Working Paper Series

Searching for Hysteresis

WP 21-03

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‡

February 2021

Abstract

We search for the presence of hysteresis, which we define as aggregate demand shocks that have a permanent impact on real GDP, in the U.S., the Euro Area, and the U.K. Working with cointegrated structural VARs, we find essentially no evidence of such effects. Within a Classical statistical framework, it is virtually impossible to detect such shocks. Within a Bayesian context, the presence of these shocks can be mechanically imposed upon the data. However, unless a researcher is willing to impose the restriction that the sign of their long-run impact on GDP is the same for all draws, which amounts to imposing the very existence of hysteresis effects, the credible set of the permanent impact uniformly contains zero. We detect some weak evidence only for the U.K., originating from an increase in labor force participation and a fall in the unemployment rate.

Keywords: Bayesian methods, transitory shocks, GDP growth.
JEL Classification: E2, E3.

*We wish to thank Rodney Strachan for very helpful suggestions and Francesco Furlanetto for sharing his work in progress. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Richmond or 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 typically regarded as transitory (e.g., monetary policy shocks) can have very long-lived, or even permanent effects. It is most often used within the context of the labor market, where long and deep recessions can lead to long-term unemployment and, through this channel, could raise the natural (or equilibrium) unemployment rate. Similarly, a broad array of demand shocks, such as a surprise expansion of the money supply, could have long-lasting effects on productivity and output through channels such as business formation or exit. The experience of the financial crisis and the Great Recession, and that of the ongoing recession caused by the COVID-19 pandemic, have rekindled interest in the possible presence of hysteresis effects in aggregate data.

In this paper we embark on an empirical search for hysteresis. While the theoretical channels for the transmission of hysteresis shocks are well-understood, empirical evidence is few and far between. Working with statistical models that allow us to decompose macroeconomic time series into permanent and transitory components, we impose structural identifying restrictions in order to search for the presence of hysteresis. Overall, we find almost no evidence of these effects in the aggregate data for the U.S., the Euro Area, and the U.K. This finding goes beyond standard uncertainty about whether the presence of hysteresis is statistically different from zero or not. Using a Classical econometric approach, it proves to be essentially impossible to find such hysteresis shocks under our identification assumptions. In contrast, within a Bayesian setting we find that the credible set of long-run impacts of the identified shocks contains zero.

Permanent-transitory decompositions of economic time series, such as that of Blanchard and Quah (1989), by definition classify shocks as either permanent or transitory. Working within this framework, the only way to search for hysteresis effects is therefore to explore whether, within the set of permanent shocks, it is possible to find disturbances that are typically regarded as transitory, such as monetary policy shocks, or more generally aggregate demand shocks. On the other hand, many (or even most) aggregate supply shocks are typically thought to have permanent effects on output: this is the case of TFP (i.e., neutral technology) shocks,[1] investment-specific technology disturbances,[2] and shocks to either population or labor force participation. Our strategy therefore hinges on searching

[1] To be precise, several papers, for instance, Barsky and Sims (2011) and Benati et al. (2020) find that *news* TFP shocks have a permanent impact on TFP and GDP, whereas non-news shocks are transitory.

[2] See, among many others, Fisher (2006).

for permanent GDP shocks with aggregate demand features.

We identify aggregate demand and aggregate supply shocks by imposing sign restrictions on both their short- and long-run impacts upon GDP and the price level. We implement this strategy within both a Classical and a Bayesian framework. Since sign restrictions are typically imposed within a Bayesian context, the core of our analysis is based on Bayesian methods. At the same time, a Classical approach allows for a somewhat cleaner test of the hypothesis that hysteresis effects do in fact exist.

One of our main findings is that, working within a Classical context, it is essentially impossible to detect permanent GDP shocks that exhibit aggregate demand features. This result then motivates the Bayesian analysis, which allows us to impose upon the data a broader set of restrictions. For standard priors about the impact of demand and supply shocks, the Bayesian evidence agrees with the Classical findings as to the absence of hysteresis effects in the data, at least for the economies and periods analyzed herein. In particular, one of our main results is that, unless a researcher is willing to impose upon the data the restriction that the sign of the long-run impact on GDP of identified demand shocks is the same for all draws from the posterior - which amounts to imposing the very existence of hysteresis effects - the credible set of such long-run impact almost always contains zero, even at the one standard deviation significance level. We regard this as a crucial piece of evidence, because it means that the only way to detect hysteresis effects is to impose upon the data their very existence by "brute force".

Our analysis goes beyond the U.S. to also consider the Euro area and the U.K. A key reason for this is the fact that, as shown by Cogley (1990), in terms of the size of the unit root in GDP, the U.S. is an outlier with all of the other countries he analyzed featuring a larger permanent component, sometimes significantly so. This suggests that, in searching for hysteresis effects on GDP, the U.S. is likely not the best candidate: for instance if, in the limit, U.S. GDP were trend-stationary, there would be no hysteresis effects by definition. Consequently, we may expect other countries to produce stronger evidence, if there is any. As it turns out, we do indeed uncover some weak evidence of hysteresis in the U.K. for some of our specifications, although it is far from conclusive.

The classic paper on hysteresis is Blanchard and Summers (1986), who introduced the concept to the economics profession and used it to explain the persistence of European unemployment in the 1980s. Although they provide some basic statistical evidence, their main argument is largely based on a theoretical model of the labor market. In a similar vein, Ljungqvist and Sargent (1998) provide a more microfounded theoretical approach to the same issue and offer corroborating evidence from labor market data. More recently, theoretical frameworks based on the idea of endogenous TFP also allow for a possible role of hysteresis effects (e.g., Anzoategui et al., 2019; or Jaimovich and Siu, 2020). These ideas also underlie the evidence in Cerra and Saxena (2008), who show that deep recessions permanently reduce the productive capacity of an economy.

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. A classic study of hysteresis is Ball (2009), who uses 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 does argue 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 DSGE 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 and price inflation experience.

Furlanetto et al. (2020) is the paper that is closest to our work. They also use a structural VAR framework that combines long-run and short-run identification as in Blanchard and Quah (1989) with sign restrictions. They do detect an important role for hysteresis effects in U.S. data. The main differences with our analysis is that they do not impose the cointegration restriction central to aggregate macroeconomic data with unit roots and consider a different set of variables. Arguably, this leads to their finding that hysteresis is quantitatively considerably more important than in our analysis. Moreover, we highlight in our Bayesian analysis the role that identification assumptions play in obtaining these conclusions. 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 briefly outlines our approach to searching for hysteresis effects on output, whereas section 3 describes the data. Section 4 presents evidence based on a Classical approach, whereas section 5 presents the Bayesian evidence. Section 6 concludes.

2 Empirical Approach

In this section we provide an outline of our empirical approach. We first discuss how we disentangle permanent and transitory GDP shocks, which is central to the identification of hysteresis effects in the data. We then introduce our empirical model as the assumed data-generating process (DGP) before delving deeper into the issue of differentiating permanent GDP shocks into demand and supply components.

2.1 Disentangling Permanent and Transitory GDP Shocks

Hysteresis is the idea that disturbances that are typically regarded as transitory can have very long-lived, or even permanent effects on real variables, such as output or unemployment. The classic example is that of a contractionary monetary policy shock, which causes a recession and leads to an increase in the unemployment rate. If the shock is sufficiently large, it is conceivable that a fraction of the newly unemployed workers could become permanently unemployed (e.g., because of waning skills), which would result in a permanent increase in the equilibrium unemployment rate. The monetary policy shock would therefore induce hysteresis in unemployment although as a nominal disturbance should have no permanent impact on unemployment or GDP according to the Classical Dichotomy.

In theoretical models, such as Ljungqvist and Sargent (1998), there is a clear causal chain linking the initial shock to the ultimate outcomes, including a specific economic mechanism inducing hysteresis. In an empirical model, this chain needs to be teased from the data. The first and crucial step to identify hysteresis effects is to reliably disentangle permanent and transitory GDP shocks. However, separating the two types of innovations is a necessary, but not sufficient condition for identifying hysteresis.

For example, assume that all aggregate supply-side shocks are permanent, while all aggregate demand-side shocks are transitory, in line with the specification in Blanchard and Quah (1989).[3] If our empirical methodology can separate permanent and transitory shocks, we would conclude correctly that, in fact, there are no demand shocks having a permanent impact on GDP; that is, there is no hysteresis. On the other hand, suppose that the methodology produces an unreliable permanent-transitory decomposition, such that shocks are randomly classified as either permanent or transitory with equal probability. Since half of the demand shocks are incorrectly classified as permanent, we would spuriously detect hysteresis when, in fact, there is none.

[3] For sake of the argument, how such shocks are defined is not important. In our empirical application, we provide specific definitions of demand- and supply-side shocks, which we use throughout the paper.

Identifying the permanent components of economic time series and the underlying shocks is a long-standing issue in applied econometrics and is fraught with difficulties (see, for instance, Faust and Leeper, 1997). In order to improve our chances of effectively disentangling permanent and transitory GDP shocks we impose cointegration restrictions in our empirical model. In addition, we require that any shock having a permanent impact on GDP also has a permanent impact on the variables it is cointegrated with. We assess the plausibility of the permanent/transitory decomposition produced by the model in terms of whether the transitory component of GDP captures well-known and widely accepted features of macroeconomic fluctuations, such as the peaks and troughs of the business cycle. We also consider whether the transitory component of GDP is strongly correlated with standard estimates of the output gap.[4] Consequently, we regard any methodology that produces implausible transitory components of GDP as unreliable for the purpose of searching for hysteresis effects in the data.

2.2 The Statistical Representation of the Economy

We employ cointegrated VARs for seven standard macroeconomic variables: real GDP, real consumption, and real investment in logarithms; a short- and a long-term nominal interest rate; either the logarithm or the log-difference of the GDP deflator depending on their integration properties; and either the logarithm of total hours worked, or its underlying determinants, that is, the unemployment rate, and the logarithms of population, the participation rate, and average hours worked per worker.

There are two main reasons for working with cointegrated VARs in the study of hysteresis. First, since our focus is on permanent GDP shocks, working with VARs that feature cointegration between GDP, consumption, and investment represents a natural choice, which enforces log-run relationships that have to hold in the data (King et al., 1991). By exploiting the information content of the cointegrated series, it allows for a more precise identification of these disturbances, compared with the case in which we had uniquely focused on GDP. In fact, this modeling framework produces plausible transitory components of GDP.

Second, our use of cointegration methods likely avoids the concerns raised by Jordà et al. (2020), who criticize evidence produced by monetary VARs as not reliably identifying the long-run impact of monetary shocks. We guard against this concern since our approach focuses on the set of shocks that have a permanent impact on GDP. Our use of cointegration

[4] This is in line with Galí (1999), who argues against the plausibility of technology shocks as the dominant driver of business-cycle fluctuations, since the components of GDP and hours uniquely driven by such shocks are difficult to reconcile with any of the postwar cyclical episodes.

sharpens the identification of these shocks by exploiting the cointegration restriction that any such shock will also have a permanent impact on consumption and investment. In contrast, standard monetary VARs tend to focus on the short- to medium-run impact of monetary shocks, and they typically do not include consumption, which contains crucial information about the permanent component of GDP (see Cochrane, 1994).[5]

2.3 Identifying Demand and Supply Shocks

Many narratives and theoretical models characterize hysteresis as originating from demand shocks, such as monetary policy surprises (e.g., Jordà et al., 2020). While it is widely recognized that many (or even most) supply shocks have a permanent impact on output, demand shocks are typically seen as being of a temporary nature, unless there is an endogenous mechanism, such as shifts in the wealth distribution or loss of skills, that induces permanent effects. We maintain this broad classification of supply and demand shocks, but we do not take a specific stand on the nature of the transmission mechanism leading to hysteresis.

Once we have identified the set of permanent GDP shocks, we then classify them based on whether they exhibit aggregate demand or aggregate supply features. To this end, we impose identifying restrictions that are motivated by a standard aggregate-demand/aggregate-supply framework as is common in many macroeconomic textbooks but is also generally consistent with the implications of standard New Keynesian models. In particular, we assume that:

1. aggregate demand is downward-sloping both in the short run and in the long run;
2. aggregate supply is upward-sloping in the short run and vertical in the long run;
3. permanent aggregate supply shocks only affect aggregate supply;
4. permanent aggregate demand shocks may or may not affect the aggregate supply curve depending on whether there are hysteresis effects.

The first three identification assumptions are standard and are consistent with a wide variety of macroeconomic frameworks ranging from the textbook AS-AD model, to the New Keynesian models and large-scale macroeconomic models used in policy institutions. The

[5] The VAR literature typically uses a specification in levels, which can, in principle, reliably capture cointegration between GDP and other series containing information about its permanent component (for instance, Sims et al., 1990, Sims and Zha, 2006, or Arias et al., 2018). The extent to which this is the case in typical sample sizes is still a concern.

fourth assumption captures the essence of the notion of hysteresis. In such a case, a negative (positive) aggregate demand shock has a negative (positive) permanent impact on long-run aggregate supply. Typical examples are the idea that deep recessions, such as the Great Recession, 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.

These assumptions imply a set of identifying restrictions for the joint dynamics of GDP and the price level. In the short run, demand shocks impact GDP and prices in the same direction, while supply shocks impact GDP and prices in opposite directions.[6] In the long run, permanent supply shocks permanently impact GDP and prices in opposite directions. However, the long-run impact of demand shocks on GDP and prices crucially hinges on whether there is hysteresis. Without hysteresis, the aggregate supply curve remains unaffected: the only impact is on the price level, while GDP remains unchanged. With hysteresis, demand shocks permanently shift GDP in the same direction, whereas their impact on prices is ambiguous. For instance, a recessionary shock permanently decreases both aggregate demand and aggregate supply and thereby unambiguously decreases GDP permanently, whereas its impact on the price level could be either negative or positive.

In the empirical analysis, we impose the short-run restrictions either only on impact or up to K periods ahead. Further, we impose the long-run restrictions that permanent supply shocks permanently affect GDP and that a negative (positive) demand shock either (*i*) permanently decreases (increases) prices leaving GDP unaffected (in the case of no hysteresis); or (*ii*) permanently decreases (increases) GDP, whereas its impact on prices is ambiguous (in the case of hysteresis). Finally, in order to sharpen the inference, we impose that any shock with a permanent impact on GDP also permanently affects consumption and investment in the same direction.

Summing up, supply shocks impact GDP and prices in opposite directions both in the short run and in the long run. Demand shocks impact GDP and prices in the same direction in the short run. As for the long run, a negative (positive) demand shock has a negative (positive) permanent impact on either of the two variables (or both).

[6] These restrictions follow the DSGE-based 'robust sign restrictions' of Canova and Paustian (2011). Their model features two demand shocks, namely a monetary policy and a taste shock. Both shocks imply impacts on GDP and inflation of the same sign.

2.4 Data

We search for hysteresis in three economies, the U.S., the Euro Area, and the U.K. Throughout the paper, we use quarterly data. The data sources are discussed in detail in Appendix A. Our list of variables includes (the logarithms of) real GDP, real consumption, and real investment. We capture the behavior of the labor market either by the logarithm of total hours worked or by their underlying determinants, i.e., population, the participation rate, unemployment, and average hours worked. Since there are no quarterly population data available for the Euro Area, we decompose employment into its underlying determinants, i.e., the labor force and the unemployment rate. Euro Area hours worked are only available since 1995Q1, so that we use employment instead. We elaborate on the respective specifications in Section 4.

We include a short- and a long-term nominal interest rate and the GDP deflator, either in log-levels or in log-differences, depending on its integration properties. The longest sample periods we consider include the financial crisis and the ZLB periods in the three economies. In order to capture the effective policy stance to the extent possible, we also consider specifications that include Wu and Xia's (2016, 2017) "shadow rate" instead of the short-term rate for any of the three economies. These rates attempt to summarize the overall monetary policy stance, in particular the impact of QE and forward guidance policies. For the Euro Area, we terminate the sample in 2017Q4, since the shadow rate is only available up to that quarter.

It is common practice in empirical studies to express trending variables in per-capita terms, as this removes one source of non-stationarity that may not be of immediate interest for the question studied. We choose not to follow this approach. For instance, after a hysteresis shock, one of the mechanisms through which aggregate supply adjusts is via migration flows.[7] This produces hysteresis effects through changes in the population and the labor force. To the extent that this effect is non-negligible, working with per-capita variables would automatically underestimate the possible presence of hysteresis effects.

Our sample periods are 1954Q3-2019Q3 for the U.S., 1970Q1-2017Q4 for the Euro Area, and 1971Q1-2016Q4 for the U.K. The different sample lengths are determined by the availability of data, whereby we select those periods with the widest coverage of the key variables as baseline samples. As a robustness check, we also consider shorter samples that exclude

[7] In the wake of the Great Recession, this mechanism has been extensively discussed. Given the scale of the downturn and the disruptions caused, it is likely that a non-negligible fraction of the foreign workers who have returned to their home countries will never return.

the zero lower bound (ZLB) period for the U.S. and the Euro Area. The respective samples are 1954Q3-2008Q4 for the U.S. and 1970Q1-2014Q2 for the Euro Area. We define the non-ZLB samples based on the requirement that the policy (or short-term) interest rate is consistently and materially above zero during the entire period. Specifically, we require the interest rate to be greater than or equal to a threshold, which we set to 25 basis points, for each single quarter. Since the monetary policy rate in the U.K. never fell below the threshold for the full sample period 1971Q1-2016Q4, we focus on this sample.

3 Classical Evidence

Our empirical results are based on two alternative methodologies, namely Classical or Bayesian methods. We begin by presenting evidence based on a Classical framework. First, we analyze the unit root and cointegration properties of the data in order to properly specify the cointegrated system. We then provide evidence on the plausibility of the permanent-transitory decomposition of GDP produced by the cointegrated VARs. The conclusion we obtain from this analysis is one of the paper's main results: in a Classical cointegration framework, it is essentially impossible to detect aggregate demand shocks that have a permanent impact on real GDP.

3.1 Unit Root and Cointegration Properties of the Data

In order to determine the order of integration of the series, we perform the unit root tests in Elliot et al. (1996). The results are reported in Table A.1 in the Appendix. To summarize, we cannot reject the null hypothesis of a unit root for any of the series, with a few exceptions. First, we can reject the null hypothesis for the log-difference of the GDP deflator for the U.K., but not for the U.S. or the Euro Area. Second, the null of a unit root is strongly rejected for the U.S. unemployment rate. We treat the remaining two exceptions, Euro Area consumption and U.K. investment, as statistical flukes[8] and proceed under the assumption that all of the remaining series feature a unit root.

In the next step, we analyze the cointegration properties of the data based on the insights from the unit root tests. We specify the VECM for the cointegration analysis with the logarithm of the GDP deflator for the U.K., whereas we work with inflation (computed as the log-difference of the GDP deflator) for the U.S. and the Euro Area. Whenever our

[8] Even if a test is correctly sized, it incorrectly rejects the null x percent of the time at the x percent level. When performing a large number of unit root tests, as is the case here, a certain fraction of such type-I errors is to be expected.

specification is the expanded system, in which either total hours worked, or employment, are decomposed into their underlying determinants, we use the logarithm of the level of unemployment[9] in the VECM for the U.S.

Standard economic theory suggests that there are at least three cointegration relationships in our empirical specification. GDP, consumption and investment are likely to share two cointegration vectors, while a third should exist between the short- and the long-term nominal interest rates. Table 1 provides evidence to this effect. It reports bootstrapped p-values[10] for Johansen's maximum-eigenvalue test[11] of 0 versus 1 cointegration vectors in three bivariate systems that feature either GDP and consumption, GDP and investment, or the short and the long rate. In all three economies, and for any of the three bivariate systems, evidence of cointegration is very strong, with the largest p-value being equal to just 0.0346.

We now perform cointegration tests for the baseline system that includes the 7 key variables and for a larger system featuring additional labor market variables (i.e., the underlying determinants of hours). The results are reported in Table 2, which shows bootstrapped p-values for Johansen's maximum eigenvalue tests of the null hypothesis of h versus $h + 1$ cointegration vectors.[12] In the case of the 7-variable systems, the results suggest three cointegration vectors for the U.S. and four for the Euro Area and the U.K. In the extended systems, we detect three cointegration vectors for the U.K. and four for the Euro Area and the U.S.

Given these results, we proceed as follows. For the U.S., we impose three and four cointegration vectors upon the 7-variable and the extended system, respectively. We assume four cointegration vectors in Euro Area data based on either specification. Finally, we impose four cointegration vectors when working with the 7-variable system for the U.K., and either three or four in the extended specification.[13]

[9] For the log-level of U.S. unemployment (as opposed to the unemployment rate) the null of a unit root cannot be rejected.

[10] The bootstrap procedure is implemented as in Cavaliere et al. (2012). Benati (2015) and especially Benati et al. (2021) provide extensive Monte Carlo evidence on the superior performance of this bootstrapping procedure.

[11] The corresponding results from the trace test are qualitatively the same and are available upon request.

[12] In all cases considered, Johansen's trace tests (not reported) overwhelmingly reject the null hypothesis of no cointegration against the alternative of one or more cointegration vectors, with bootstrapped p-values ranging between 0.000 and 0.005.

[13] We consider four cointegration vectors in the larger system for the U.K. since these relationships should carry over from the smaller system, despite decomposing total hours worked into its underlying determinants. In fact, this makes essentially no difference because the evidence for the U.K. obtained by imposing either three or four cointegration vectors in the larger system is qualitatively the same. For reasons of brevity, we only report and discuss evidence based on imposing four cointegration vectors, but the alternative set of results is available upon request.

3.2 Decomposing GDP into Permanent and Transitory Components

We estimate the VECMs under the identified cointegration relationships for the economies and specifications discussed above. Specifically, we first estimate the reduced-form cointegrated VECM based on the Johansen (1991) estimator as detailed in Hamilton (1994, pp. 630-645), whereby we impose the identified number of cointegration vectors in the estimation. We then impose the restriction that there are only two shocks that have a permanent impact on GDP. Having identified such shocks, we compute the transitory components of GDP by "re-running history" as in Blanchard and Quah (1989), that is, by conditioning on transitory GDP shocks only.

Figure 1 reports the point estimates of the transitory components of GDP for the U.S. and the U.K. For reference, we also include output gap estimates from the U.S. Congressional Budget Office and the U.K. Office for Budget Responsibility.[14] In addition, we show the peaks and troughs of the business cycle in the U.S. as vertical blue lines, as established by the NBER Business Cycle Dating Committee. The figure shows that in both countries the transitory component of GDP produced by the cointegrated VAR comoves closely with the output gap estimates. In the U.S., the relationship appears closer after 1980, although there are some discrepancies after the end of recessions, such as in the early 1980s. At the same time, fluctuations in the transitory component of GDP capture well the NBER business cycle peaks and troughs. The take-away from this exercise is that the permanent-transitory decomposition produced by the cointegrated VARs provides a plausible and reasonable approach to identifying permanent movements.

3.3 Identifying Hysteresis Shocks

We now turn to one of the key elements of the paper. Having established a decomposition of our time series in permanent and transitory components, we investigate whether we can identify aggregate demand disturbances within the set of shocks that permanently impact GDP. We disentangle aggregate demand and aggregate supply shocks by imposing sign restrictions on both their short- and their long-run impacts on GDP and the price level.

Specifically, we consider the set of all possible rotations of the two permanent GDP shocks, which we obtain by applying the rotation matrix,

$$R = \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ \text{-}\sin(\omega) & \cos(\omega) \end{bmatrix},$$

[14] We do not include a comparable measure for the Euro Area since it is not publicly available.

to the corresponding columns of the impact matrix of the cointegrated structural VAR (SVAR) under the point estimates, with $\omega \in [-\pi, \pi]$. We identify demand and supply shocks based on assumptions (1)-(4) described in section 2.3. Importantly, we impose the signs of the short-run responses of GDP and the GDP deflator only on impact.

Figure 2 shows the results from this exercise. For each of the three economies, we show the number of identified demand and supply shocks over the domain of the rotation angle ω. The results appear fairly clear. In every economy considered, an appropriate rotation of the two permanent GDP shocks allows us to find at least one shock that exhibits aggregate supply features. In fact, for some values of ω, we uncover two such shocks in the U.S. and the Euro Area. In contrast, the evidence for aggregate demand, i.e., hysteresis, shocks is considerably weaker: there is no identified permanent demand shock in the case of the U.S., and only very weak evidence of one shock for the Euro Area and the U.K.

One obvious concern about this analysis is that it is based on the point estimates of the statistical model. The latter can be thought of as the reduced form of a structural model that we identify using the scheme above. However, the point estimates of the model may not be fully representative of the full range of likely outcomes when compared to plausible, nearby perturbations of the reduced form. In order to address this concern, we conduct the following exercise based on the bootstrap procedure of Cavaliere et al. (2012). For each country, we simulate the cointegrated VAR 10,000 times, conditional on the identified number of cointegration vectors. We then estimate a cointegrated VAR on each artificial sample, where we impose the same number of cointegration vectors as in the simulation, and we identify two permanent GDP shocks. Finally, we compute the average number of identified demand and supply shocks within the set of all possible rotations of the two permanent shocks, i.e. over the interval $[-\pi, \pi]$ for ω. The results from this exercise are consistent with the findings based on the point estimates. The median of the bootstrapped distribution of the average number of identified demand shocks is equal to 0.000, 0.018, and 0.156 for the U.S., the Euro Area, and the U.K., respectively.

In summary, the evidence for hysteresis shocks, namely identified aggregate demand shocks that have a permanent effect on GDP, is at best fairly weak. Even when considering plausible perturbations of the reduced form around the point estimates, detecting demand shocks appears to be exceedingly difficult in a Classical VAR setting. We now turn to evidence obtained with a Bayesian approach, which provides more leeway to guide the estimation and dig deeper into the apparent lack of hysteresis.

4 Bayesian Evidence

In this section we perform a similar empirical analysis on the data sets under investigation, but from a Bayesian perspective. We start by discussing technical details pertaining to the identification of the cointegration vectors, the estimation of the reduced-form, and the imposition of the identifying restrictions. We then move on to presenting the evidence and discussing some caveats.

4.1 Methodology

Our approach is adapted from methods proposed by Strachan and Inder (2004) and Koop et al. (2010), which we discuss in the following sub-sections.

4.1.1 Bayesian Identification of Cointegration Vectors

When estimating cointegrated VARs, a key issue is the identification of the cointegration vectors. Since only the cointegration space is uniquely identified, the cointegration vectors are identifiable up to a rotation. Consider the VECM representation:

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

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

For the r cointegration vectors to be uniquely identified, each of the r columns of the matrix β ought to feature at least r restrictions (see Kleibergen and van Dijk, 1994, or Bauwens and Lubrano, 1996). Within a Classical context, the standard solution proposed by Johansen (1988, 1991) is to rely on the identification method for reduced-rank regression models introduced by Anderson (1951). In a Bayesian setting, several methods have been proposed, as discussed by Koop et al. (2006). 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.[15] This restriction is imposed by defining a semi-orthogonal matrix $H = H_g(H_g' H_g)^{-1/2}$, which is used to center the prior for the cointegration space around the value that a researcher considers the most plausible. Since H and H_g

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

span the same space, this is obtained by appropriately selecting the columns of H_g based on a priori information, for instance derived from economic theory.

Our baseline specification is a 7-variable system represented by equation (1), where $Y_t = [y_t, c_t, R_t, r_t, p_t \text{ (or } \pi_t), i_t, h_t \text{ (or } e_t)]'$. y_t, c_t, i_t, p_t, h_t, and e_t are, respectively, the logarithms of GDP, consumption, investment, the GDP deflator, total hours worked, and employment; $\pi_t = p_t - p_{t-1}$ is inflation; and R_t and r_t are the short- and the long-term nominal interest rates. As an example, with three cointegration vectors, H_g is set to:

$$H_g' = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{bmatrix}. \quad (3)$$

The three columns encode the three cointegration relationships that we identified in Table 1: between GDP and consumption, GDP and investment, and the short and the long rate. Similarly, when we impose four cointegration vectors, we set H_g to:

$$H_g' = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4)$$

The fourth row encodes our conjecture that the system only features the three cointegration vectors that would be suggested by standard economic theory, despite the statistical evidence we previously obtained. Finally, when working with larger systems, we appropriately modify either (3) or (4) by augmenting them with a row of zeros for each of the underlying determinants of either hours or employment.

4.1.2 Bayesian Estimation of the Cointegrated VAR

We estimate the cointegrated VAR using 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 we then discard. We then generate 200,000 draws, which we "thin" by sampling every 100 draws in order to reduce their autocorrelation. This leaves us with 2,000 draws from the ergodic distribution, which we use for inference.[16]

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

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:

$$\alpha \mid \beta, B_0, B_1, ..., B_p, \Sigma, \tau, \nu \sim MN\left(0, \nu(\beta' P_{1/\tau}\beta)^{-1} \otimes G\right), \tag{5}$$

where MN denotes the matricvariate-normal distribution; $P_{1/\tau} = HH' + \tau^{-1}H_\perp H_\perp'$; and G is a square matrix that is set to $G = \Sigma$ in line with 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. In what follows, we set $\tau = 1$, which corresponds to a flat non-informative prior on β, and $\nu = 1000$, corresponding to a very weakly informative prior for α.[17]

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

$$p(\beta) \propto |P_\tau|^{-r/2}|\beta'(P_\tau)^{-1}\beta|^{-N/2}, \tag{6}$$

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:

$$p(\Sigma) \propto |\Sigma|^{-(N+1)/2}. \tag{7}$$

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.[18]

In the estimation step, we impose the restriction that the maximum deviation of the implied transitory component of GDP from an external output gap estimate is equal to at most k percent. This allows us to impose some discipline on the permanent-transitory decomposition of GDP as a basis for identifying hysteresis effects to ensure that the implied decomposition offers some degree of plausibility.[19] More specifically, we implement this restriction with a simple accept/reject step at the end of each Gibbs-sampling pass. This is conceptually in line with Cogley and Sargent's (2002) approach to imposing stationarity within a time-varying parameter VAR context. If at the end of step (*iv*) for pass j of the Gibbs sampler the restriction on the transitory component of GDP is violated, we reject all steps (*i*)-(*iv*) of pass j. Rather than moving to pass $j+1$, we repeat the entire cycle.

We report evidence based on $k = 5$ percent, but we also experimented with several alternative values of k, from 3 to 7 percent. All tend to produce very similar results.

[17] The standard non-informative prior for α would be obtained by setting $1/\nu = 0$.

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

[19] Since the transitory component of GDP is obtained by imposing "hard", namely zero long-run restrictions on the estimated reduced-form VECM, the mapping between the statistical model and the implied transitory component of GDP is one-for-one. Therefore, any restriction on the latter automatically maps into a restriction on the former.

Whether we do impose this restriction or not does not make much of a difference for our findings about the presence of hysteresis, which arguably speaks to the robustness of our results. Nevertheless, we prefer results under this restriction since they stem from a more plausible permanent-transitory decomposition of GDP.

Finally, we jointly impose the zero long-run restrictions and the short- and long-run sign restrictions, based on the by now standard methodology proposed by Arias et al. (2018). Details are provided in Appendix B. We impose the short-run sign restrictions both on impact and for the subsequent eight quarters.

4.1.3 Drawing from the Posterior Distribution

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

$$\Sigma \mid \alpha, \beta, B_0, B_1, ..., B_p, Y_t \sim IW\left(u_t' u_t + \nu\alpha(\beta' P_{1/\tau}\beta)^{-1}\alpha', T + r\right), \tag{8}$$

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

$$vec(A) \mid \alpha, \beta, \Sigma, \tau, \nu, Y_t \sim N\left\{(\Sigma^{-1} \otimes Z'Z + \lambda^2 I)^{-1}(\Sigma^{-1} \otimes Z'Z)vec(A'), (\Sigma^{-1} \otimes Z'Z + \lambda^2 I)^{-1}\right\}, \tag{9}$$

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:

$$b \mid a, B_0, B_1, ..., B_p, \Sigma, \tau, \nu \sim MN\left(0, (a'G^{-1}a)^{-1} \otimes \nu P_\tau\right), \tag{10}$$

$$p(a) \propto |G|^{-r/2}|a'G^{-1}a|^{-N/2}. \tag{11}$$

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

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

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

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

4.2 Empirical Evidence

We discuss results for two specifications. The first is the benchmark 7-variable system, which contains the key macroeconomic aggregates. The second specification decomposes hours (or employment) into their underlying labor market determinants, which allows us to obtain a more fine-grained view, in particular with respect to (un)employment.

As a first pass we compute the transitory component of GDP with and without the restriction on its maximum deviation from the external output gap estimate. The results are reported in the Appendix in Figures A.I.3 and A.I.4 with and without the restriction, respectively. Estimates for the former are very close to the external output gap estimates, and are narrowly clustered around them, whereas for the latter there are more discrepancies and a greater extent of uncertainty. Given these results, we prefer to work with the models imposing the restrictions on the transitory components of GDP.[20]

4.2.1 Baseline Specification

We discuss two sets of results that differ in terms of a key identification assumption, which is at the heart of the hysteresis discussion. In our first exercise we do not impose the restriction that the sign of the long-run impact of demand shocks on GDP is the same for all draws. The rationale for this restriction is that it effectively amounts to imposing the very existence of hysteresis shocks.[21] In contrast, in the second exercise we do impose the restriction, thus postulating that hysteresis effects do in fact exist. Our main finding is that unless such restriction is imposed, the credible set of the long-run impact of demand shocks on GDP almost always contains zero for all economies and model specifications.

Figures 3-5 show, for each of three economies we are analyzing, the impulse-response functions (IRFs) to the identified demand and supply shocks for horizons up to 15 years after impact. We report the median IRFs of the seven endogenous variables with the 16-84

[20] Qualitatively, the results produced by the two sets of estimates are very similar in terms of impulse-response functions and variance decompositions.

[21] From a practical standpoint, each draw j from the posterior (i.e., each model j) is a "candidate truth". As long as for draw j the long-run impact of demand shocks on GDP is not *exactly* equal to zero, hysteresis is, strictly speaking, always present.

and 5-95 percentiles of the posterior distributions.[22] For the U.S. and the Euro Area the estimates are based on the full samples. They also include Wu and Xia's "shadow rates", instead of the standard short-term nominal interest rate. The Appendix contains results for the 'non-ZLB' sample period, which are qualitatively the same and quantitatively close to those for the full sample.

Similar to the results based on Classical methods, Figures 3-5 do not reveal much evidence of hysteresis for any of the three economies. Notably, for all economies the credible sets of the IRFs for GDP, consumption, and investment consistently contain zero, even when narrowly defined based on the 16-84 percentiles of the posterior distribution. We provide further evidence in Table 3a, which shows the fractions of draws from the posterior associated with a positive impact of either demand or supply shocks on any of the variables in the long run. It is only in the case of the U.K. that there is some evidence of hysteresis, as the fraction of draws associated with a positive long-run impact on GDP is equal to 80 percent.

For completeness, we report in Figures A.I.7 to A.I.9 in the Appendix the results obtained by imposing the restriction that the sign of the long-run impact of demand shocks on GDP is the same for all draws. As discussed above, this restriction assumes the presence of hysteresis effects by definition. In a sense, the outcome is therefore mechanical and perhaps somewhat uninformative.[23] The IRFs show that permanent demand shocks do imply hysteresis effects, but this is by assumption. Consequently, we regard as meaningful results only those in Figures 3-5, which show that, consistent with the Classical evidence provided before, the data provide little support for the existence of hysteresis unless one is willing to assume it.

4.2.2 Extension: Decomposing Hours and Employment

We now consider a key extension of our benchmark specification. One potential limitation of the baseline model is that it could miss permanent impacts of demand shocks on the underlying determinants of either hours worked or employment. This is particularly pertinent since many, if not most, theoretical models of hysteresis derive it within a model of the labor market through the impact of unemployment spells. In principle, the long-run impacts of

[22] Figures A.I.10-A.I.12 in the Appendix report the fractions of forecast error variance explained by demand and supply shocks.

[23] Nevertheless, there is a caveat. As long as a researcher is willing to take a sufficiently large number of draws from the statistical model in a Bayesian context, it is in principle possible to conjure essentially any shock, i.e., impose on the data. Whereas the existence of technology shocks is arguably not in question, the fact that hysteresis effects do or do not exist is on the other hand entirely open to question.

demand shocks on the determinants of hours worked or employment could cancel out in the aggregate, which would therefore present an incomplete picture of the overall impact of these shocks.[24] We therefore present estimates based on expanded systems, in which we decompose either total hours worked or employment into their underlying determinants.

Let H_t be total hours worked and $H_t = E_t h_t$, where E_t and h_t are, respectively, employment and average hours worked. Let P_t and L_t be population and the labor force, so that $L_t = E_t + U_t$, where U_t is unemployment. The unemployment rate is therefore defined as $u_t = U_t/L_t$. We then have:

$$H_t = h_t(1 - u_t)lfp_t\ P_t, \tag{12}$$

where $lfp_t = L_t/P_t$ is the labor force participation rate. We can then write the logarithm of total hours as (approximately):

$$\ln H_t \simeq \ln h_t - u_t + \ln lfp_t + \ln P_t, \tag{13}$$

where we make use of the fact that $\ln(1 - u_t) \simeq -u_t$. In the case of the Euro Area, for which quarterly data on population are not available, we similarly write:

$$E_t = (1 - u_t)L_t \ \Rightarrow \ \ln E_t \simeq \ln L_t - u_t. \tag{14}$$

For the U.K., we replace $\ln H_t$ with $\ln h_t$, u_t, $\ln lfp_t$, $\ln P_t$; for the U.S., we use $\ln h_t$, $\ln Part_t$, $\ln P_t$, and $\ln U_t$;[25] and for the Euro Area $\ln E_t$ is replaced by $\ln L_t$ and u_t. Table 3*b* reports the corresponding fractions of draws associated with a positive long-run impact of either shock on any of the variables.

We report the results for the extended system in Figures 6-8. Overall, our findings are in line with those from the baseline specification. We detect no evidence of hysteresis effects on GDP for the U.S. and the Euro Area. In addition, there is no evidence of a permanent impact of demand shocks on any underlying determinants of hours and employment. Results for the U.K. are qualitatively the same as those in the baseline. However, we find somewhat stronger evidence of hysteresis. The fraction of draws associated with a positive long-run impact on GDP (and consumption and investment) is now around 95 percent, as opposed to the corresponding figure of 80 percent in Table 3*a*.

The evidence in Table 3*b* therefore provides some evidence that in the U.K. hysteresis may originate from a permanent increase in the participation rate and a permanent decrease

[24] In addition, a VECM system that does not include variables such as labor force participation might be informationally insufficient in the sense of Forni and Gambetti (2014).

[25] Recall for the log-level of unemployment the null of a unit root cannot be rejected.

in the unemployment rate with fractions of draws associated with positive permanent impacts of demand shocks of 78 and 23 percent, respectively. While this is in line with some theoretical models of hysteresis and also consistent with, for instance, the evidence provide by Erceg and Levin (2014) or Yagan (2019), we find this result too inconclusive to support the notion of hysteresis in the aggregate data.

5 Conclusion

We have gone on a search for hysteresis, looking for demand shocks in the aggregate data that have a permanent impact on real GDP. Methodologically, our search was primarily empirical and statistical, 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 pursued a two-prong approach. First, a Classical investigation of the hypothesis of hysteresis, and second, a Bayesian implementation, which allowed for flexibility in guiding the data and a more encompassing representation of the inherent uncertainty of the hysteresis question.

We are hard-pressed to find evidence of hysteresis effects, which are at best weak or simply non-existent. Within our Classical framework, it is virtually impossible to detect such effects for the three economies we considered, the U.S., the Euro Area, and the U.K. Within a Bayesian context, the presence of hysteresis shocks can be mechanically imposed upon the data *via* a strong prior belief. However, this requires a researcher to impose willingly the restriction that the sign of the long-run impact of hysteresis shocks on GDP is the same for all draws. In practice, this amounts to imposing upon the data the very existence of hysteresis. Without this restriction, we find that the credible set of the permanent impact on GDP nearly always contains zero. A partial exception is the U.K., for which we detect some evidence of hysteresis originating from an increase in labor force participation and a fall in the unemployment rate.

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 Dan F. Waggoner (2018): "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications". *Econometrica*, pp. 685-720.

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

[5] Barsky, Robert, and Eric Sims. (2011): "News Shocks and Business Cycles". *Journal of Monetary Economics*, 58(3), 273-289

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

[7] Benati, Luca. (2015): "The Long-Run Phillips Curve: A Structural VAR Investigation". *Journal of Monetary Economics*, 76, pp. 15-28.

[8] Benati, Luca, Joshua Chan, Eric Eisenstat, and Gary Koop (2020): "Identifying Noise Shocks". *Journal of Economic Dynamics and Control*, 111, 1-17

[9] Benati, Luca, Robert. E. Lucas, Jr., Juan-Pablo Nicolini, and Warren Weber (2021): "International Evidence on Long-Run Money Demand". *Journal of Monetary Economics*, forthcoming.

[10] 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.

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

[12] 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.

[13] Cavaliere, Giuseppe, Anders Rahbek, and A.M. Robert Taylor (2012): "Bootstrap Determination of the Cointegration Rank in Vector Autoregressive Models". *Econometrica*, 80(4), pp. 1721-1740.

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

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

[16] Cogley, Timothy (1990): "International Evidence on the Size of the Random Walk in Output". *Journal of Political Economy*, 98(3), pp. 501-518.

[17] Cogley, Timothy, and Thomas Sargent (2002): "Evolving Post-World War II U.S. Inflation Dynamics". In: *NBER Macroeconomics Annual 2001*, 16, pp. 331-388.

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

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

[20] Faust, Jon, and Eric Leeper (1997): "When Do Long-Run Identifying Restrictions Give Reliable Results?" *Journal of Business and Economic Statistics*, 15(3), pp. 345-353.

[21] Fisher, Jonas (2006): "The Dynamic Effects of Neutral and Investment-Specifc Technology Shocks?" *Journal of Political Economy*, 114, pp. 413-451

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

[23] Furlanetto, Francesco, Ørjan Robstad, Pål Ulvedal, and Antoine Lepetit (2020): "Estimating Hysteresis Effects". Norges Bank Working Paper 2020-13.

[24] Galí, Jordi (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89(1), pp. 249-271.

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

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

[27] Geweke, John (1996): "Bayesian Reduced Rank Regression in Econometrics". *Journal of Econometrics*, 75(1), pp. 121-146.

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

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

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

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

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

[33] 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.

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

[35] Koop, Gary, Rodney W. Strachan, Herman van Dijk, and Mathias Villani (2006): "Bayesian Approaches to Cointegration". In: K. Patterson and T. Mills (eds.): *The Palgrave Handbook of Theoretical Econometrics*, Palgrave MacMillan.

[36] Koop, Gary, Robert 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.

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

[38] Sims Christopher A., James H. Stock, and Mark M. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots". *Econometrica*, 58(1), pp. 113-144.

[39] Sims, Christopher A., and Tao Zha (2006): "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, 96(1), pp. 54-81.

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

[41] 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.

[42] Wu, Jing Cynthia, and Fan Dora Xia (2017): "Time Varying Lower Bound of Interest Rates in Europe". Manuscript.

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

Table 1: Bootstrapped p-values[a] for Johansen's maximum eigenvalue[b] tests for bivariate systems

Economy	Log GDP and log Consumption	Log GDP and log Investment	Short Rate and Long Rate
United States	9.0e-4	0.0016	3.0e-4
Euro Area	7.0e-4	0.0000	0.0021
United Kingdom	0.0060	6.0e-4	0.0346

[a] Based on 10,000 bootstrap replications of the Cavaliere et al. (2012) procedure. [b] Test of 0 versus 1 cointegration vectors.

Table 2: Bootstrapped p-values for Johansen's maximum eigenvalue tests of h versus $h+1$ cointegration vectors[a]

	Test:				
Economy	*0 versus 1*	*1 versus 2*	*2 versus 3*	*3 versus 4*	*4 versus 5*
	7-variable systems:				
United States	0.087	0.013	0.052	0.635	–
Euro Area	4.0e-4	0.004	0.017	0.008	0.147
United Kingdom	0.006	0.002	0.021	0.052	0.160
	Larger systems including population and labor market variables:				
United States	0.037	0.000	0.017	0.034	0.467
Euro Area	1.0e-4	0.005	0.015	0.065	0.319
United Kingdom	0.013	0.000	3.0e-4	0.144	–

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

Table 3*a* Fractions of draws associated with a positive long-run impact of demand- and supply-side permanent GDP shocks, based on 7-variables models

	United States				*Euro area*				*United Kingdom*	
	Including ZLB		Excluding ZLB		Including ZLB		Excluding ZLB			
	Demand	Supply	Demand	Supply	Demand	Supply	Demand	Supply	Demand	Supply
GDP[a]	0.528	1.000	0.495	1.000	0.511	1.000	0.515	1.000	0.805	1.000
Short rate	0.675	0.299	0.652	0.371	0.553	0.677	0.545	0.675	0.624	0.414
Long rate	0.678	0.210	0.677	0.250	0.557	0.485	0.550	0.462	0.357	0.011
GDP deflator	0.945	0.000	0.896	0.001	0.910	0.000	0.906	0.007	0.556	0.000
Hours	0.571	0.923	0.496	0.960	0.533	0.817	0.500	0.837	0.756	0.978

[a] By construction, the fractions for consumption and investment are identical to those for GDP.

Table 3*b* Fractions of draws associated with a positive long-run impact of demand- and supply-side permanent GDP shocks, based on extended models with population and labor market variables

	United States				*Euro area*				*United Kingdom*	
	Including ZLB		Excluding ZLB		Including ZLB		Excluding ZLB			
	Demand	Supply	Demand	Supply	Demand	Supply	Demand	Supply	Demand	Supply
GDP[a]	0.455	1.000	0.510	1.000	0.626	1.000	0.633	1.000	0.943	1.000
Short rate	0.570	0.656	0.585	0.601	0.604	0.808	0.613	0.811	0.694	0.527
Long rate	0.610	0.476	0.589	0.395	0.601	0.657	0.616	0.654	0.361	0.151
GDP deflator	0.900	0.007	0.885	0.003	0.825	0.012	0.844	0.020	0.491	0.000
Unemployment rate[b]	0.571	0.349	0.512	0.096	0.424	0.547	0.430	0.551	0.233	0.058
Population	0.441	0.980	0.429	0.867					0.546	0.651
Participation rate	0.579	0.814	0.601	0.326					0.775	0.771
Average hours worked	0.358	0.694	0.405	0.882					0.675	0.883
Labor force					0.615	0.810	0.619	0.811		

[a] By construction, the fractions for consumption and investment are identical to those for GDP.
[b] For the U.S., logarithm of unemployment.

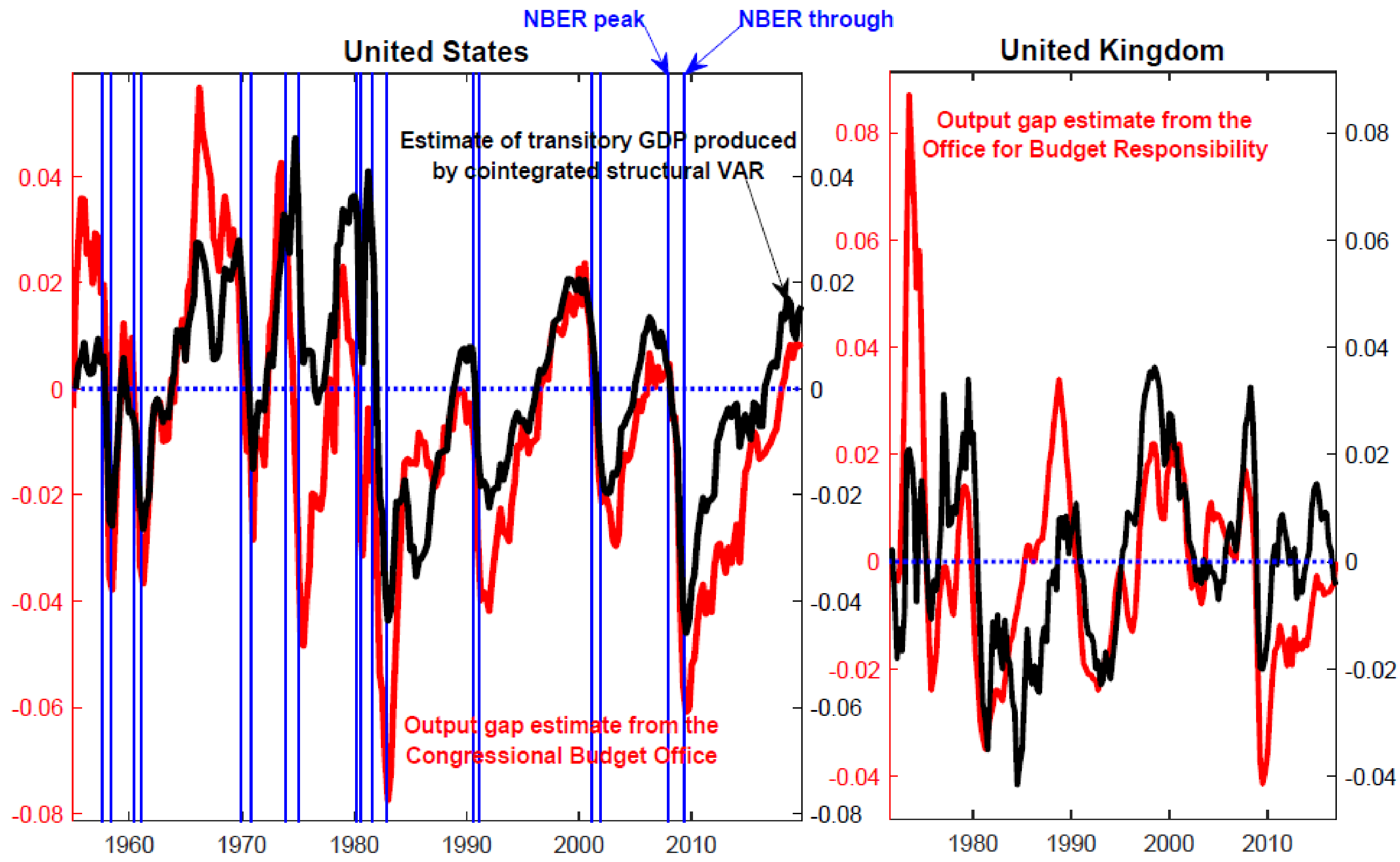

Figure 1 Assessing the plausibility of the permanent-transitory decomposition of GDP produced by the cointegrated SVARs

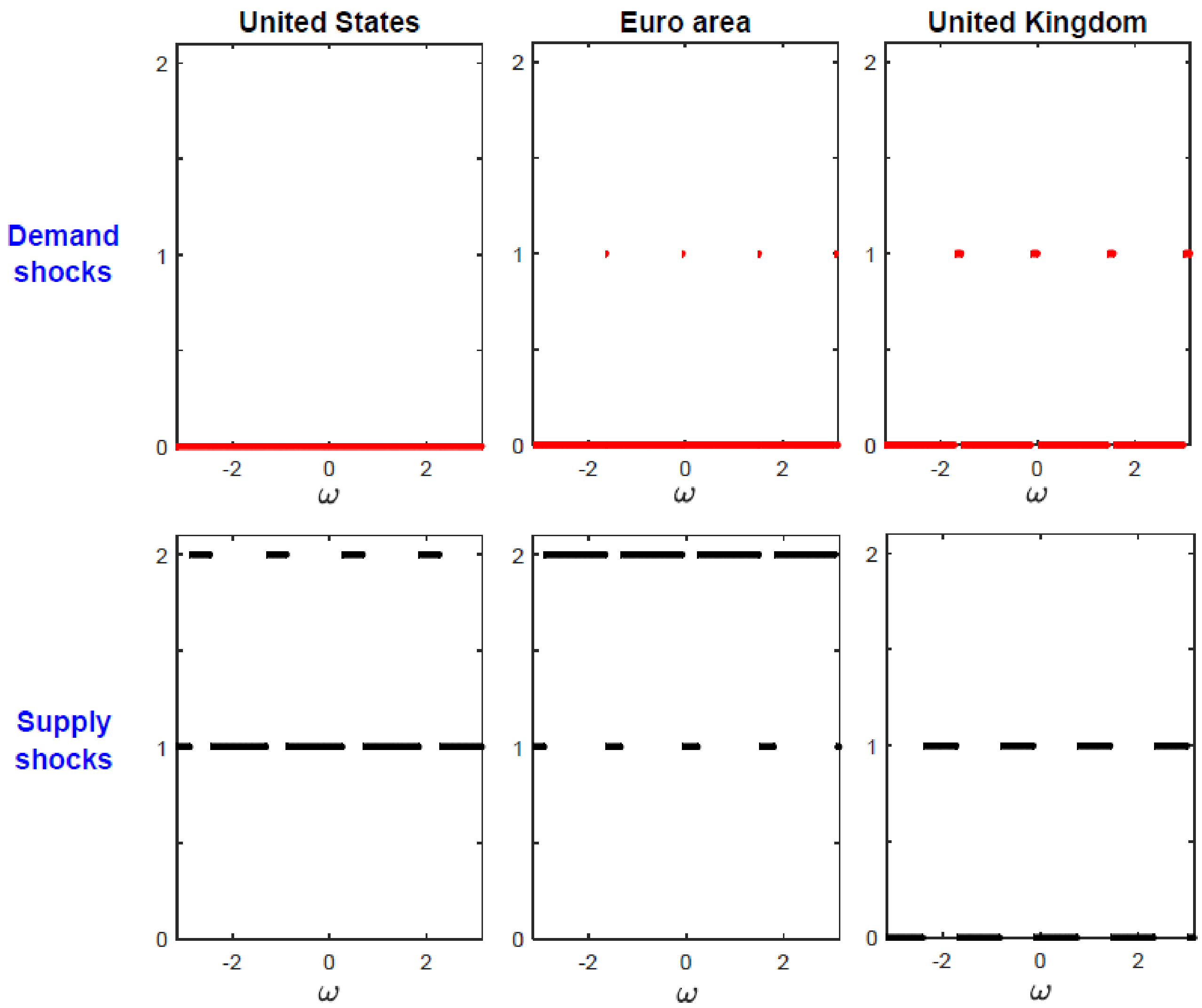

Figure 2 Classical evidence based on the point estimates of the cointegrated VARs: number of identified demand and supply shocks for each rotation angle $\boldsymbol{\omega}$

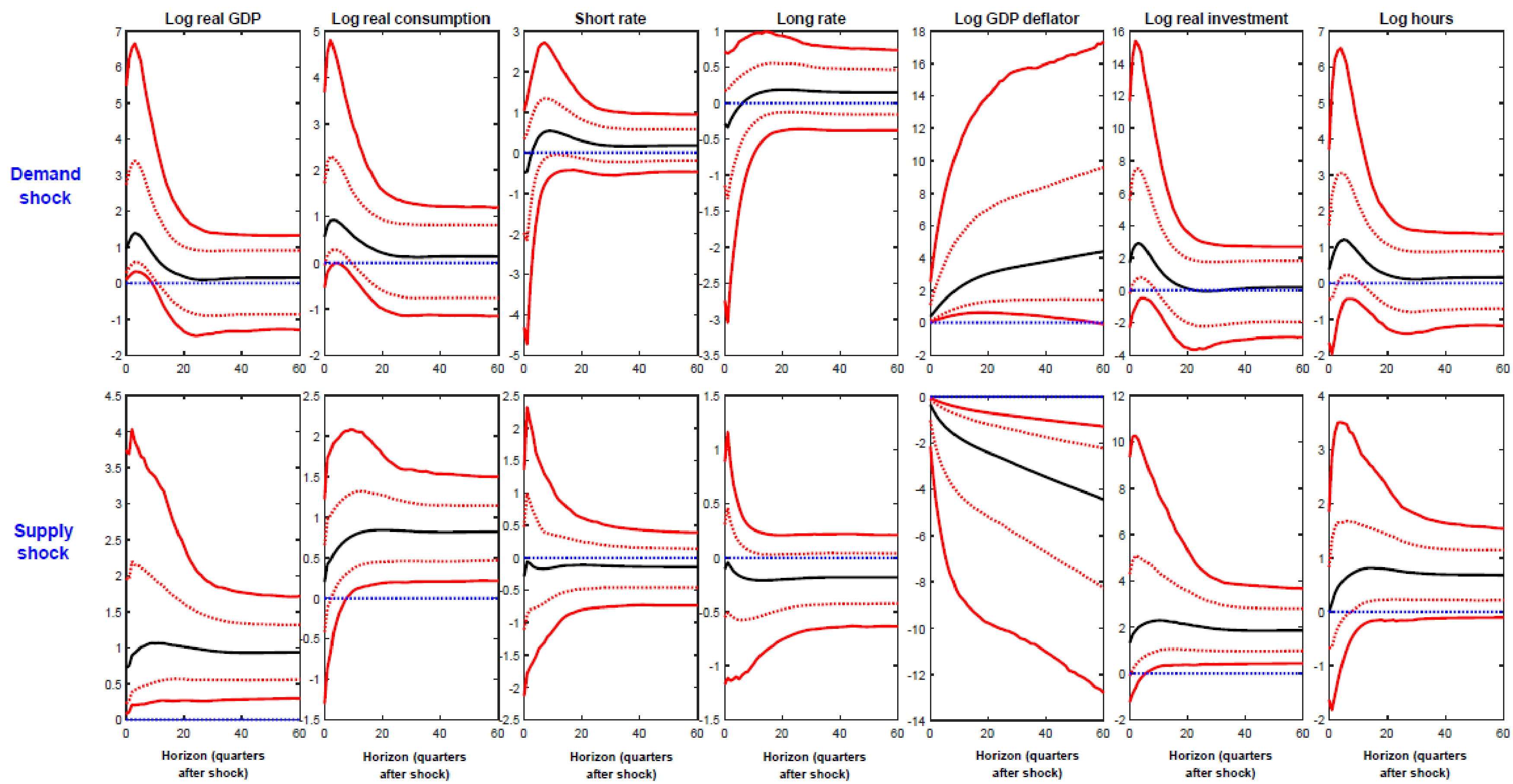

Figure 3 Bayesian evidence for the U.S.: impulse-response functions to demand and supply shocks, without imposing the restriction that the signs of the long-run impacts of demand shocks on GDP are the same for all draws (medians, and 16-84 and 5-95 percentiles)

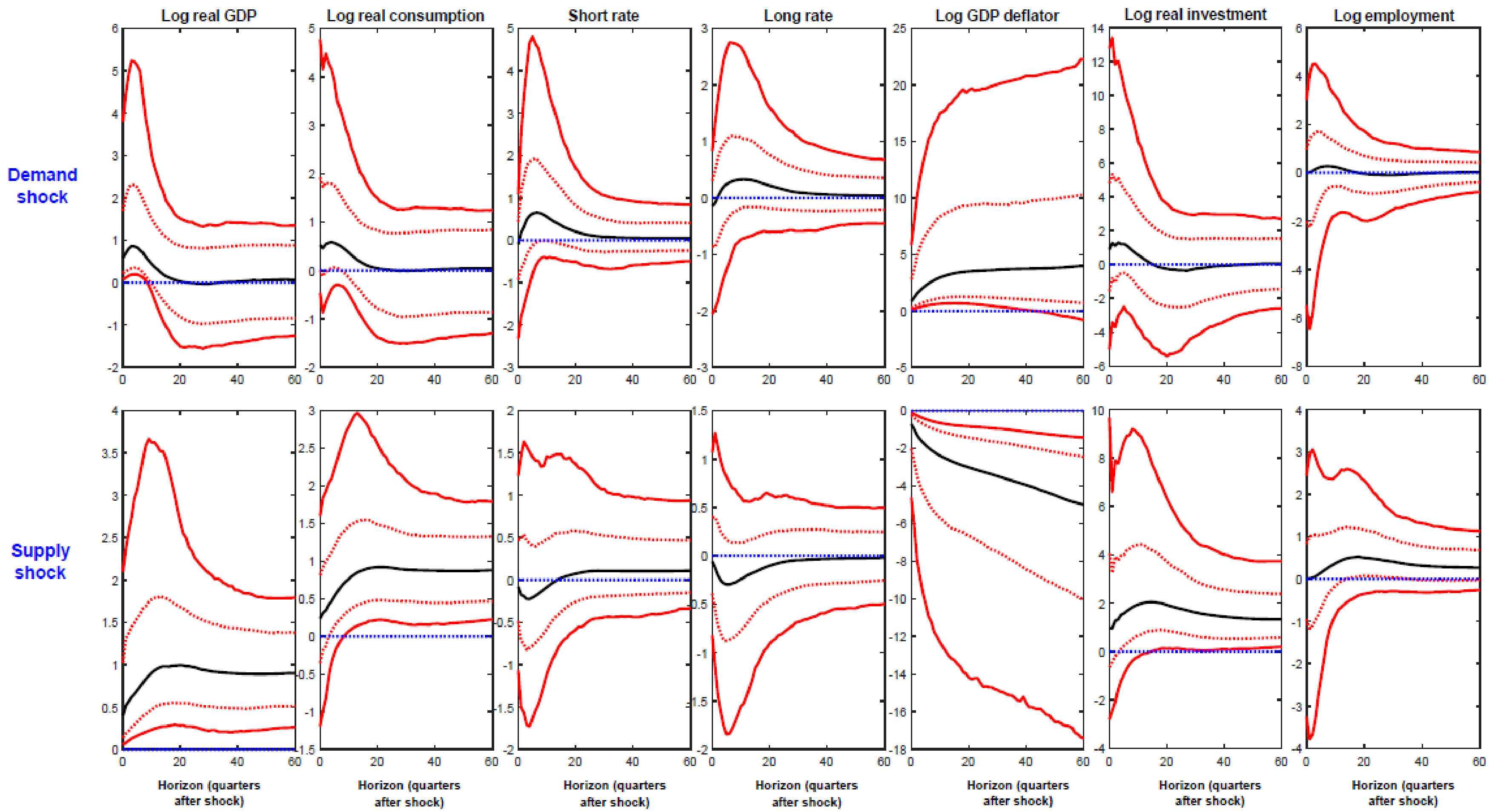

Figure 4 Bayesian evidence for the Euro area: impulse-response functions to demand and supply shocks, without imposing the restriction that the signs of the long-run impacts of demand shocks on GDP are the same for all draws (medians, and 16-84 and 5-95 percentiles)

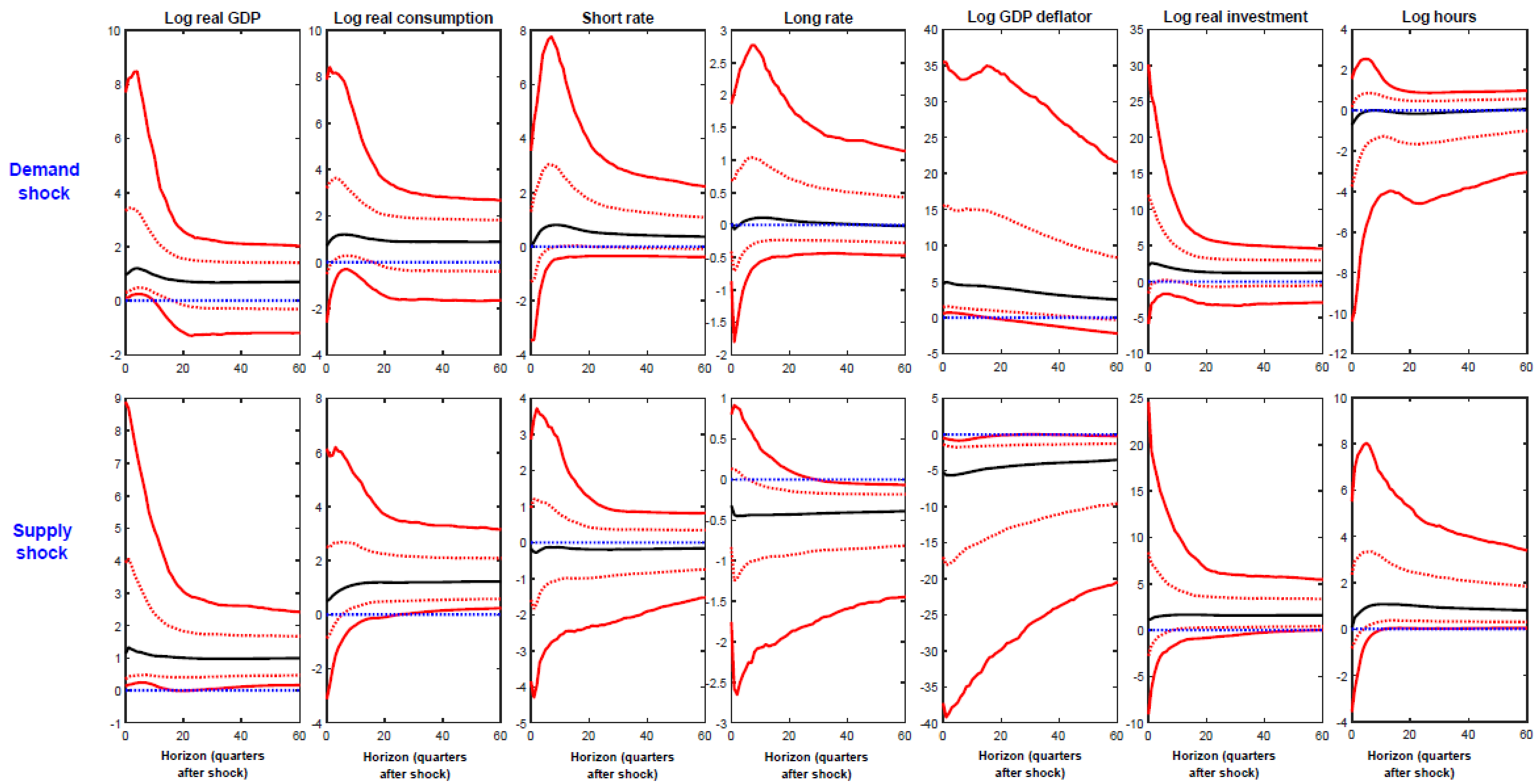

Figure 5 Bayesian evidence for the U.K.: impulse-response functions to demand and supply shocks, without imposing the restriction that the signs of the long-run impacts of demand shocks on GDP are the same for all draws (medians, and 16-84 and 5-95 percentiles)

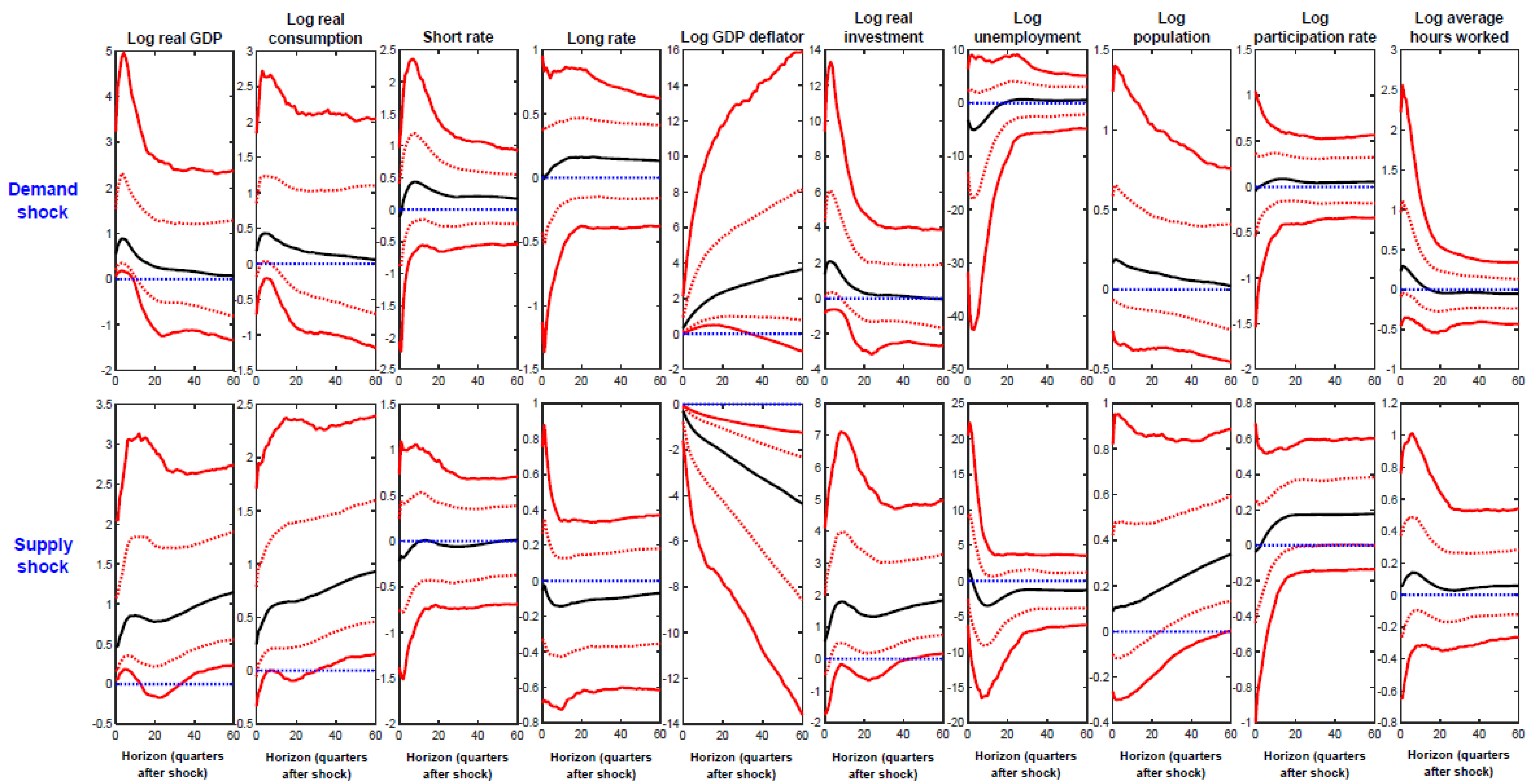

Figure 6 Bayesian evidence for the U.S. based on the expanded model: impulse-response functions to demand and supply shocks, without imposing the restriction that the signs of the long-run impacts of demand shocks on GDP are the same for all draws (medians, and 16-84 and 5-95 percentiles)

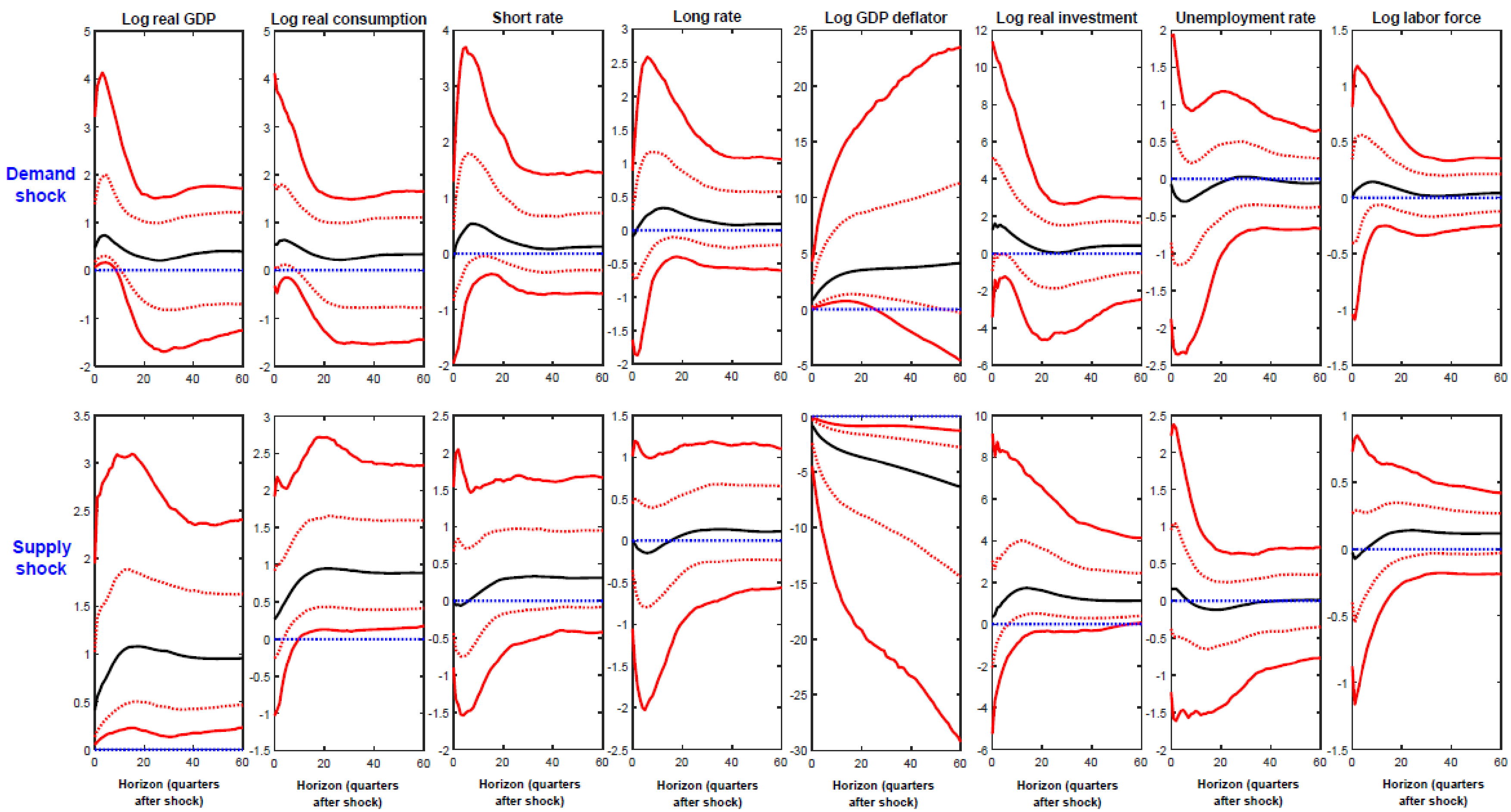

Figure 7 Bayesian evidence for the Euro area based on the expanded model: impulse-response functions to demand and supply shocks, without imposing the restriction that the signs of the long-run impacts of demand shocks on GDP are the same for all draws (medians, and 16-84 and 5-95 percentiles)

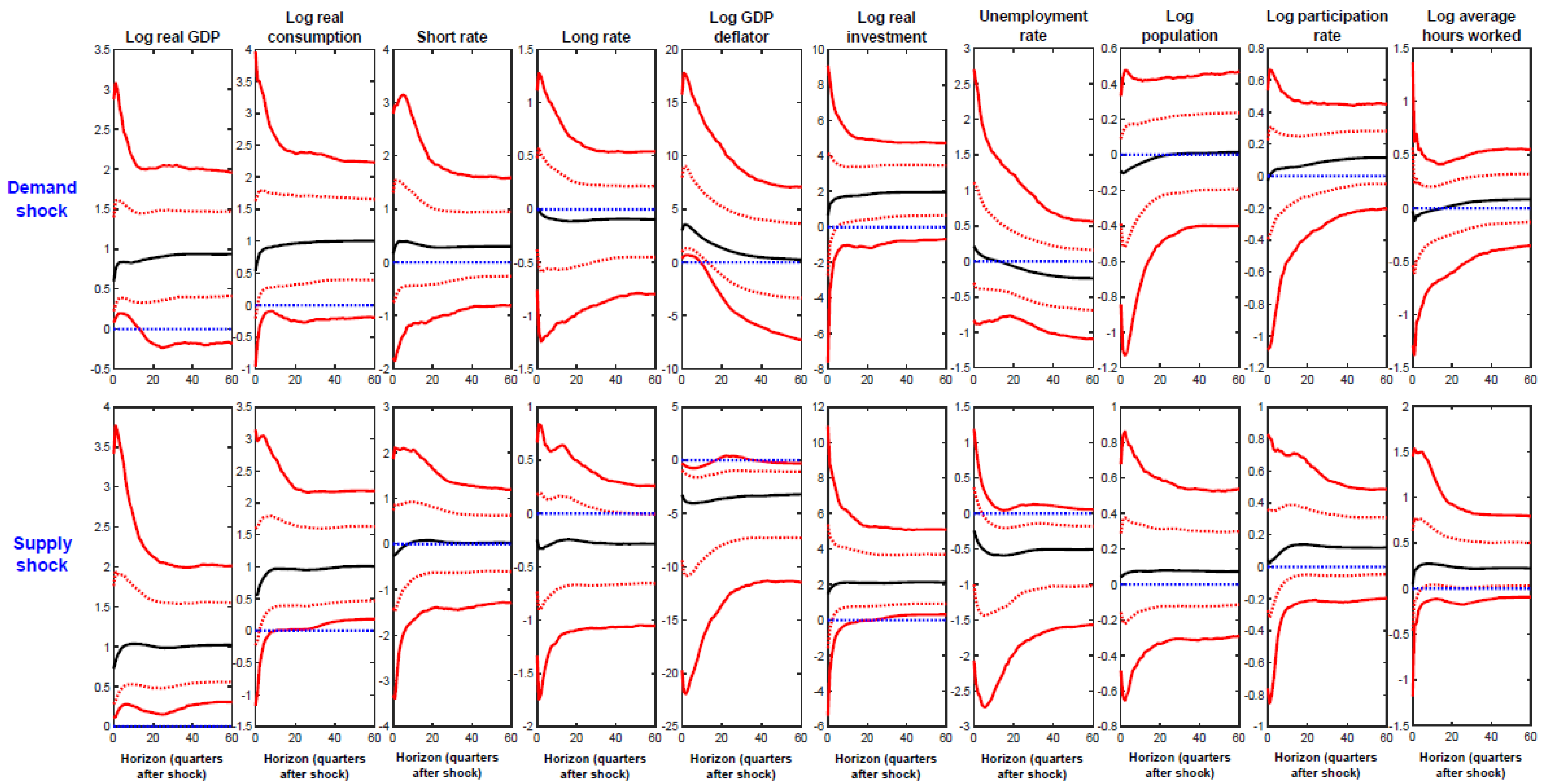

Figure 8 Bayesian evidence for the U.K. based on the expanded model: impulse-response functions to demand and supply shocks, without imposing the restriction that the signs of the long-run impacts of demand shocks on GDP are the same for all draws (medians, and 16-84 and 5-95 percentiles)