Working Paper Series

Searching for Hysteresis

WP 22-05

Luca Benati
University of Bern

Thomas A. Lubik
Federal Reserve Bank of Richmond

**FEDERAL RESERVE BANK
OF RICHMOND**®

Richmond • Baltimore • Charlotte

Searching for Hysteresis[*]

Luca Benati
University of Bern[†]

Thomas A. Lubik
Federal Reserve Bank of Richmond[‡]

April 1, 2022

Abstract

We search for hysteresis in aggregate post-WWII U.S. data. Defining hysteresis as the presence of aggregate demand shocks with a permanent impact on real GDP, we identify such shocks based on Bayesian VARs via 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, and they only appear when including the period following the collapse of Lehman Brothers. Even then, these effects only account for at most 10 per cent of the long-run variance of GDP. We show 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 issue with a Monte Carlo-based approach.

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

---

1 Introduction

The term hysteresis captures the notion that economic disturbances that are typically regarded as transitory, for instance, monetary policy shocks, can have very long-lived or even permanent effects. Hysteresis is most often associated with 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. 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 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 the theoretical channels for the transmission of hysteresis shocks are well-understood, empirical evidence is scarce.

In this paper we search for hysteresis, which we define as the presence of aggregate demand shocks with a permanent effect on output. 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 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. We then apply our identification scheme to Bayesian VARs estimated based on post-WWII U.S. data.

Our paper makes three main contributions. First, we detect a non-negligible, but modest extent of hysteresis, which explains roughly 20 percent of the frequency zero (i.e., long-run) variance of real GDP. Estimates, however, are characterized by a very large extent of uncertainty.

Second, 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 third contribution lies in developing an approach to test for the presence of hysteresis and to estimate its extent. Our approach uses Monte-Carlo analysis to simulate specific statistics, such as the fraction of the long-run variance of output due to hysteresis shocks, conditional on alternative values of hysteresis that are imposed upon the estimated

Bayesian VAR. We then estimate 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. Specifically, we use as a metric capturing the closeness of the simulated and actual distributions the Kolmogorov-Smirnov statistic. This also allows us to quantify the implied 'correction' to the simple estimates produced by the Bayesian VAR due to the mis-identification problem. When applying this approach to our dataset we estimate an even more modest extent of hysteresis, equal to at most 10 percent of the long-run variance of GDP. Further, when excluding from the sample the period following the collapse of Lehman Brothers evidence is compatible with the notion of no hysteresis.

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) argued that the evidence is not sufficiently clear-cut 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/price inflation experience.

Furlanetto et al. (2021) is the paper that is closest to our work. They use a structural VAR framework that combines long-run zero restrictions and short-run sign restrictions. 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. 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 robustly identifying 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 that allows for hysteresis. Section 6 discusses our proposed Monte Carlo-based solution to this problem, and presents and discusses the evidence obtained by applying the proposed approach to the data. 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

In both theoretical and empirical macroeconomics it is often assumed that permanent increases in GDP originate on the supply side of the economy, for instance in the form of permanent shocks to total factor productivity (TFP). The key idea underlying the notion of hysteresis is that some permanent output shocks 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 ultimately transitory output fluctuations. We now discuss the identifying restrictions that allow us to accomplish this.

We assume that the permanent component of real GDP is driven by 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. They are also consistent with a wide range of theoretical frameworks, such as the medium- or large-scale DSGE models that are used in policy institutions. Specifically, we make the following assumptions:
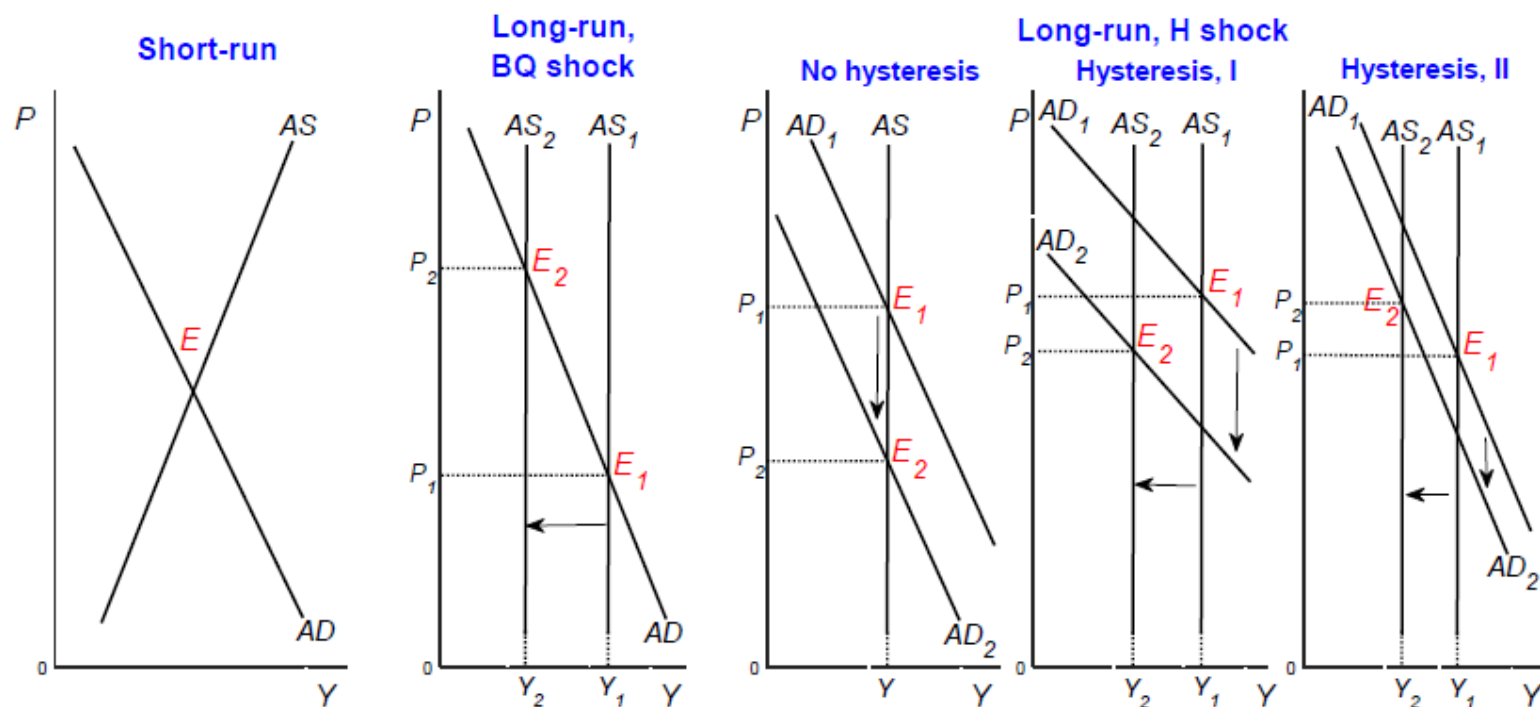
4

Figure 1 A simple illustration of the identifying 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, with a negative (positive) H shock having 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 macro-economic frameworks, such as the textbook AD-AS model, simple New Keynesian models, and 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 identifying assumptions in the output (Y)-price level (P) space.

In the first panel, in the short run BQ shocks shift the AS curve, thus moving the economy along the AD curve. This causes output and prices to move in opposite directions. By the same token, H shocks move the economy along the AS curve, so that output and prices move in the same direction. Consequently, the first set of restrictions we impose in order to disentangle the two shocks is that both on impact, and over the subsequent 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*, whereas 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 (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,

---

[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 impacts on output and inflation have the same sign.

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, as 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 shown in the fourth and fifth panels. In the limit case of no output hysteresis 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.

Finally, in order to achieve a stronger identification of the two shocks we impose the restriction that, for either of them, the sign of its long-run impact on GDP ought to be the same as the signs of its long-run impacts on consumption and investment.

We summarize the previously discussed restrictions in Table 1. We now turn to discussing 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.

Table 1: Identifying Restrictions				
	Shock:			
Impact on:	BQ	H	BQ	H
	Short run:		Long run:	
Output	+	+	+	0 +
Price level	−	+	−	+ ?/+

3 Empirical Approach

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 rate. For the short-term rate we use the 'shadow rate' from Wu and Xia (2016), in order to capture the effective policy stance during the zero-lower bound (ZLB) 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 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). On the other hand, for inflation, i.e. $\pi_t = p_t - p_{t-1}$ the null of a unit root is rejected.

We proceed with the empirical analysis as follows. Based on the evidence from the unit root tests we take as our baseline specification a Bayesian cointegrated VAR for Y_t, as previously defined. Since the notion that interest rates and hours per capita are I(1) may be questioned on conceptual grounds, in Online Appendix D we also consider a second specification in terms of stationary variables. In brief, evidence from this alternative specification is qualitatively similar to that produced by the baseline model.

Our approach to Bayesian cointegration is based on methods introduced by Strachan and Inder (2004) and Koop et al. (2010). Let the cointegrated vector error correction (VECM) representation of the VAR be

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

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, 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, i.e.

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

where I_r is the $r \times r$ identity matrix. This restriction is imposed by defining a semi-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 the short- and the long-term nominal interest rate. Similarly, the same tests identify three cointegration vectors for the baseline 7-variables system for $Y_t = [y_t, c_t, i_t, R_t, r_t, p_t, h_t]'$ for the period 1954Q3-2019Q4, and two for the period 1948Q1-2008Q4. For the latter period, however, lack of identification of a third cointegration vector is marginal, with p-values for the trace and maximum eigenvalue tests both equal to 0.11. In what follows 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. 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. Our numerical choices for H_g reflect 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 percent of the increase for the latter. We want to stress, however, that given the uninformativeness of the prior for β (discussed below, and in Appendix A), the centering encoded in this choice for H_g imposes, in fact, extremely weak restrictions.

We estimate the reduced-form 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 (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

8

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.

Appendix A 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 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. The Online Appendix also reports evidence on the convergence of the Markov chain, by showing Geweke's (1992) inefficiency factors (IFs) of the draws for each individual parameter. For all parameters the IFs 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 (see Table 1). We jointly impose all of the restrictions based on the methodology proposed by Arias et al. (2018). Details on the procedure can be found in 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) to BQ and H shocks produced by the baseline specification, whereas Figure 3 reports the fractions of forecast error variance (FEV) of the variables explained by the two shocks. 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. Table 2 reports the medians and selected percentiles of
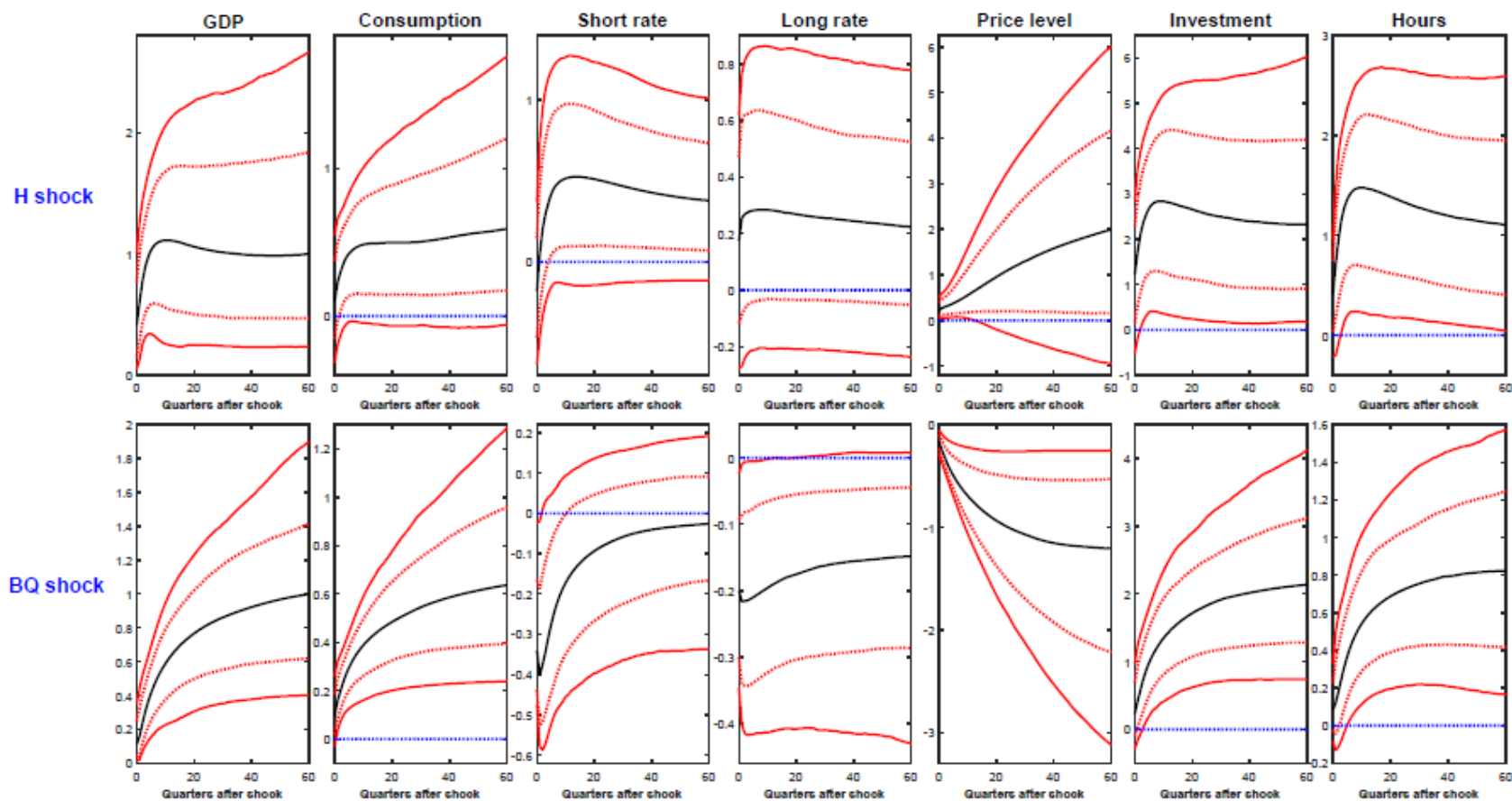
Figure 2 Impulse-response functions to H and BQ shocks, based on the 7-variables
cointegrated VAR (including the Zero Lower Bound)

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.

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 near-uniformly as expected, and in fact for GDP, consumption, investment and the price level they are the product of the restrictions we imposed. Specifically, a BQ shock permanently increases GDP, consumption, and investment and permanently 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 are 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.[2]

Turning to the price level, the IRFs for the two samples obtained without imposing restrictions on the long-run impact of H shocks on prices reveal an interesting finding. For the 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. Further, at the frequency zero 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.

For the sample excluding the ZLB (Figure A.4), at the 15 years horizon exactly 50 percent of the draws are associated with a negative impact of H shocks on prices, whereas at the frequency zero the corresponding fraction is equal to 65.6 percent. Taking these results at face value, they imply that in this sample the possibility illustrated in the panel labelled as 'Hysteresis II' in Figure 1 is significantly more likely than in the full sample. A second possible interpretation—which, to anticipate, is in fact more compatible with our whole body of evidence—is instead that in this sample H shocks are (virtually) absent, and that this result is simply the figment of the fact that our identifying restrictions impose upon the data their very existence by 'brute force'. The intuition is straightforward. Suppose that the DGP only features BQ shocks. Since for the price level we only impose *short-run*

---

[2]The IRFs produced by stationary SVARs (see Online Appendix D) are closely in line with those produced by the cointegrated SVARs. The main difference is that, by construction, the response of hours mean-reverts to zero.
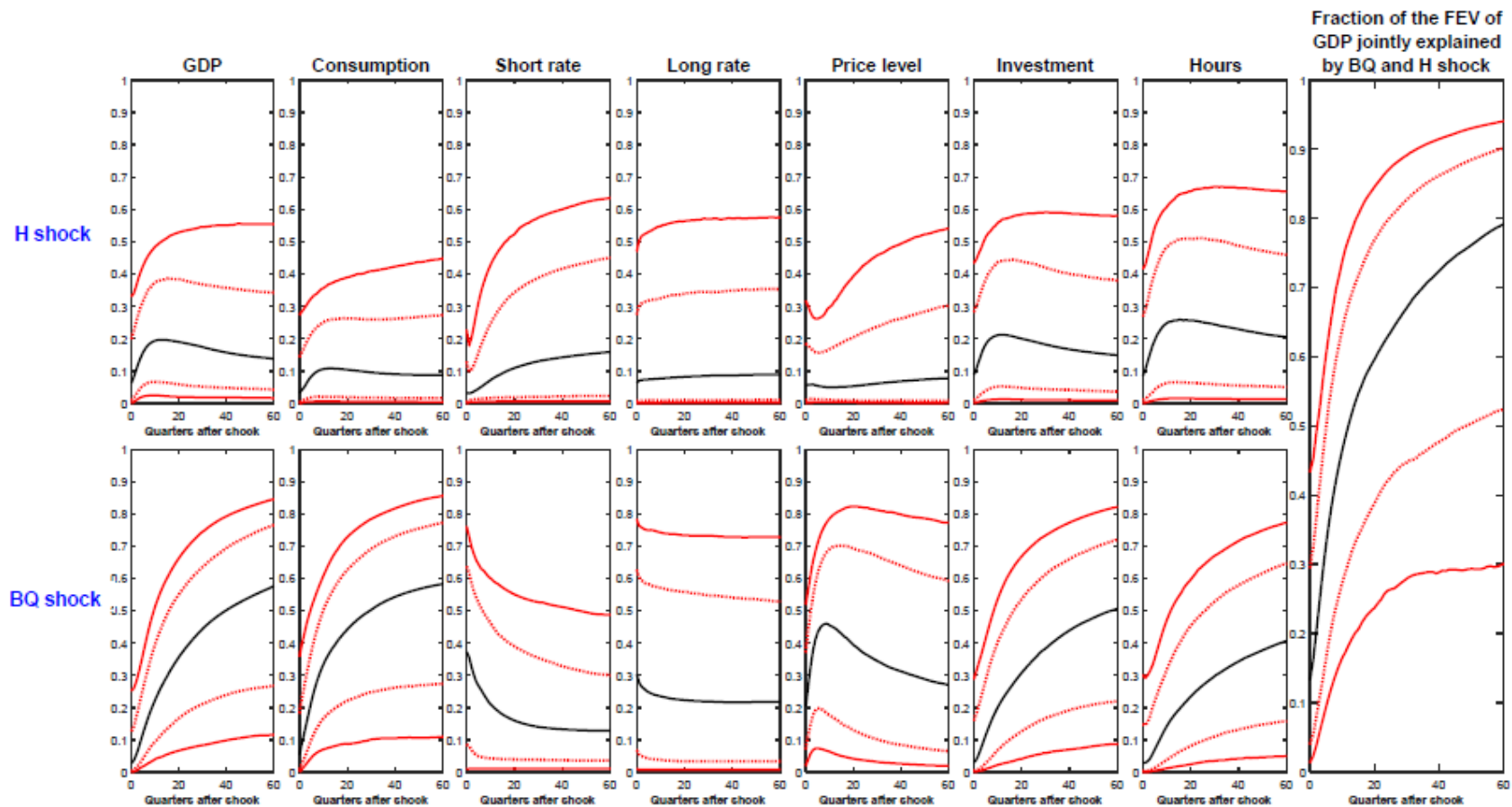
Figure 3 Fractions of forecast error variance explained by H and BQ shocks, based
on the 7-variables cointegrated VAR (including the Zero Lower Bound)

restrictions on its 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. However, since we do not impose any restriction on the long-run impact of H shocks on prices, and by assumption in the DGP all permanent GDP shocks cause a permanent decrease in the price level, in the long-run the IRF of prices to H shocks will tend to be negative. This insight suggests that there is the possibility of spuriously identifying hysteresis when in fact the DGP features none. We will investigate this issue in section 5.

4.2 Forecast Error Variance

Figure 3 shows the fractions of FEV explained by H and BQ shocks for the baseline specification. The two shocks jointly explain very large fractions of the FEV of GDP at long horizons, i.e. about 80 percent. A similar picture emerges from the other specifications reported in the Online Appendix, i.e. for the sample excluding the ZLB period without imposing restrictions on the long-run impact of H shocks on prices, and for either sample when restricting such impact. A consistent finding is also that H shocks explain between 10 and 20 percent of the FEV of 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 FEV of consumption explained by H shocks is fairly small, ranging between about 4 and 10 percent, whereas the corresponding fraction explained by BQ shocks ranges between 60 and 65 percent. This is an important finding since, as discussed by Cochrane (1994), consumption is very close to GDP's stochastic trend, and should therefore be more 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 is a first crucial piece of evidence suggesting that if hysteresis effects are present in post-WWII U.S. data, they are likely small-to-negligible. In particular, based on the full sample and without restricting the long-run impact of H shocks on prices, H shocks explain 13.8 percent of the overall fraction of the FEV of GDP at the 15 years horizon that is jointly explained by BQ and H shocks, whereas the corresponding figure for the sample excluding the ZLB period is 5.6 percent.

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]

We now dig deeper into the role played by H shocks at very long horizons, by focusing on the frequency-zero (or long-run) variance of GDP and consumption. Table 2 reports the medians and the 16-84 and 5-95 percentiles of the posterior distributions of the fractions of the long-run variance of GDP and consumption explained by H shocks. In line with the previous discussion, the estimates for consumption are consistently smaller than those for GDP. Although the difference is small for the full sample, when excluding the ZLB period it is non-negligible: based on our preferred model, i.e. 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.

Focusing on consumption, evidence suggests that the 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 prices. When imposing such restrictions, the corresponding values are 0.218 and 0.056. This strongly suggests that for the sample excluding the ZLB period hysteresis effects are very small to negligible, and that such effects only appear when including the financial crisis and the Great Recession. In a sense this is not surprising, since this is the period that has most likely seen hysteresis effects at play, because of (e.g.) historically high long-term unemployment. For the period excluding the ZLB restricting the long-run impact of H shocks on prices typically produces smaller estimates of the fraction of the frequency-zero variance of either GDP or consumption due to H shocks. This is especially apparent based on stationary VARs (see Online Appendix D).

Overall, the evidence produced by Bayesian VARs points towards a modest, but non-negligible extent of hysteresis, which in our preferred specification, based on cointegrated VARs, is equal to about 20 percent of the frequency-zero variance of GDP for the sample including the ZLB period. For the sample excluding the ZLB period evidence is mixed, but

overall it points towards a smaller role played by H shocks in driving long-horizon variation in GDP and consumption, with evidence that this is the case being stronger for the latter variable.

An important caveat to these results is that our approach imposes upon 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 identifed is a genuine result, or it is rather the artifact of our identification strategy. In the next section we show, based on a standard model featuring the possible presence of hysteresis, that the problem is indeed potentially serious. 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 approach designed to address 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 detect hysteresis when in fact there is none in the underlying DGP. The reason for this is straightforward. Since this strategy jointly identifies both a BQ and H shock, in fact it imposes upon the data the very existence of hysteresis by assumption. Ultimately, how problematic in fact this is in an assessment of hysteresis is an empirical issue. To that end, we conduct a simulation study that explores and quantifies the potential for misidentification.

5.1 The Data-Generating Process

The theoretical model we use 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 D. In the following we therefore only provide an outline of its structure.

Households solve the optimization problem:

$$U_0^* = \underset{C_t, N_t}{Max} E_0 \sum_{t=0}^{\infty} \beta^t \left[\ln(C_t - bC_{t-1}) - \frac{N_t^{1+\phi}}{1+\phi} \right], \tag{4}$$

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 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.[3]

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 $\sigma_a = 0.007$.[4] The model

---

[3] Online Appendix D 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.

[4] 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, within the present context, a_t is the random-walk component of log real GDP.

is augmented with three transitory AR(1) disturbances: two of them (v_t and u_t) can be thought of as supply shocks, whereas 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 $\sigma_z = 0.005$, so that $y_t = a_t + (1-\alpha)n_t + z_t$, where $y_t = \ln(Y_t)$ and $n_t = \ln(N_t)$. The autoregressive parameters are calibrated as $\rho_v = \rho_u = \rho_e = 0.75$, 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 $\sigma_u = 0.001$, $\sigma_v = 0.005$ and $\sigma_e = 0.045$.

Based on this calibration the permanent technology shock (ϵ_t^a) explains one-third of the FEV of log GDP on impact. In the absence of hysteresis, when $\delta = 0$, it explains about 96 percent 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 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. 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 percent of the long-run variance of output, while at long horizons the other three shocks contribute essentially nothing to output's variation.

Figure 4 shows the IRFs 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 $\delta = 0.1256$.[5] The evidence in the figure suggests that the model is an appropriate DGP for assessing the ability of the identifying restrictions in Table 1 to effectively recover the extent of hysteresis in the data (if any). First, for either of the two values of δ (as well as 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) 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 Monte Carlo exercises below we will impose this restriction in the identification of H shocks.

---

[5] 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.
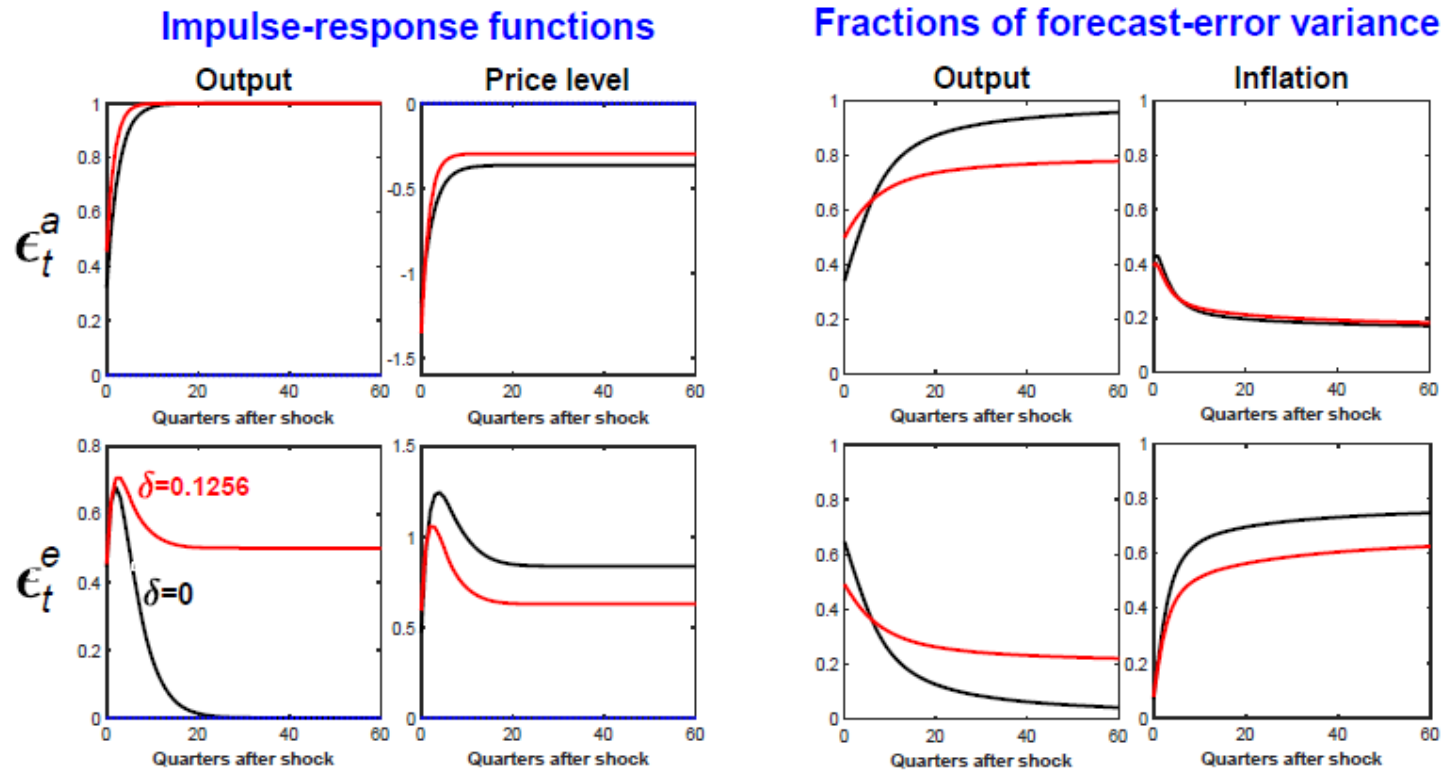
Figure 4 Theoretical impulse-response functions and fractions of forecast error variance for the RBC model for output and prices, for ϵ_t^a (BQ) and ϵ_t^e (H) shocks

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) based 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.[6] 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 (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 we perform 10,000 Monte Carlo simulations and then conduct inference by taking 10,000 draws from the posterior distribution.

5.2.1 Imposing Model-Consistent Identifying Restrictions With No Hysteresis in the DGP

As a first exercise we consider as DGP the model featuring no hysteresis (i.e., with $\delta=0$), and we impose upon the VAR the correct identifying restriction, i.e. that the DGP features a single shock (ϵ_t^a) with a permanent impact on output. Our objective is to ascertain whether under these circumstances the VAR can recover the main features of the DGP, in terms of IRFs and fractions of FEV. For 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, and we compute IRFs and fractions of FEV. We then extract and store the median and the 5th, 16th, 84th, and 95th percentiles of their posterior distributions. In this way we build up the Monte Carlo distributions of the medians and the percentiles of the IRFs and fractions of FEV. The evidence is shown in Figure A.3 in the Online Appendix. The figure shows the means, across all of the the Monte Carlo simulations, of the median IRFs and fractions of FEV, and of the 5th, 16th, 84th, and 95th

---

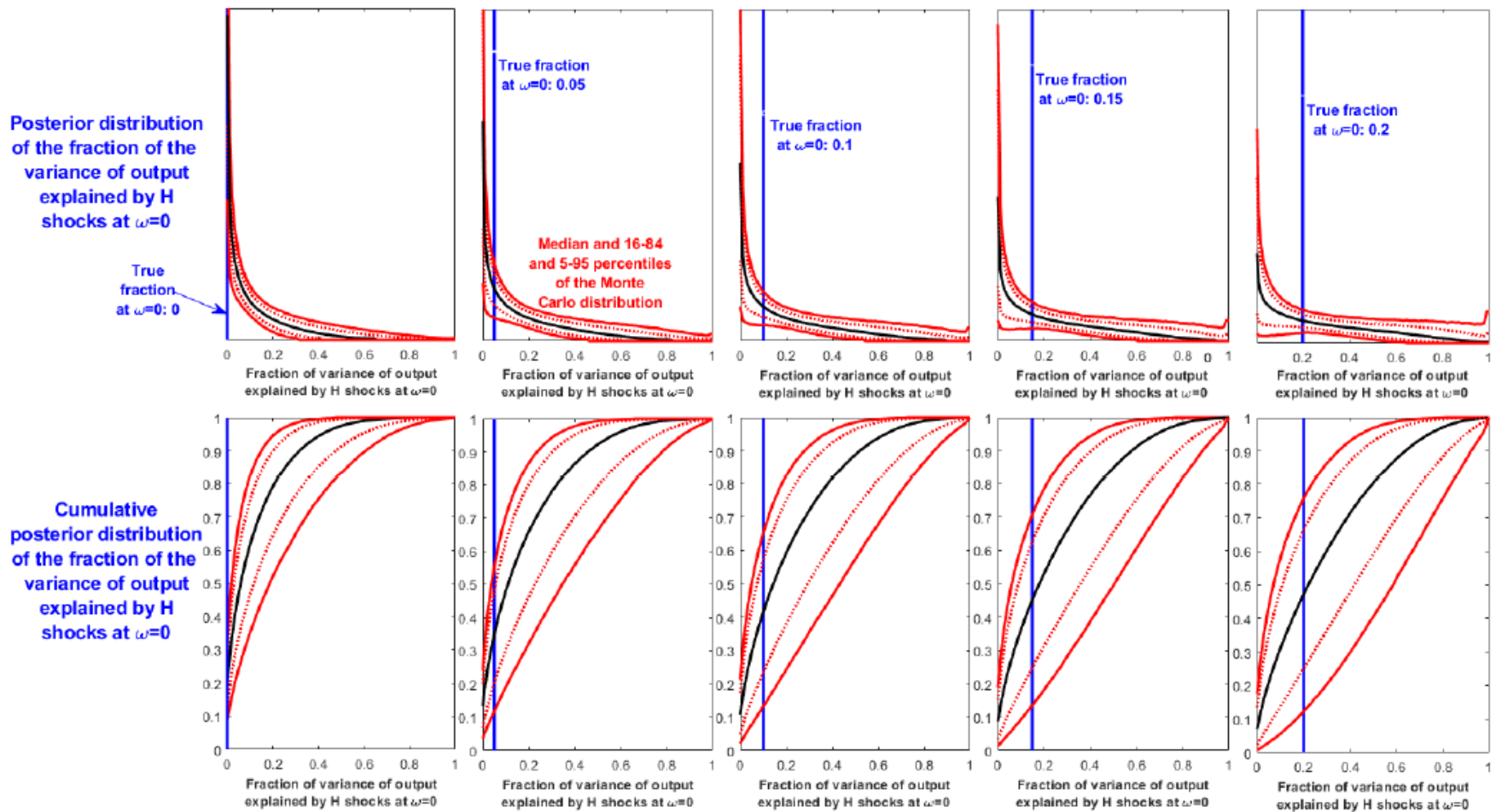[6]See the discussion in Cochrane (1994) about alternative ways of estimating cointegrated systems.

Figure 5 True fractions of the variance of output explained by ϵ_t^e (H) shocks at $\omega=0$ 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

percentiles. In brief, evidence shows that the SVAR recovers the main features of the DGP with great precision. In particular, this is the case for the main series of interest within the present context, namely output.

5.2.2 Imposing the Baseline Identifying Restrictions

In the next exercise we impose the identifying restrictions reported in Table 1 and discussed in section 2. We first take as DGP the model featuring no hysteresis, and we explore whether jointly identifying both BQ and H shocks leads to spuriously detecting hysteresis. We then consider the model featuring hysteresis, and we investigate whether the identifying restrictions allow to correctly capture the true extent of hysteresis in the DGP.

No hysteresis in the DGP Once again we jointly impose the zero long-run restrictions, and the short- and long-run sign restrictions, via the methodology proposed 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 5 shows the results. The two panels show the averages, across all of the Monte Carlo simulations, of the medians and the percentiles of the posterior distribution (top panel) and posterior cumulative distribution (bottom panel) of the statistic of interest, namely the fraction of the frequency-zero variance of output explained by the identified H shocks, together with the corresponding true fractions explained by ϵ_t^e in the theoretical model.[7] The conclusion to be drawn from this exercise appears as fairly clear-cut: the identifying restrictions in Table 1 tend to detect hysteresis when the DGP features none with non-negligible, and in fact quite sizeable probability. The likely reason for this is that, by its very logic, the identifying restrictions impose upon the data the very existence of both BQ *and* H shocks.

Hysteresis in the DGP Columns 2 to 5 of Figure 5 report the corresponding sets of results for four different values of the true fraction of the long-run variance of output explained by ϵ_t^e. What emerges from the figure is that the identifying restrictions have a low power to discriminate between alternative extents of hysteresis. This is especially apparent

---

[7]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, thus building up the Monte Carlo distribution of these objects. Figure 5 shows the means of these objects taken across the Monte Carlo distributions.

from the evidence in the top row, which shows the posterior distributions of the fraction of the long-run variance of output explained by H shocks. These distributions 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 distribution in the third panel in the top row still exhibits a clear peak at 0, but values in excess of 20, 30, or even 40 percent are associated with non-negligible probabilities.

Summing up, the analysis in this section shows that our identifying restrictions (and likely other identification schemes, such as Furlanetto et al.'s, 2021) are prone to spuriously detect hysteresis when in fact the DGP features none. Moreover, they have a low power to discriminate between alternative extents of hysteresis. Because of these mis-classification issues, which are conceptually akin to Type-I and Type-II errors within a frequentist context, in the next section we develop a Monte-Carlo based approach in order to correct the 'simple' estimates produced by the Bayesian VARs.

6 A Monte Carlo-Based Approach to Identifying Hysteresis

The identifying restrictions in Table 1 produce a posterior distribution for the fraction of the frequency-zero variance of GDP that is explained by H shocks, which we label Φ. We denote as $F^*(\Phi)$ the posterior distribution obtained by imposing the identifying restrictions upon the Bayesian VAR estimated based on the actual data. By the same token, $F_C^*(\Phi)$ is the corresponding cumulative posterior distribution. In what follows we describe a Monte Carlo-based approach designed to correct for the mis-classification problems documented in the previous section. We start by providing a broad outline of the general idea. We then discuss the simulation algorithm in more detail, and we present the evidence obtained by applying the proposed approach to the actual data.

6.1 General Idea

The idea behind the proposed approach is to simulate estimated Bayesian cointegrated SVARs upon which we have imposed specific extents of hysteresis, and to then estimate, based on such artificial data, the same SVAR we have estimated based on the actual data. We take Φ as the relevant statistic. We consider a grid of K values for Φ from 0 (no hysteresis) to $\Phi_{Max} = 0.5$, for which H shocks explain half of the frequency-zero variance of GDP. The individual values of Φ in the grid are labelled as Φ_k, with $k = 1, 2, ..., K$. For each simulated, artificial sample s, with $s = 1, 2, ..., S$, we estimate the same Bayesian VAR we have estimated based on the actual data, and we impose the identifying restrictions in Table

1. This produces, for each value of Φ_k and for each simulation s, a posterior distribution $F^s(\Phi_k)$, and a corresponding cumulative posterior distribution $F^s_C(\Phi_k)$, of the fraction of the long-run variance of GDP explained by H shocks.

By repeating this process for each Φ_k and for all s we thus build up the Monte Carlo distributions of the $F^s_C(\Phi_k)$. The evidence in Figure 5 suggests that the larger Φ_k, the more the Monte Carlo distributions of the $F^s_C(\Phi_k)$ shift to the right. For each Φ_k and each s we then compute the Kolmogorov-Smirnov (henceforth, KS) statistic between $F^s_C(\Phi_k)$ and the cumulative posterior distribution computed based on the actual data, $F^*_C(\Phi)$, which provides a statistical measure of the distance between the two distributions. In this way, for each Φ_k we build up the Monte Carlo distribution of the KS statistic. We will use these distributions in order to assess the true extent of hysteresis in the data taking into account of the distortions documented in the previous Section. We now discuss the proposed approach in more detail.

6.2 The Monte-Carlo Approach to Hysteresis in More Detail

Let the cointegrated structural VAR representation of (1) be

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

with A_0 being the impact matrix of the structural shocks ϵ_t, with $\epsilon_t = A_0^{-1} u_t$, where the u_t are the reduced-form shocks from (1). We define the matrix of the long-run impacts of the structural shocks as $L = C \times A_0$, where $C = \beta_{OC}[\alpha'_{OC}(I_N - B)^{-1}\beta_{OC}]\alpha'_{OC}$; $B = B_1 + B_2 + ... + B_p$; and α_{OC} and β_{OC} are the orthogonal complements of α and β. Finally, let D be the number of draws from the posterior distribution. We now describe how we impose specific values of Φ in the Bayesian structural VARs estimated based on the actual data, and how we use the thus-obtained models as DGPs in the Monte Carlo exercise.

The first step is to select among the D models the one that is closest to the 'median model'. Let $\bar{A}_0$, $\bar{B}$, and $\bar{L}$ be the medians of the posterior distributions of A_0, B, and L. Our selection criterion minimizes

$$\Gamma_d = \sum_{i=1}^{N}\sum_{j=1}^{N}\left[(A_{0,d}^{ij} - \bar{A}_0^{ij})^2 + (L_d^{ij} - \bar{L}^{ij})^2\right] + \sum_{i=1}^{N}\sum_{j=1}^{Np}(B^{ij} - \bar{B}^{ij})^2 \tag{9}$$

for $d = 1, 2, ..., D$, namely 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 $\bar{L}$, where N is the number of variables in the VAR and p the lag order. We thus obtain the model

$$\Delta Y_t = \tilde{B}_0 + \tilde{B}_1 \Delta Y_{t-1} + ... + \tilde{B}_p \Delta Y_{t-p} + \tilde{\alpha}\tilde{\beta}' Y_{t-1} + \tilde{A}_0 \epsilon_t. \tag{10}$$

Finally, we impose each value of $\Phi_k \geq 0$ in the grid upon model (10) by appropriately rescaling the column of $\tilde{A}_0$ associated with H shocks.[8] We then use the resulting model as DGP in the Monte Carlo exercise, thus building up the Monte Carlo distribution of the KS statistic for each value of Φ_k. The KS statistic for $F_C^*(\Phi)$ and $F_C^s(\Phi)$ is defined as:

$$KS = \sup_{\Phi} |F_C^*(\Phi) - F_C^s(\Phi)|, \tag{11}$$

with $\Phi \in [0, 1]$, i.e. as the maximum absolute value of the difference between the two cumulative distributions. We compute univariate KS statistics for GDP, consumption, and investment. For each simulated sample s, with $s = 1, 2, ..., S$, the most plausible value of the extent of hysteresis in the actual data is the value of Φ_k that minimizes (11). Accordingly, in the next sub-section we will focus upon the value of Φ_k corresponding to the minimum of the median of the Monte Carlo distribution of the KS statistic.

We also extend the univariate KS statistic in equation (11) to the bivariate case as follows. Let $F_C^*(\Phi_y, \Phi_c)$ and $F_C^s(\Phi_y, \Phi_c)$ be the joint cumulative posterior distributions of the fractions of long-run variance of GDP (Φ_y) and consumption (Φ_c) explained by H shocks. The joint KS statistic for Φ_y and Φ_c is:

$$KS_J = \sup_{\Phi_y, \Phi_c} |F_C^*(\Phi_y, \Phi_c) - F_C^s(\Phi_y, \Phi_c)|, \tag{12}$$

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 both the full sample and the sample excluding the ZLB period. Figure 6 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 individual statistics for GDP, consumption, and investment, and the joint statistic for GDP and consumption. As in the previous exercises there is a fairly high degree of uncertainty, which is intrinsic to the analysis we are performing. When we focus on the median estimates, two main results emerge from the figure.

For the sample excluding the ZLB period evidence uniformly suggests that the most likely value of the extent of hysteresis is zero. For the full sample period the evidence is

---

[8]For $k=1$, if we imposed $\Phi_1 = 0$ the model would become stochastically singular, so that based on the simulated data it would be impossible to estimate the reduced-form model (1). Accordingly, we therefore impose $\Phi_1 = 10^{-6}$ instead.
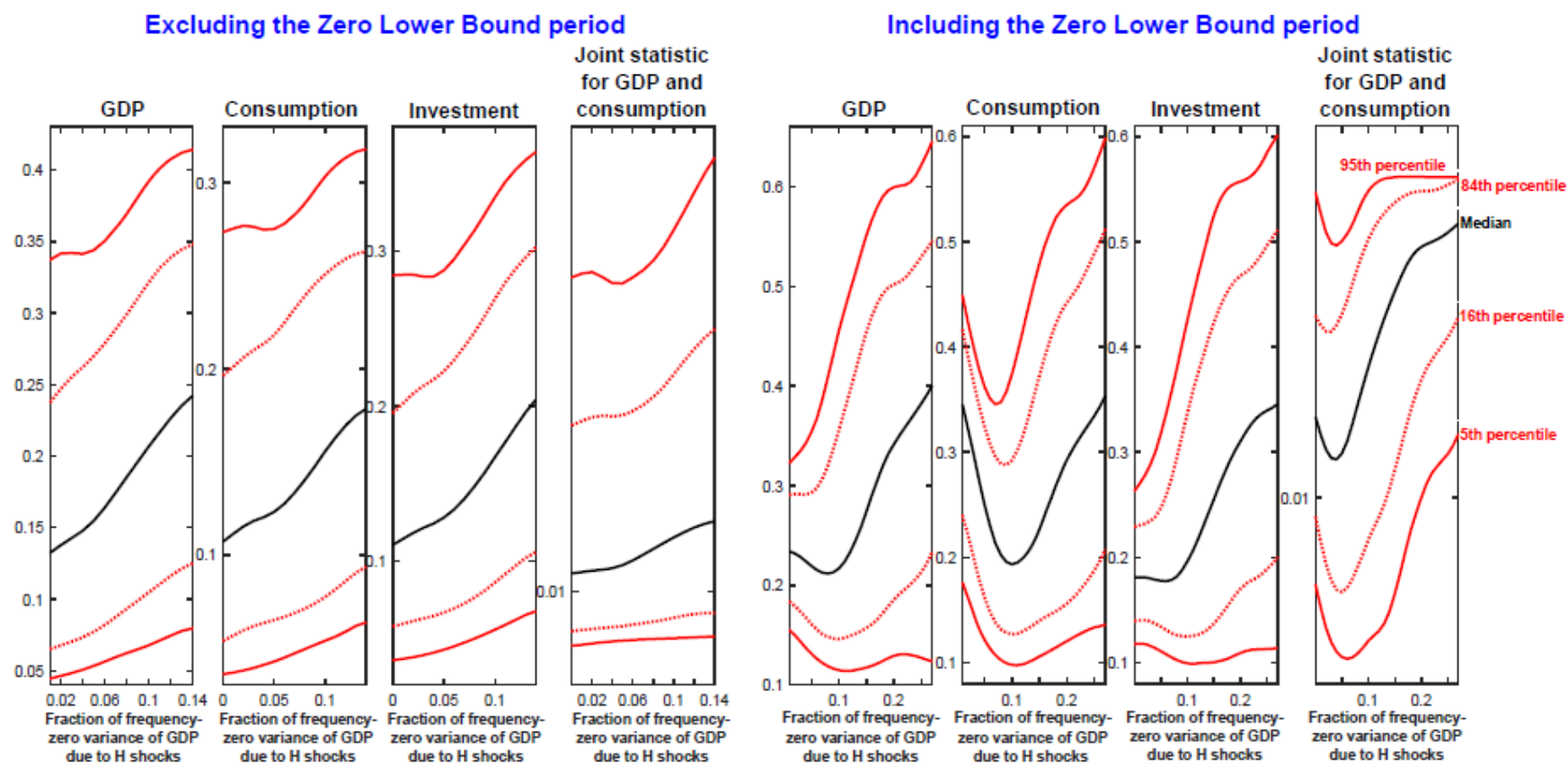
Excluding the Zero Lower Bound period

GDP Consumption Investment

Joint statistic
for GDP and
consumption

Including the Zero Lower Bound period

GDP Consumption Investment

Joint statistic
for GDP and
consumption

Fraction of frequency-zero variance of GDP due to H shocks

95th percentile
84th percentile
Median
16th percentile
5th percentile

Figure 6 Median and selected percentiles of the Monte Carlo distribution of the Kolgomorov-Smirnov statistic

less clear-cut, but overall it points towards a modest extent of hysteresis. This is the case based on both the joint KS_J statistic and the individual KS statistic 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, whereas the individual statistic for consumption reaches its minimum around 10 percent. Evidence for the individual statistic for GDP is broadly consistent with these values. The statistic for investment, on the other hand, is essentially flat until about 7-8 percent and it then starts to increase rapidly, which suggests that larger values are less plausible. 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 surprising that evidence for consumption is sharper than for GDP.

7 Conclusion

We have gone on a search for hysteresis, looking for aggregate demand shocks that have a permanent impact on real GDP. Methodologically, our search is primarily empirical and statistical, and it is mainly based on cointegrated structural VARs that identified such shocks via a combination of zero long-run restrictions and both short- and long-run sign restrictions. 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. Conceptually, we propose a simple approach to test for the presence and extent of hysteresis. It 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.

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 including the period following the collapse of Lehman Brothers. Even then, however, these effects only account for at most 10 percent of the frequency-zero variance of GDP. An important role in obtaining our results is played by a Monte Carlo-based approach 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.

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 rate as one possible hysteresis channel.

References

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

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

[8] 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.

[9] 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.

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

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

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

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

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

[15] 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.

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

[17] 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.

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

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

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

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

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

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

[24] 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.

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

[26] 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.

[27] 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.

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

[29] 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.

[30] 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.

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

[32] 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.

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

[34] 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.[9]

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):

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

---

[9]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 $\lambda^2 = 10$.

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 $a^{(*)} = \alpha^{(*)}(\alpha^{(*)'}\alpha^{(*)})^{-1/2}$; 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.19 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.