory component of output, and $\delta \geq 0$ captures 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 is conceptually the same as in Jordà et al. (2020). It 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 interest-rate smoothing:

$$R_t = \rho R_{t-1} + (1 - \rho)\phi_\pi \pi_t, \tag{7}$$

where the notation is standard.[4]

We calibrate the structural parameters as follows, based on standard values in the literature: $\beta = 0.99$, $\alpha = 1/3$, $b = 0.8$, $\phi = 1$, $\rho = 0.9$, $\phi_\pi = 1.5$, and $\sigma_a = 0.007$.[5] The model is augmented with three transitory AR(1) disturbances; two of them (v_t and u_t) can be thought of as supply shocks, while the third (e_t) is a demand shock. We also include an additive disturbance to the log of the production function, $z_t \sim N(0, \sigma_z^2)$, with $\sigma_z = 0.005$. The autoregressive parameters are calibrated as $\rho_v = \rho_u = \rho_e = 0.75$, while 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 $\sigma_u = 0.001$, $\sigma_v = 0.005$ and $\sigma_e = 0.045$.

Based on this calibration the permanent technology shock (ϵ_t^a) explains one-third of the forecast error variance of log GDP on impact. In the absence of hysteresis, when $\delta = 0$, it explains about 96 per cent of GDP's FEV 15 years ahead. These numbers are broadly in line with the evidence produced by the structural VAR literature (see for example Cochrane, 1994). Moreover, the demand innovation ϵ_t^e is essentially the only driver of the transitory component of output. In that sense, this validates the identifying restrictions in Table 1. In the Monte Carlo exercise we consider a grid of values for the hysteresis parameter δ, ranging from $\delta = 0$ (no hysteresis) to $\delta = 0.1256$. In the latter case, the demand shock ϵ_t^e explains 20 per cent of the long-run variance of output, while at long horizons the other three shocks contribute essentially nothing to the variation in output.

Figure 5 shows impulse responses for output and prices 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 $\delta = 0$ and $\delta = 0.1256$.[6] The dynamic behavior of the model suggests that it is an appropriate DGP for assessing the ability of the identifying restrictions in Table 1 to recover the extent of hysteresis in the data. First, for either of the two values of δ (as well as for any value inbetween) 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 $\delta = 0$, ϵ_t^e is the dominant driver of output at short horizons and of prices at medium-to-long horizons. Second, it is apparent from the impulse responses how ϵ_t^a and ϵ_t^e satisfy the restrictions for

[4] Online Appendix C reports the log-linearized equations of the stationarized model. Output and the real wage inherit the unit root in a_t, and therefore need to be stationarized for standard solutions methods for linear rational expectations models 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. $\tilde{y}_t$ is thus the component of output that is driven by the transitory disturbances.

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

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

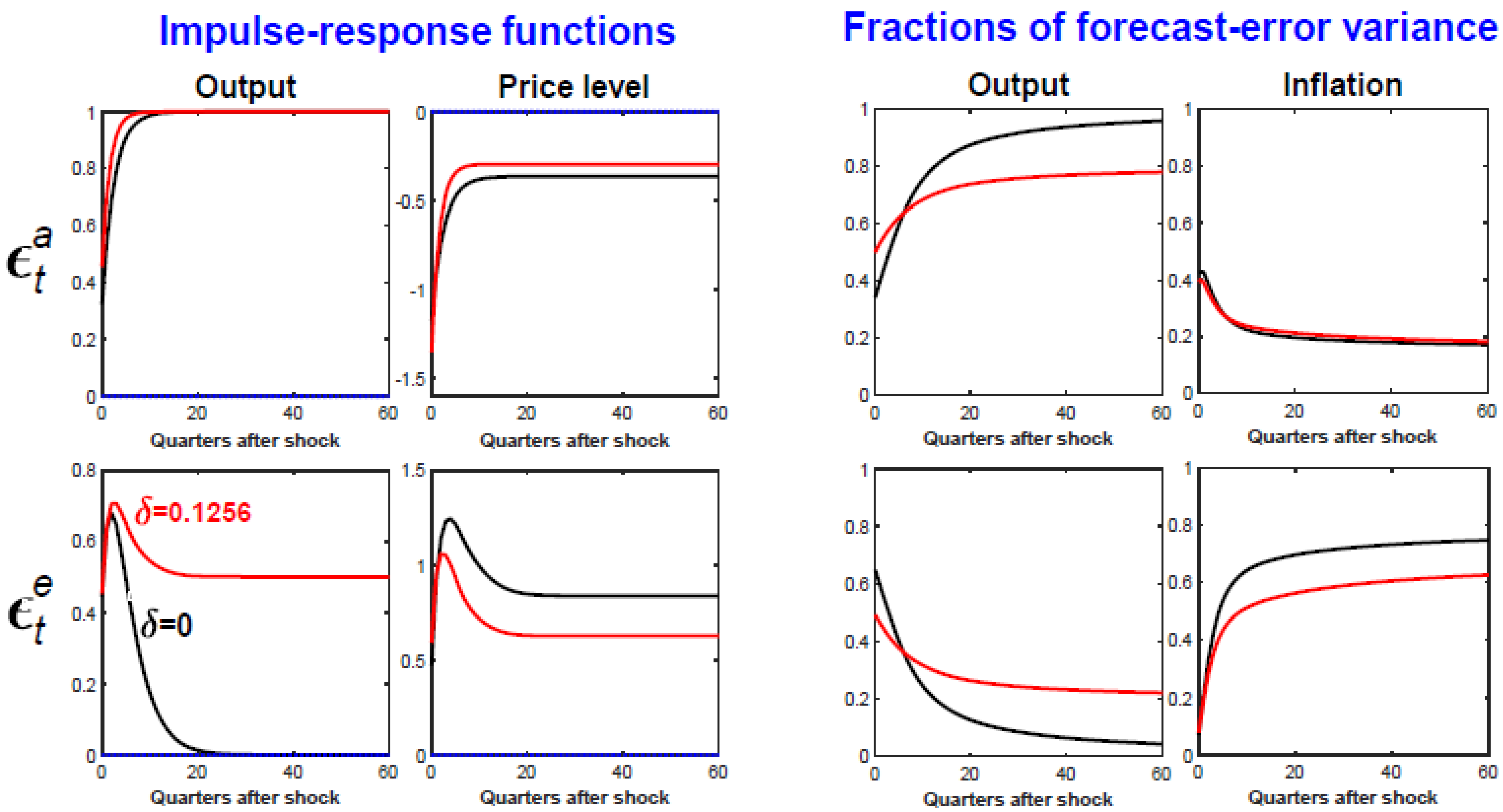

Figure 5 Theoretical impulse-response functions and fractions of forecast error variance for the RBC model for output and prices, for ϵ_t^a (BQ) and ϵ_t^e (H) shocks

BQ and H shocks. In particular, with $\delta = 0.1256$ (and in fact for any $\delta > 0$) the long-run impact of a positive H shock ϵ_t^e on the price level is positive instead of being ambiguous as in the last two panels in Figure 1. Accordingly, we impose this restriction for the identification of H shocks in the Monte Carlo exercises below.

5.2 Empirical Assessment

We simulate the model for samples of length equal to the actual U.S. sample 1954Q3-2019Q4, i.e., 261 observations. For each simulation we run a pre-sample of 100 observations that we then discard. We then estimate Bayesian VARs for output (y_t), the real wage (ω_t), hours (n_t), inflation (π_t), and technology (a_t) on each artificial sample. Whereas n_t and π_t are I(0), y_t and ω_t inherit the unit root in a_t. They are both cointegrated with log technology with cointegration vector $[1, -1]'$.

In what follows we work with $Y_t = [\Delta y_t, (y_t - a_t), (\omega_t - a_t), \pi_t, n_t]'$, so that cointegration between a_t and either y_t and ω_t is imposed directly on the estimated system as in King et al. (1991).[7] For each Monte Carlo simulation we choose the lag order as the maximum of the lag orders chosen by the Schwartz and Hannan-Quinn criteria, based on VARs estimated by OLS. We estimate the Bayesian VARs as in Uhlig (2005). For each value of δ we perform 10,000 Monte Carlo simulations and then conduct inference by taking 10,000 draws from the posterior distribution.

5.2.1 Model-Consistent Identifying Assumptions

As a first exercise, we assume that there is no hysteresis in the model and impose the correct identification assumption on the VAR. We simulate the model for $\delta = 0$ and assess whether the VAR correctly recovers the key features of the DGP when we impose the correct identifying restrictions, i.e., that the DGP has a single shock (ϵ_t^a) with a permanent impact on output. In each Monte Carlo simulation and for each draw from the posterior distribution of the Bayesian VAR's reduced-form coefficients we identify the permanent shock as in Blanchard and Quah (1989), that is, as the only shock having a permanent impact on y_t.

We then compute impulse responses and fractions of FEV and report the median and various percentiles of their posterior distributions to build up the Monte Carlo distributions of the medians and the percentiles of the IRFs and fractions of FEV. The results are reported in Figure A.3 in the Online Appendix. In brief, the evidence confirms that the

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

VAR recovers the main features of the DGP in terms of both IRFs and fractions of FEV with great precision. This is, in particular, the case for the main series of interest within the present context, namely output.

5.2.2 Baseline Identifying Assumptions

In the next exercise we impose the identifying assumptions as summarized in Table 1 and discussed in section 2. We first impose the no-hysteresis specification and analyze whether imposing joint identification of two permanent shocks leads to a spurious detection of hysteresis. We then simulate the model under hysteresis and assess whether the identification schemes come to the correct conclusion and to what extent the identification schemes lead to a spurious detection.

No hysteresis in the theoretical model In our baseline we jointly identify a BQ and an H shock, which imposes the presence of hysteresis a priori. Given that the DGP features no hysteresis, we explore to what extent the restrictions in Table 1 tend to spuriously detect it. We jointly impose the zero long-run restrictions and the short- and long-run sign restrictions, implemented using methodology introduced by Arias et al. (2018). We impose the short-run sign restrictions both on impact and for the subsequent eight quarters. Since the long-run impact of a positive hysteresis shock ϵ_t^e on the price level is consistently positive for all values of δ, we impose this restriction in the identification of H shocks.

The first column of Figure 6 shows the results. The two panels show the averages of the medians and percentiles of the posterior distribution (top panel) and posterior cumulative distribution (bottom panel) of the statistics of interest, namely the fraction of the output variance explained by identified H shocks at $\omega = 0$, together with the corresponding true fractions of the output variance explained by ϵ_t^e in the theoretical model.[8] The conclusion drawn from this exercise seems fairly clear-cut. Our baseline identification approach tends to detect hysteresis when there is none in the DGP with non-negligible and quite sizeable probability. The likely reason for this is that by its very logic this scheme imposes on the data the very existence of both BQ and H shocks.

Hysteresis in the theoretical model Columns 2 to 4 of Figure 6 report the corresponding sets of results for four different values of the true fraction of the output variance

[8] The Bayesian SVAR produces a posterior distribution of the fraction of the variance of output explained by H shocks at $\omega = 0$. For each simulation we extract the median and the percentiles of this distribution and thus build up the Monte Carlo distribution of these objects. Figure 6 shows the means of these objects taken across the Monte Carlo simulations.

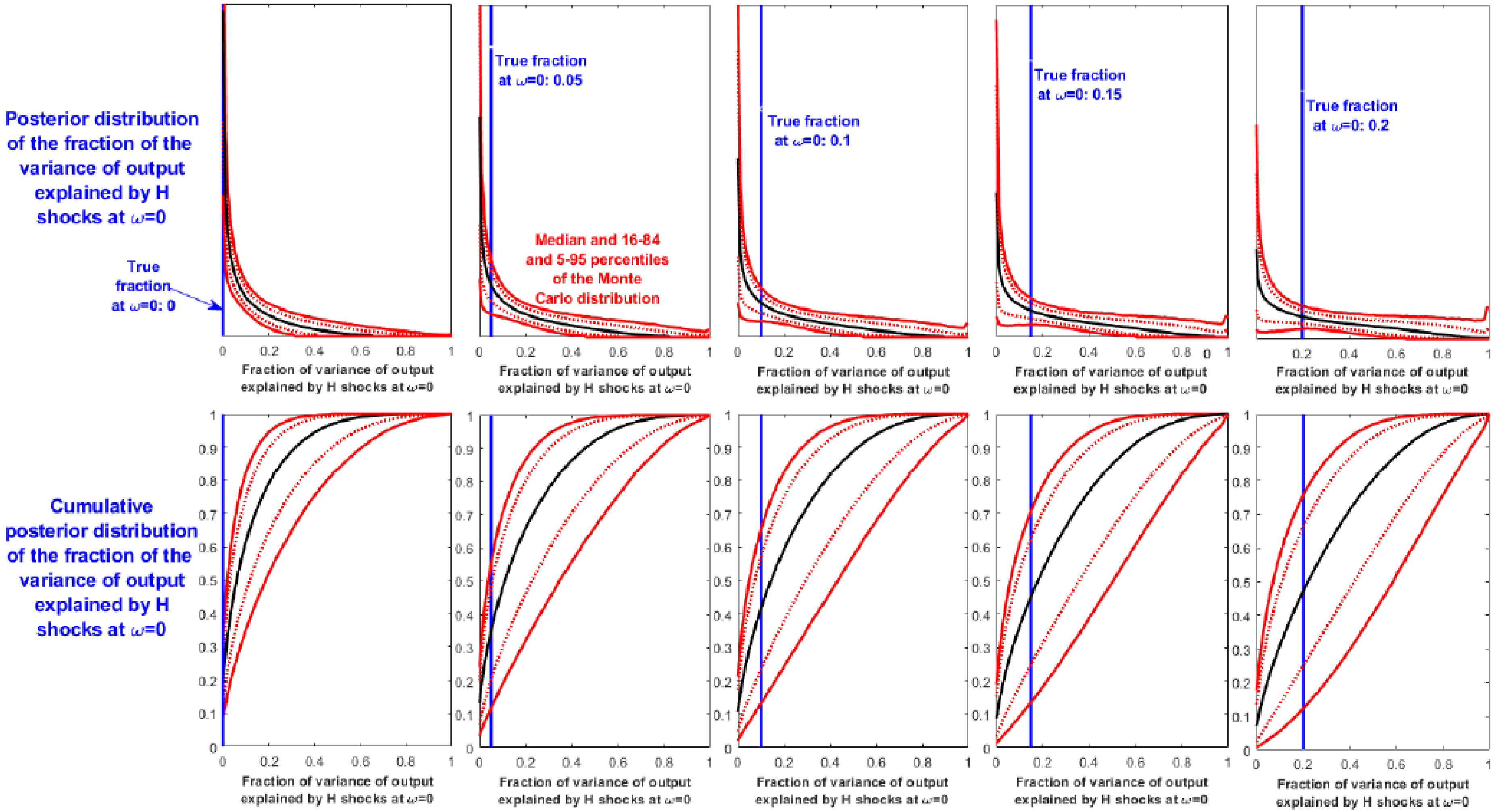

Figure 6 True fractions of the variance of output explained by ϵ_t^e (H) shocks at ω=0 in the RBC model, and medians and 16-84 and 5-95 percentiles of the Monte Carlo distributions of the estimated fractions, based on VARs in differences

explained by ϵ_t^e in the long run. What emerges from the figure is that our baseline identification approach has low power to discriminate between alternative extents of hysteresis. This is especially apparent from the evidence in the top row, which shows the posterior distributions of the fraction of the output variance explained by H shocks. These are consistently spread out, with materially different values being associated with quite similar probabilities. For example, when the true fraction is 10 percent, the mode of the posterior in the upper right-hand panel still has a clear peak at 0, but values in excess of 20, 30, or even 40 percent are associated with non-negligible probabilities.

In summary, the analysis in this section provides evidence that our baseline identification scheme (and likely others, such as Furlanetto et al., 2021) are prone to spuriously detect hysteresis when, in fact, there is none. Moreover, they have low power in discriminating between alternative extents of hysteresis. Given these misclassification issues, which we can think of as Type-I and Type-II errors in a frequentist context, we develop a Monte-Carlo based approach to correct for these errors in the next section.

6 A Monte Carlo-Based Approach to Identifying Hysteresis

We now propose a correction to the potential bias in assessing the degree of hysteresis in the data. As we discuss above, the bias stems from the fact that we impose a priori the presence of two permanent shocks as drivers of GDP in the long run. While this serves as an identifying assumption for hysteresis (in contrast with Furlanetto et al., 2021, who do not impose such long-run restrictions), it leaves room for the estimation algorithm to put posterior probability mass on the presence of H shocks even when there are none. We account for this bias by means of a Monte-Carlo based correction. We first give a broad outline of the idea, then discuss the simulation algorithm in more detail, and finally present empirical results.

6.1 General Idea

We focus on the fraction of the frequency-zero or long-run variance of output that is explained by H shocks as the key measure for the presence of hysteresis. We label this statistic Φ. The identifying restrictions in Table 1 result in a posterior distribution $F(\Phi)$. We denote $F^*(\Phi)$ as the posterior distribution obtained by estimating the Bayesian SVAR on the actual data, whereas $F_C^*(\Phi)$ is the corresponding cumulative distribution. The general idea is to construct counterparts to these distributions from Monte Carlo simulations of the estimated Bayesian SVAR for a range of Φ values that are imposed upon the data-generating process (DGP). Our bias-corrected estimate of hysteresis is then the value of Φ for which the simulated posterior distribution is closest to the empirical one.

We start from the cointegrated Bayesian SVAR with identifying restrictions detailed above. Inference based on the actual data results in a posterior distribution over the model coefficients. In order to set the benchmark DGP on which the Monte-Carlo analysis is based, we select the draw from the posterior distribution, for which the SVAR coefficients are closest to their median posterior values. We then keep these coefficients fixed for the remainder of the Monte-Carlo analysis. We consider a range of K values for Φ from 0 (no hysteresis) to $\Phi_{Max} = 0.5$, that is, when H shocks explain half of the long-run output variance. Individual values of Φ are labelled Φ_k, with $k = 1, 2, ..., K$. We impose different values of Φ by appropriately scaling the column of the impact matrix A_0 associated with H shocks.

We simulate the DGP for each Φ in the range, denoting each artificial sample s, with $s = 1, 2, ..., S$, whereby we maintain the sample length, specification and identification assumptions as in our empirical model. For each value of Φ_k, the cointegrated Bayesian VAR is estimated for each artificial sample s. This produces a posterior distribution $F^s(\Phi_k)$ and a corresponding cumulative posterior distribution $F^s_C(\Phi_k)$. Repeating this process for all s and for each Φ_k, builds up the Monte-Carlo distribution of the $F^s_C(\Phi_k)$. The evidence presented in Figure 6 suggests that the larger Φ_k, the more the cumulative distributions of the $F^s_C(\Phi_k)$ shift to the right.

In the final step, we assess the statistical distance between the empirical distribution $F^*_C(\Phi)$ and the Monte-Carlo distribution, $F^s_C(\Phi_k)$, for each Φ_k and each s. Among the K values of Φ_k we then choose as our bias-corrected estimate of Φ the one for which the induced $F^s_C(\Phi_k)$ is closest to $F^*_C(\Phi)$. As a metric for the statistical distance between these distributions, we use the Kolmogorov-Smirnov (henceforth, KS) statistic. For each Φ_k we thus build up the Monte Carlo distribution of the KS statistic based on the S artificial samples. This allows us to assess the true extent of hysteresis in the data while correcting for the previously discussed misclassification errors. The value of Φ that minimizes the median of the Monte-Carlo distribution of the KS statistic is our bias-corrected estimate. We now discuss our approach in more detail.

6.2 The Monte-Carlo Approach to Hysteresis

The key idea behind our proposed approach is that we simulate estimated Bayesian VARs, on which we impose specific degrees of hysteresis. The simulation results allow us to assess what the full posterior shape of hysteresis would have been had it been generated by a specific DGP that we control. We then compare the outcomes from the Monte Carlo simulations to those observed from the actual data. Although we use Φ as the relevant statistic, our approach is generally applicable to other statistics derived from a posterior distribution.

Recall the cointegrated structural VAR representation of the model in equation (1):

$$\Delta Y_t = B_0 + B_1 \Delta Y_{t-1} + ... + B_p \Delta Y_{t-p} + \alpha\beta' Y_{t-1} + A_0 \epsilon_t. \tag{8}$$

A_0 is the impact matrix of the structural shocks ϵ_t, with $\epsilon_t = A_0^{-1} u_t$, and the u_t are the reduced-form shocks from equation (1). The matrix of the long-run impacts of the structural shocks is given by:

$$L = C \times A_0 = \beta_{OC}[\alpha'_{OC}(I_N - B)^{-1}\beta_{OC}]\alpha'_{OC} \times A_0, \tag{9}$$

where $B = B_1 + B_2 + ... + B_p$, and α_{OC} and β_{OC} are the orthogonal complements of α and β. We now describe how we select a DGP for the Monte Carlo analysis from the posterior distribution of the estimated Bayesian SVAR and how we impose specific values of Φ on the DGP.

Our objective is to select a DGP that is closest to the median outcome of the posterior. However, we cannot simply take the median values of posterior coefficients since they would likely stem from different 'models', that is, the resulting median draws would be associated with different parameterizations. Instead, we adopt the following approach. Let D be the number of draws from the posterior distribution. We select the specification among the D draws that is closest to the 'median model'.[9]

Specifically, let $\bar{A}_0$, $\bar{B}$, and $\bar{L}$ be the medians of the posterior distributions of A_0, B, and L. Among all D draws from the posterior we select the draw associated with $\tilde{A}_0$, $\tilde{B}_0, ..., \tilde{B}_p$, and $\widetilde{L}$ for the DGP by minimizing the following criterion:

$$\Gamma_d = \sum_{i=1}^{N}\sum_{j=1}^{N}\left[(A_{0,d}^{ij} - \bar{A}_0^{ij})^2 + (L_d^{ij} - \bar{L}^{ij})^2\right] + \sum_{i=1}^{N}\sum_{j=1}^{Np}(B^{ij} - \bar{B}^{ij})^2, \text{ for } d = 1, 2, ..., D. \tag{10}$$

This minimizes the sum of the squared deviations of the individual elements of $A_0, B_0, ..., B_p$, and L from the corresponding elements of $\bar{A}_0$, $\bar{B}$, and $\bar{L}$. Recall that N is the number of variables in the VAR and p is the lag order. We fix the chosen parameters, which thereby

[9] We performed the same steps for the 'modal model'. The results are virtually identical as far as the Monte-Carlo simulations are concerned.

establishes our DGP. The simulations are then accomplished by drawing from the *i.i.d.* Gaussian process ϵ_t. We thus obtain the DGP:

$$\Delta Y_t = \tilde{B}_0 + \tilde{B}_1 \Delta Y_{t-1} + ... + \tilde{B}_p \Delta Y_{t-p} + \tilde{\alpha}\tilde{\beta}' Y_{t-1} + \tilde{A}_0 \epsilon_t, \tag{11}$$

for 'optimal' parameters $\tilde{A}_0$, $\tilde{B}_0, ..., \tilde{B}_p$, and $\widetilde{L}$.

We implement varying degrees of hysteresis, as measured by the contribution of H shocks in explaining the FEVD of output, by imposing the values of $\Phi_k \geq 0$ over the chosen range. We do so by appropriately rescaling the column of $\tilde{A}_0$ associated with H shocks.[10]. We then use the resulting specifications as DGP in the Monte Carlo exercise. We set the total number of simulations for each Φ_k to $S = 1000$. For each simulation, we estimate the Bayesian cointegrated VAR on the simulated data and impose the identifying restrictions in Table 1. For each s we obtain the cumulative posterior distribution of Φ, $F^s_C(\Phi)$.

The SVAR estimated on actual data produces a posterior distribution $F^*_C(\Phi)$, from which we report the posterior median as our measure of hysteresis. As we discuss above, there is an inherent bias in this metric as it likely puts posterior probability mass on the presence of H shocks even when there is no underlying hysteresis. We correct for this misattribution by comparing the entire empirical posterior distribution to those produced by the Monte-Carlo simulations for different values of Φ.

We compare the two cumulative posterior distributions using the Kolmogorov-Smirnov (KS) statistic, which is defined as the maximum absolute value of the difference between them:

$$KS = \sup_{\Phi} \left| F^*_C(\Phi) - F^s_C(\Phi) \right|. \tag{12}$$

This provides a measure of the statistical distance between (i) the cumulative distribution of the fraction of the long-run variance of GDP explained by H shocks estimated on actual data, $F^*_C(\Phi)$, and (ii) the corresponding Monte-Carlo cumulative distribution computed for sample s conditional on the imposed value Φ_k. Our 'corrected' estimate of hysteresis is the value of Φ that minimizes the median of the Monte-Carlo distribution of the KS statistic.

We compute a univariate KS statistic for GDP, consumption, and investment. We then extend the univariate KS statistic in equation (12) to the bivariate case (GDP, consumption) as follows. Let $F^*_C(\Phi_y, \Phi_c)$ and $F^s_C(\Phi_y, \Phi_c)$ be the joint cumulative posterior distributions of the fractions of long-run variance of GDP (Φ_y) and consumption (Φ_c) explained by H

[10] As a side note, if we imposed $\Phi_1 = 0$ the model would become stochastically singular and therefore impossible to estimate the reduced-form model (1). We therefore impose $\Phi_1 = 10^{-6}$ instead.

shocks. The joint KS statistic for Φ_y and Φ_c is:

$$KS_J = \sup_{\Phi_y,\Phi_c} \left| F_C^*(\Phi_y, \Phi_c) - F_C^s(\Phi_y, \Phi_c) \right|, \tag{13}$$

namely the maximum absolute value of the difference between the two cumulative distributions. We note that both $F_C^*(\Phi_y, \Phi_c)$ and $F_C^s(\Phi_y, \Phi_c)$ are two-dimensional surfaces, rather than one-dimensional lines.

6.3 Results

We perform the Monte Carlo exercise for the full sample and for the sample excluding the ZLB period. Figure 7 shows the median and the 16-84 and 5-95 percentiles of the Monte Carlo distributions of the KS statistics for the two samples. We report both the joint statistic for GDP and consumption and individual statistics for GDP, consumption, and investment. As in the previous exercises, there is a fairly high degree of uncertainty, which is intrinsic to the analysis we are performing.

Focusing on the median estimates, two key observations emerge from the figure. First, the evidence suggests that the most likely value of the extent of hysteresis is zero in the sample excluding the ZLB period. Second, and in contrast, the evidence for the full sample period is less clear-cut. It suggests a non-zero, but modest extent of hysteresis. This is the case for both the joint KS_J statistic and the individual KS statistics for consumption. In particular, the minimum of the KS_J statistic corresponds to a value of hysteresis of about 5 percent of the long-run variance of GDP, while the individual statistic for consumption suggests around 10 percent. Evidence from the individual statistic for GDP is broadly consistent. The statistic for investment, on the other hand, is essentially flat until about 7-8 percent and then starts to increase rapidly, which suggests that larger values are less plausible.[11]

The Monte-Carlo results confirm our intuition that imposing two permanent drivers on GDP a priori leads the estimates to detect hysteresis even when there is none. From a Bayesian perspective this is simply a matter of having posterior probability mass on such an outcome since inference is not robust enough to cleanly distinguish between either one or two sources of long-run movements. From a frequentist perspective, the statistical model is misspecified since under the null hypothesis of no hysteresis the second permanent shock is not identifiable. Our Monte-Carlo approach is one way of correcting for this since it looks

[11] The comparison between the shapes of the median estimates for GDP and consumption points towards a natural interpretation. Since consumption is very close to the permanent component of GDP (Cochrane, 1994), it is not necessarily surprising that evidence for consumption is sharper than for GDP.

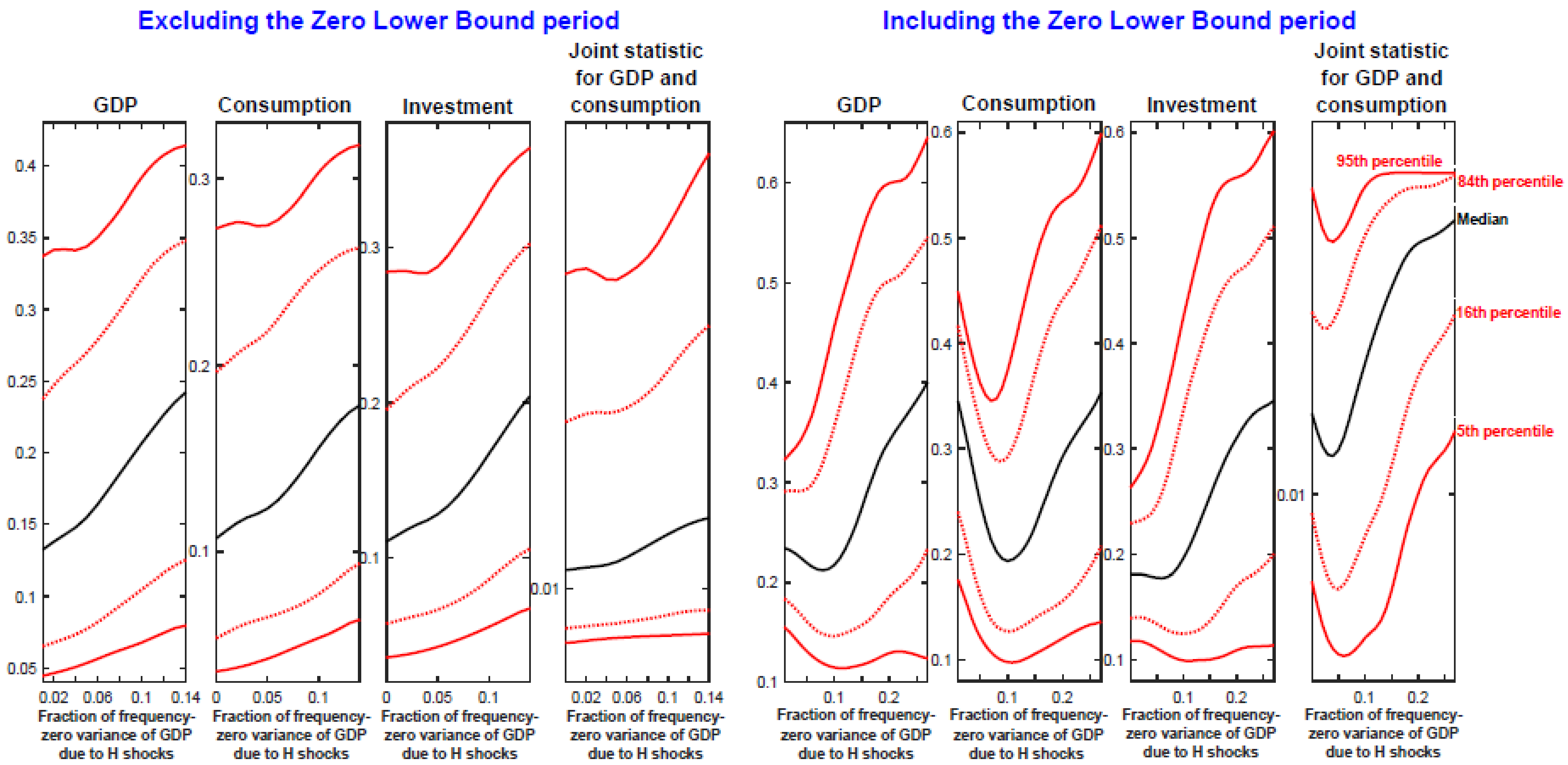

Figure 7 Median and selected percentiles of the Monte Carlo distribution of the Kolgomorov-Smirnov statistic

at the set of potential outcomes over the entire posterior distribution. We thus conclude that evidence of hysteresis is scant in U.S. data and is only detectable when a long and deep recession is included in the sample, such as the Great Recession.

7 Conclusion

We have embarked on a search for hysteresis, looking for aggregate demand shocks that have a permanent impact on real GDP and other macroeconomic aggregates. For this purpose, we develop a cointegrated structural VAR framework, where we identify shocks via a combination of zero long-run restrictions and both short- and long-run sign restrictions. We regard this, to the best of our knowledge, as a key methodological contribution to the Bayesian VAR literature. These restrictions are consistent with a wide range of theoretical models and are broad enough to leave room for identifying hysteresis shocks if they exist.

Our evidence suggests that in the post-WWII U.S. hysteresis effects have been virtually absent for samples excluding the financial crisis and the Great Recession. They only appear when we extend the sample to include the period following the collapse of Lehman Brothers. Even then, however, these effects only account for at most 10 percent of the long-run variance of GDP. While not as large as other estimates in the literature such as Jordà et al. (2020), our estimate nevertheless points towards the potential presence of a mechanism that converts economic demand disturbances into permanent outcomes for output. The difference between the results for the various sample periods also suggest that hysteresis shocks are operational in times of extreme economic distress, such as the Great Recession, which came with a dramatic increase in the long-term unemployment as one possible hysteresis channel.

Conceptually, we propose a simple approach to test for the presence and extent of hysteresis. An important role in obtaining our results is played by a Monte Carlo-based correction to the 'simple' Bayesian estimates, designed to control for the fact that, as we show via DSGE models, there is a high probability of detecting hysteresis effects even when the DGP features none by construction. Our correction is based on simulating specific statistics (e.g., the fraction of long-run variance of GDP due to hysteresis shocks) conditional on alternative extents of hysteresis imposed on the estimated VAR. The hysteresis statistics from the resulting Monte Carlo distributions are then compared to the corresponding distributions computed from the actual data via the Kolgomorov-Smirnov statistic. This approach corrects for a misidentification of hysteresis shocks when there is none in the underlying data. Methodologically, our approach can be applied to other scenarios where there is uncertainty about the presence of various permanent components.

References

[1] Aikman, David, Mathias Drehmann, Mikael Juselius, and Xiaochuan Xing (2022): "The Scarring Effects of Deep Contractions". Working Paper.

[2] Anderson, Theodore W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions". *Annals of Mathematical Statistics*, 22, pp. 327-351.

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

[4] Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (2018): "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications". *Econometrica*, 86, pp. 685-720.

[5] Ball, Laurence M. (2009): "Hysteresis in Unemployment: Old and New Evidence". In: Jeff Fuhrer et al. (eds.): Understanding In‡ation and the Implications for Monetary Policy. MIT Press, Cambridge, MA.

[6] Bauwens, Luc, and Michel Lubrano (1996): "Identification Restrictions and Posterior Densities in Cointegrated Gaussian VAR Systems". Advances in Econometrics 11, Part B (JAI Press, Greenwich), pp. 3-28.

[7] Blanchard, Olivier J. (2018): "Should We Reject the Natural Rate Hypothesis?" *Journal of Economic Perspectives*, 32(1), pp. 97-120.

[8] Blanchard, Olivier J., and Lawrence Summers (1986): "Hysteresis and the European Unemployment Problem". *NBER Macroeconomics Annual*, 1, pp. 15-7.

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

[10] Canova, Fabio, and Matthias Paustian (2011): "Measurement with Some Theory: Using Sign Restrictions to Evaluate Business Cycle Models". *Journal of Monetary Economics*, 58, pp. 345-361.

[11] Cerra, Valerie, Antonio Fatás, and Sweta Chaman Saxena (2022): "Hysteresis and the Business Cycle". Forthcoming, *Journal of Economic Literature*.

[12] Cerra, Valerie, and Sweta Chaman Saxena (2008): "Growth Dynamics: The Myth of Economic Recovery". *American Economic Review*, 98(1), pp. 439-57.

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

[14] Elliot, Graham, Thomas J. Rothenberg, and James. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root". *Econometrica*, 64(4), pp. 813-836.

[15] Forni, Mario, and Luca Gambetti (2014): "Testing for Sufficient Information in Structural VARs". *Journal of Monetary Economics*, 66, pp. 124-136.

[16] Furlanetto, Francesco, Antoine Lepetit, Ørjan Robstad, Juan Rubio-Ramirez, and Pal Ulvedal (2021): "Estimating Hysteresis Effects". Norges Bank Working Paper 2020-13.

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

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

[19] Geweke, John (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, pp. 169-193.

[20] Geweke, John (1996): "Bayesian Reduced Rank Regression in Econometrics". *Journal of Econometrics*, 75, pp. 121-146.

[21] Hamilton, James D. (1994): *Time Series Analysis.* Princeton, NJ, Princeton University Press.

[22] Jaimovich, Nir, and Henry Siu (2020): "Job Polarization and Jobless Recoveries". *Review of Economics and Statistics*, 102(1), pp. 129-147.

[23] Johansen, Søren (1988): "Statistical Analysis of Cointegration Vectors". *Journal of Economic Dynamics and Control*, 12, pp. 231-254.

[24] Johansen, Søren (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models". *Econometrica*, 69, pp. 111-132.

[25] Jordà, Òscar, Sanjay R. Singh, and Alan M. Taylor (2020): "The Long-Run Effects of Monetary Policy". *NBER Working Paper* No. 26666.

[26] King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson (1991): "Stochastic Trends and Economic Fluctuation". *American Economic Review*, 81(4), pp. 819-840.

[27] Kleibergen, Frank. and Herman K. van Dijk (1994): "On the Shape of the Likelihood/Posterior in Cointegration Models". *Econometric Theory*, 10, pp. 514-551.

[28] Kwiatkowski, Denis, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root". *Journal of Econometrics*, 54, pp. 159-178.

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

[30] Ljungqvist, Lars, and Thomas J. Sargent (1998): "The European Unemployment Dilemma". *Journal of Political Economy*, 106(3), pp. 514-550.

[31] Sim, Jae (2022): "Demand Shocks, Hysteresis and Monetary Policy". Finance and Economics Discussion Series 2022-080. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2022.080.

[32] Stock, James H., and Mark W. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations". *Journal of Business and Economic Statistics*, 14(1), pp. 11-30.

[33] Stock, James H., and Mark W. Watson (1998): "Asymptotically Median Unbiased Estimation of Coefficient Variance in a Time Varying Parameter Model". *Journal of the American Statistical Association*, 93(441), pp. 349-358.

[34] Strachan, Rodney W. and Brett Inder (2004): "Bayesian Analysis of the Error Correction Model". *Journal of Econometrics*, 123, pp. 307-325.

[35] Uhlig, Harald (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure". *Journal of Monetary Economics*, 52, pp. 381-419.

[36] Watson, Mark W. (1986): "Univariate Detrending with Stochastic Trends". *Journal of Monetary Economics*, 18, pp. 49-75.

[37] Wu, Jing Cynthia, and Fan Dora 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.

Appendix

A Computational Details for the Bayesian Cointegrated VAR

A.1 Choice of 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:

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

where MN denotes the matricvariate-normal distribution; $P_{1/\tau} = HH' + \tau^{-1}H_\perp H_\perp'$; and G is a square matrix that we set to $G = \Sigma$ following Strachan and Inder (2004). In addition, τ is a scalar between 0 and 1, and ν is a scalar that controls the extent of shrinkage.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:

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

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:

$$p(\Sigma) \propto |\Sigma|^{-(N+1)/2}. \tag{C.3}$$

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.[12]

A.2 Drawing from the Posterior

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

$$\Sigma \mid \alpha, \beta, B_0, B_1, ..., B_p, Y_t \sim IW\left(u_t'u_t + \nu\alpha(\beta' P_{1/\tau}\beta)^{-1}\alpha', T + r\right), \tag{C.4}$$

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

$$vec(A) \mid \alpha, \beta, \Sigma, \tau, \nu, Y_t \sim N\left\{(\Sigma^{-1} \otimes Z'Z + \lambda^2 I)^{-1}(\Sigma^{-1} \otimes Z'Z)vec(A'), (\Sigma^{-1} \otimes Z'Z + \lambda^2 I)^{-1}\right\}, \tag{C.5}$$

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

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 $T \times 1 + Np$ 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:

$$b \mid a, B_0, B_1, ..., B_p, \Sigma, \tau, \nu \sim MN\left(0, (a'G^{-1}a)^{-1} \otimes \nu P_\tau\right), \tag{C.6}$$

$$p(a) \propto |G|^{-r/2}|a'G^{-1}a|^{-N/2}. \tag{C.7}$$

The conditional draws for α and β, that is, steps (i) and (ii) of the Gibbs-sampling algorithm discussed in the main text of the paper, 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 $\alpha = a^{(*)}(b'b)^{1/2}$.

Figure A.15 of the Online Appendix report evidence on the convergence of the Markov chain, by showing Geweke's (1992) inefficiency factors (IFs) of the draws for each individual parameter. The inefficiency factors are defined as the inverse of the relative numerical efficiency measure of Geweke (1992),

$$RNE = (2\pi)^{-1}\frac{1}{S(0)}\int_{-\pi}^{\pi} S(\omega)d\omega, \tag{C.8}$$

where $S(\omega)$ is the spectral density of the sequence of draws from the Gibbs sampler for the quantity of interest at frequency ω. We estimate the spectral densities via the lag-window estimator as described in chapter 10 of Hamilton (1994). We also considered an estimator based on the fast-Fourier transform, and results were very close. For all parameters the IFs are equal to at most 8, that is, well below the values of 20-25 which are typically taken to indicate problems in the convergence of the Markov chain.