The Long-Run Phillips Curve: A Structural VAR Investigation*

Luca Benati
University of Bern†

Abstract

I use both Classical and Bayesian structural VARs identified based on either long-run restrictions, or a combination of long-run and sign restrictions, to investigate the long-run trade-off between inflation and the unemployment rate in the U.S., the Euro area, the U.K., and Canada over the post-WWII period.

Results based on Classical VARs featuring a single permanent inflation shock do not allow to reject the null hypothesis of a vertical long-run Phillips curve for either country, with both the modes and the medians of the bootstrapped distributions of the long-run impact on unemployment of a one per cent permanent shock to inflation being close to zero. Results based on Bayesian VARs allowing for four permanent inflation shocks, which are sorted out from one another by means of DSGE-based robust sign restrictions, produce a very similar picture. The overall extent of uncertainty is however substantial, especially in the latter case, thus suggesting that the data are compatible with a comparatively wide range of possible slopes of the long-run trade-off.

For all countries, Johansen's cointegration tests point towards the presence of cointegration between either inflation and unemployment, or inflation, unemployment, and a short-term interest rate, with the long-run Phillips trade-off implied by the estimated cointegrating vectors being negative and sizeable. I argue however that this evidence should be discounted, as, conditional on the estimated structural VARs—which, by construction, do not feature cointegration between any variable—Johansen's procedure tends to spuriously detect cointegration a non-negligible, and sometimes large, fraction of the times.

Keywords: Inflation; unemployment; Phillips curve; unit roots; cointegration; Bayesian VARs; structural VARs; long-run restrictions; sign restrictions.

*I wish to thank Fabio Canova and Russell Cooper for helpful discussions, and Harald Uhlig and seminar participants at CREST, the European University Institute, Norges Bank, and the Swiss National Bank's 2012 research conference 'Policy Challenges and Developments in Monetary Economics' for helpful comments. Usual disclaimers apply.

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

1 Introduction

In spite of the central role played by the unemployment-inflation trade-off in shaping the evolution of both macroeconomic thinking[1] and policymaking over the last several decades, surprisingly little econometric work has been devoted to investigating the nature of the long-run trade-off. In particular, as I discuss more extensively below, to the best of my knowledge the only existing investigation of the long-run Phillips trade-off based on structural VAR methods is King and Watson (1994)'s 'revisionist econometric history' of the post-WWII U.S. Phillips curve, which has produced evidence of a negative, and statistically significant long-run trade-off conditional on aggregate demand-side shocks.

In this paper I use both Classical and Bayesian structural VARs identified based on either long-run restrictions, or a combination of long-run and sign restrictions, in order to investigate the long-run trade-off between inflation and the unemployment rate in the United States, the Euro area, the United Kingdom, and Canada over the post-WWII period.

Results based on Classical VARs featuring a single permanent inflation shock do not allow to reject the null hypothesis of a vertical long-run Phillips curve for either country, with both the modes and the medians of the bootstrapped distributions of the long-run impact on unemployment of a one per cent permanent shock to inflation being close to zero. Applying the same identification strategy within a Bayesian context produces results which are numerically very close to those produced by Classical methods, pointing, once again, towards no long-run unemployment-inflation trade-off.

Since, in principle, these results are not incompatible with the notion that *some* of the shocks exerting a permanent impact on inflation may induce a non-zero long-run Phillips trade-off, working within a Bayesian context I then proceed to disentangle permanent inflation shocks into demand- and supply-side ones, by imposing Canova and Paustian's (2011) DSGE-based 'robust sign restrictions' on their impact on the endogenous variables at t=0. Overall, results are qualitatively similar to the one produced by VARs featuring a single permanent inflation shock. In particular,

(*i*) for all countries, and for either shock, the 90%-coverage percentiles of the posterior distributions of the long-run impact on the unemployment rate of a one per cent permanent shock to inflation contain zero, thus implying that the notion of a vertical long-run Phillips curve cannot be rejected at conventional significance levels.

(*ii*) For either shock, both the modes and the medians of the posterior distributions of the long-run impact on unemployment of a one per cent permanent shock to inflation are, in general, close to zero.

An important point to stress, however, is that the overall extent of uncertainty is substantial, thus suggesting that the data are compatible with a comparatively wide range of possible slopes of the long-run trade-off. This is the case both for the Classical and Bayesian VARs featuring a single permanent inflation shock, and, especially,

[1] See in particular Lucas (1972a), Lucas (1972b), and Lucas (1973).

for the Bayesian VARs allowing for multiple shocks exerting a permanent impact on inflation. For the United States, for example, the 90% bootstrapped confidence interval for the estimated long-run impact on the unemployment rate of a one per cent permanent shock to inflation produced by Classical VARs featuring a single permanent inflation shock stretches between -0.56 and 0.15. The key reason for such a comparatively large extent of uncertainy is that the feature of the data we are attempting to estimate pertains to the infinite long run, and, as it is well known—see e.g. Faust and Leeper (1997)—this is bound to produce imprecise estimates, unless the researcher is willing to impose upon the data very strong restrictions (which, in general, is not advisable). In the case of the VARs allowing for four permanent inflation shocks, this problem is compounded by our use of sign restrictions, which, as stressed by Fry and Pagan (2007), are intrinsically 'weak information', and should therefore not be expected to produce strong inference.

For all countries, Johansen's cointegration tests point towards the presence of cointegration between either inflation and unemployment, or inflation, unemployment, and a short-term interest rate, with the long-run Phillips trade-off implied by the estimated cointegrating vectors being negative and sizeable. As I show *via* Monte Carlo, this is not the product of the comparatively short samples I am working with, as the fraction of simulations for which the bootstrapped trace statistic incorrectly rejects the null of no cointegration between two independent random walks at a given significance level ranges between 11.3 and 11.9 per cent at the 10 per cent level; between 5.5 and 6.0 per cent at the 5 per cent level; and between 1.1 and 1.3 per cent at the 1 per cent level, thus pointing towards an excellent performance of the cointegration procedure I am using herein (which largely originates from the bootstrap's ability to effectively take into account of the specific characteristics of the data generation process under investigation). I argue however that this evidence should be discounted, as, conditional on the estimated structural VARs—which, by construction, do not feature cointegration between *any* variable—Johansen's bootstrapped procedure tends to spuriously detect cointegration a non-negligible fraction of the times. For example, for the Euro area and the United Kingdom, conditional on taking the VARs featuring four permanent inflation shocks as data generation processes, the fractions of bootstrapped p-values for Johansen's trace statistic for testing the null of no cointegration between inflation and the unemployment rate which are smaller than 10 per cent are equal to 0.269 and 0.236 per cent respectively. This means that, if the estimated structural VARs were the true data-generation process, Johansen's trace test would incorrectly reject the null of no cointegration between inflation and unemployment at the 10 per cent level about one-fourth of the times.

1.1 Related literature

To the very best of my knowledge, the only existing investigation of the long-run Phillips trade-off based on structural VAR methods is King and Watson (1994)'s

'revisionist econometric history' of the post-WWII U.S. Phillips curve. King and Watson (1994) estimate a bivariate VAR for the first differences of CPI inflation and the unemployment rate for the period 1954-1992, and explore the long-run trade-off induced by aggregate demand-side permanent shocks to inflation based on three alternative identification schemes. Results based on the identification scheme they regard as more reliable (which they label as 'Rational Expectations Monetarist') point towards a negative, statistically significant, and comparatively flat long-run Phillips trade-off for either the full sample period, or the 1970-1992 sub-sample (with point estimates equal to -0.29 and -0.23, respectively), and to a steeper trade-off for the 1954-1969 sub-sample (with a point estimate of -0.47).

Nearly two decades after King and Watson (1994), there are several reasons why it is of interest to reconsider this issue.

First, their finding of a negatively sloped, statistically significant, and comparatively flat long-run Phillips curve has radical implications for the conduct of monetary policy, as it implies that the current consensus, within the central banking community, that there is no long-run trade-off between inflation and economic activity—with its corollary that the central bank should focus on delivering low and stable inflation—is misplaced. Current monetary frameworks have been built around the notion that there is no long-run trade-off which can be exploited by monetary policy: in spite of its strong conceptual appeal, it is important to know whether such a notion is in fact supported by empirical evidence.

Second, in the years since 1994 structural VAR econometrics has seen important developments in terms of identification. When King and Watson wrote, short-run restrictions were still either of the 'inertial' type—that is, based on imposing zeros in the impact matrix of the structural shocks at t=0—or they were based on the notion of 'calibrating' some of these impacts based on information extraneous to the VAR.[2] In recent years, several contributions have highlighted the dangers associated with the former approach,[3] whereas the reliability of the latter crucially hinges, as a matter of logic, on just how credible the numbers the researcher is imposing in the VAR's structural impact matrix truly are. Since imposing a specific number entails making a very strong assumption—implying an extent of knowledge we typically

[2] Indeed, this is how King and Watson (1994) achieved identification in their preferred specification. As stressed by Evans (1994) in his comment on King and Watson (1994),

> '[i]dentification of the supply and demand shocks is achieved by imposing a value for λ [the parameter which determines the impact of demand shocks on the unemployment rate at $t = 0$] a priori in the empirical analysis.'. (See Evans, 1994, p. 222.)

[3] The work of Fabio Canova and his co-authors, in particular (see, first and foremost, Canova and Pina (2005)) has demonstrated that, since inertial restrictions are, in general, incompatible with the structure of general equilibrium models—in the specific sense that, within DSGE models, the impact matrix of the structural shocks at t=0 is, in general 'full', i.e., it has no zero entries—imposing such zeros can lead to dramatically distorted inference, for example 'uncovering' price puzzles which are not in the data generation process.

do not have—an alternative style of identification based on weaker informational requirements might be regarded as preferrable. Several researchers[4] have therefore proposed sign restrictions—that is, restrictions on the signs of the impacts of the structural shocks at t=0, and possibly on their impulse-response functions at longer horizons—as the best (or least bad ...) way of achieving identification based on short-run restrictions. As shown by Canova and Paustian (2011), indeed, DSGE models often imply a robust pattern of signs for the impacts of the structural shocks at t=0 (where 'robust' means that such pattern holds true for alternative sub-classes of DSGE models, and for a wide range of plausible parameters' configurations), which is often sufficient to disentangle the structural shocks from one another. In fact, when seen from the perspective of DSGE models, a specific pattern of signs for the impacts of the structural shocks at t=0 is typically the *only* kind of information we can be reasonably confident about, whereas the specific values taken by such impacts are, in general, much more uncertain, thus raising doubts on the reliability of an approach to identification based on the notion of calibrating such impacts.[5]

Third, King and Watson's analysis was entirely based on a bivariate VAR for the first differences of inflation and the unemployment rate, but, as shown by Evans (1994) in his comment,[6] even based on their identification strategy, evidence based on trivariate VARs was sometimes significantly different, pointing in some cases towards a vertical long-run Phillips curve. This naturally suggests reconsidering the issue based on VARs featuring a broader informational content, in particular about the stance of monetary policy and the state of the business cycle.[7]

The paper is organized as follows. The next section discusses methodological issues related to the estimation, from a vector of time series, of the slope of the long-run unemployment-inflation trade-off. The following two sections discuss the choice of the sample periods, and present results from unit root tests for inflation, the unemployment rate, and the short rate. Section 5 presents results based on either Classical or Bayesian VARs featuring a single permanent inflation shock, whereas Section 6 presents evidence based on Bayesian VARs allowing for four permanent inflation shocks. Section 7 discusses the issue of the possible (in principle) presence of cointegration between either inflation and the unemployment rate, or inflation, the

[4]See in particular Faust (1998), Canova and de Nicolo (2002), and Uhlig (2005).

[5]An approach to identification based on sign restrictions is not without problems of its own. As extensively discussed by Fry and Pagan (2007), in particular, sign restrictions suffers from the shortcoming that they are intrinsically 'weak information', and therefore they should not be expected to produce strong inference.

[6]See Evans (1994, Section 3.2, and in particular the results reported in Figure 2).

[7]For the reason discussed, e.g., by Sargent (1987)—that is: the first-difference filter wipes out most of the variance at the business-cycle frequencies—the fact that, as it is well known, the *level* of the U.S. unemployment rate is highly informative about the state of the U.S. business cycle logically implies that its *first difference* is not. This means that a VAR for the U.S. which, beyond the first-difference of the unemployment rate, does not include other indicators of real economic activity, does not contain strong information about the state of the business cycle.

unemployment rate, and the short rate. Section 8 concludes.

2 Methodological Issues

How should we estimate the slope of the long-run Phillips curve? Before Sargent (1971) and Lucas (1972a), the standard approach, due to Solow (1968) and Tobin (1968), was to estimate the Phillips curve regression

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-1} + \alpha_2 \pi_{t-2} + ... + \alpha_p \pi_{t-p} + \gamma_1 U_{t-1} + \gamma_2 U_{t-2} + ... + \gamma_p U_{t-p} + \epsilon_t \quad (1)$$

—with π_t and U_t being inflation and the unemployment rate, respectively—and to test whether $\alpha_1 + \alpha_2 + ... + \alpha_p = \alpha(1) = 1$. Failure to reject the null hypothesis that $\alpha(1)$=1 was regarded as evidence in favor of the notion of a vertical long-run Phillips curve, whereas an estimate of $\alpha(1)$ significantly smaller than 1 was regarded as pointing towards a negative long-run trade-off.

2.1 Why single-equation methods are misleading

Sargent and Lucas highlighted the fallacy intrinsic to such an approach. As pointed out by Lucas (1972a),

> '[...] the natural rate hypothesis, correctly formulated, has *no* implications for the coefficients of distributed lags Phillips curves, or for any other single-equation expression of the empirical inflation-real output trade-off.'[8]

He went on to further stress that

> '[...] a valid test of the natural rate hypothesis involves a test of a restriction on the parameters across equations of a complete simultaneous equations model. The esistence of a natural rate is thus a *systems property*, like stability or identifiability.'[9]

But if $\alpha(1)$ cannot capture the slope of the long-run Phillips curve, *what* does it capture, in fact? As stressed by Sargent (1971), it captures inflation persistence.[10] As he pointed out (see Sargent, 1971, p. 724),

[8] Emphasis in the original.

[9] Emphasis added.

[10] This can be illustrated *via* the following simple example. Let aggregate supply be given by a Lucas (1972b)-type supply curve incorporating long-run neutrality (that is, a vertical long-run Phillips curve), $y_t = \delta[\pi_t - \pi_{t|t-1}] + u_t$, with π_t being inflation, $\pi_{t|t-1}$ being its rational expectation based on information at time t-1, y_t being the deviation of output from its natural level, and δ being the slope of the short-run Phillips curve. Inflation is postulated to be equal to money growth, μ_t. Finally, money growth is set by the policymaker according to the following AR(1) process,

'[...] even the most casual glance at the price history of the United States makes it clear that the inflation rate has not been a strongly drifting variable. [...] It is not surprising, therefore, that most empirical studies have estimated α to be markedly less than unity. Unfortunately, as usually interpreted, these estimates tell us virtually nothing about the validity of the accelerationist thesis.'

2.2 An empirical illustration based on an estimated DSGE model

How relevant is this problem is practice? Since, as I have previously documented (see Benati (2008)), under inflation targeting regimes inflation persistence has essentially disappeared—to the point that, in most cases, inflation has been, so far, statistically indistinguishable from white noise[11]—by Sargent and Lucas' argument we should logically expect that, by applying the Solow-Tobin approach to data generated under such regimes, we would 'uncover' very flat long-run Phillips curves. In this sub-section I therefore estimate a simple New Keynesian DSGE model featuring a vertical long-run Phillips curve based on data from Sweden, the U.K., and Canada under inflation-targeting, and I show that the application of the Solow-Tobin approach to this data-generating processes (henceforth, DGPs) would lead a researcher to uncover flat long-run Phillips curves for either country, in spite of the fact that, by construction, the model encodes a vertical long-run Phillips curve.

The New Keynesian model I am using is near-identical to the one I estimated in Benati (2008), and is described by the following equations,

$$y_t = \gamma y_{t+1|t} + (1-\gamma)y_{t-1} - \sigma^{-1}(R_t - \pi_{t+1|t}) + \epsilon_{y,t} \tag{2}$$

$$\pi_t = \frac{\beta}{1+\alpha\beta}\pi_{t+1|t} + \frac{\alpha}{1+\alpha\beta}\pi_{t-1} + \kappa y_t + \epsilon_{\pi,t} \tag{3}$$

$\mu_t = \rho_\mu \mu_{t-1} + z_t$, and all of the shocks are postulated to be white noise, with innovation variances respectively equal to σ_u^2 and σ_z^2. The model's solutions for π_t and y_t are given by $\pi_t = \rho_\mu \pi_{t-1} + z_t$ and $y_t = \delta z_t + u_t$, from which the theoretical value of the coefficient on π_{t-1} in the Phillips curve regression of inflation on lags of itself and of the deviation of output from its natural level turns out to be equal to $\hat{\alpha}_{OLS} = \rho_\mu$. This implies that if $\rho_\mu \to 1$, so that money growth and inflation tend to unit root processes, $\hat{\alpha}_{OLS} \to 1$, and the Solow-Tobin approach therefore 'uncovers' the truth of a vertical long-run Phillips curve. However, if inflation is weakly persistent the Solow-Tobin approach will turn out to be misleading. In the extreme case in which $\rho_\mu = 0$, so that money growth and inflation are white noise processes, $\hat{\alpha}_{OLS} = 0$, thus pointing towards a perfectly flat long-run Phillips curve.

[11] Updating Benati's (2008) estimates to 2011Q4, Hansen (1999) 'grid bootstrap' estimates of the sum of the autoregressive coefficients in univariate AR(p) representations for GDP deflator inflation are equal to -0.15 in the U.K., to 0.06 in Sweden, and to 0.13 in Canada, with 90%-coverage bootstrapped confidence intervals equal to [-0.43 0.15], [-0.26 0.38], and [-0.09 0.34] respectively. A unit root in inflation can therefore be strongly rejected, and evidence points instead towards inflation being white noise.

$$R_t = \rho R_{t-1} + (1-\rho)[\phi_\pi \pi_t + \phi_y y_t] + \epsilon_{R,t} \tag{4}$$

where π_t, y_t and R_t are inflation, the output gap, and the nominal rate; γ is the forward-looking component in the intertemporal IS curve; α is price setters' extent of indexation to past inflation, and $\epsilon_{\pi,t}$, $\epsilon_{y,t}$ and $\epsilon_{R,t}$ are white noise disturbances.[12] All of the variables in (2)-(4) are expressed as log-deviations from a non-stochastic steady-state. I set β=1, thus imposing a vertical long-run Phillips curve, and I estimate the model based on Swedish, U.K., and Canadian data from the inflation-targeting regimes. Both the Bayesian methodology and the priors are identical to those I used in Benati (2008), to which the reader is referred to. The VAR representation of the estimated model for Sweden conditional on the median estimates is given by[13]

$$\begin{bmatrix} R_t \\ \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} 0.788 & 0.007 & 0.208 \\ -0.230 & 0.037 & 0.288 \\ -0.147 & -0.001 & 1.004 \end{bmatrix} \begin{bmatrix} R_{t-1} \\ \pi_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} u_{R,t} \\ u_{\pi,t} \\ u_{y,t} \end{bmatrix} \tag{5}$$

Consistent with my previous results on the disappearance of inflation persistence under inflation-targeting, the coefficients on π_{t-1} in the equation for π_t is equal to 0.037. Interpreting the second equation of the VAR as an estimated short-run Phillips curve, by applying the Solow-Tobin approach we can immediately back out the estimate of the slope of the long-run Phillips curve, which is equal to 0.299, in spite of the fact that the underlying DGP features a vertical Phillips curve by construction. Results for the other two countries are qualitatively the same, being equal to 0.306 for the United Kingdom, and to 0.245 for Canada.

These results provide a straighforward explanation for Svensson's (2012) finding of a comparatively flat long-run Phillips trade-off in Sweden under the inflation targeting regime based on the Solow-Tobin approach. Specifically, Svensson (2013) estimates a version of (1) for the period 1997Q4-2011Q4,[14] 'backs out' the estimate of the long-run Phillips trade-off from his short-run estimates, and performs the Solow-Tobin test that $\alpha(1) = 1$. He thus summarizes his results, which point towards a negatively-sloped long-run Phillips curve:

> 'From the estimates in table 2 it follows that the slope of the long-run Phillips curve is about 0.8. With a standard error of 0.19, it is fairly precisely estimated. A 95-percent confidence interval of the slope is the interval from 0.43 to 1.18. Clearly, the hypothesis of a vertical Phillips curve, an infinite slope, can be rejected.'

[12] The assumption that all of the structural disturbances are white noise is made uniquely for the sake of simplicity, it can be easily relaxed, and it plays no role in my results.

[13] The covariance matrix of the VAR's reduced-form innovations is not reported for the sake of simplicity. By the same token, I do not report the corresponding VAR representations of the estimated DSGE models for the other three countries, but they are available upon request.

[14] See Svensson (2013, Table 2).

In the light of the previous discussion of Sargent and Lucas' criticism of the Solow-Tobin approach, and of the fact that, under inflation-targeting, inflation in Sweden is, today, very close to white noise[15], Svensson's rejection of a vertical long-run Phillips curve based on the test that $\alpha(1) = 1$ is to be expected. At the same time, however, as pointed out by Sargent (1971) in the previous quotation, this evidence is uninformative about the authentic slope of Sweden's long-run Phillips curve.

2.3 King and Watson's (1994) alternative approach

Since the results produced by single-equation methods—especially when applied to data generated by monetary regimes which cause inflation to be strongly mean-reverting, such as inflation targeting, European Monetary Union (henceforth, EMU), and the Swiss 'new monetary policy concept'—are unreliable, in this paper I follow King and Watson (1994), and I use structural VAR methods to estimate the permanent impact (if any) on the unemployment rate of permanent shocks to inflation. As stressed by King and Watson (1994, p. 177), indeed, if inflation has a unit root the data are in fact informative about the slope of the long-run Phillips curve, since they do contain the relevant 'experiment':

> 'When inflation is stationary, [...] as stressed by Lucas and Sargent, the relevant experiment—permanent changes in the rate of inflation—are absent from the inflation data and so the long-run Phillips trade-off could not be estimated [...]. In the unit root case, by contrast, [...] the relevant experiments are present in the data: variation in $\bar{\pi}_t$ [i.e.: the permanent component of inflation] allows the long-run slope to be determined.'

As a simple illustration of King and Watson's point, suppose that the economy is described by a Lucas (1972b)-type supply curve possibly incorporating a long-run non-neutrality (that is, a non vertical long-run Phillips curve),

$$y_t = \delta[\pi_t - \beta\pi_{t|t-1}] + u_t, \tag{6}$$

with π_t being inflation, $\pi_{t|t-1}$ being its rational expectation based on information at time t-1, y_t being the deviation of output from its natural level, δ being the slope of the short-run Phillips curve, and β being the parameter upon which the slope of the long-run Phillips curve crucially depends. Expression (6) implies that the long-run impact on y_t of a permanent change in π_t is equal to δ(1-β). Finally, let's assume that inflation is a pure random walk, $\pi_t = \pi_{t-1} + z_t$. The VAR representation of the model is given by

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} 0 & \delta(1\text{-}\beta) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_t \\ z_t \end{bmatrix} \tag{7}$$

[15] See footnote 11.

from which the long-run impact on y_t of a permanent shock to π_t at $t = 0$ can be immediately computed as

$$\lim_{t\to\infty} \frac{\partial y_t}{\partial z_0} = \delta(1-\beta), \tag{8}$$

which is precisely the slope of the long-run Phillips curve.

2.4 Why a DSGE-based approach is not a viable option

In principle, an alternative approach would be to estimate a (DSGE) structural macroeconomic model in which the slope of the long-run trade-off is encoded in one or more of the model's structural parameters, and then to test whether, conditional on the model's estimates, it is possible to reject the null hypothesis that the long-run Phillips curve is vertical. There are two key drawbacks to this approach.

First, and least importantly, as a matter of logic results would be, in general, model-dependent (that is: they would be conditional on the specific structural model which is being estimated), and they could not be regarded as possessing a general validity.

Second—and this is key—this approach crucially rests on the assumption that the parameters of the estimated model are truly structural in the sense of Lucas (1976), and, in particular, that they are invariant to changes in the characteristics of the inflationary environment—first and foremost, equilibrium inflation. This is key because *if and only if* this assumption is satisfied we can be confident that the long-run Phillips trade-off extracted from a structural macroeconomic model estimated based on a sample period during which average inflation has been equal to, say, two per cent, will hold at higher equilibrium inflation rates. In recent years, however, evidence has been accumulating that several DSGE models' supposedly structural parameters are in fact, most likely, not structural in the sense of Lucas (1976). Benati (2008), for example, has shown that the indexation parameter in New Keynesian backward- and forward-looking Phillips curves changes systematically with the monetary regime, whereas the recent work of Fernandez-Villaverde and Rubio-Ramirez (and their co-authors) has documented instability in several structural parameters. Fernández-Villaverde and Rubio-Ramírez (2008), in particular, have shown that the extent of price stickiness negatively and systematically co-moves with trend inflation. All of this implies that the notion of extracting, from an estimated DSGE model, a long-run Phillips trade-off which can safely be regarded as invariant to changes in equilibrium inflation rests, most likely, on a leap of faith, *unless* we will be able, in the future, to effectively model the way in which those specific DSGE models' features which crucially determine the slope of the long-run trade-off (first and foremost, price stickiness and the extent of backward-looking indexation) co-move systematically with equilibrium inflation.

The attractiveness of King and Watson's approach, on the other hand, is that the *only* assumption it crucially hinges upon is that inflation has a unit root, and,

conditional on this assumption, the slope of the long-run trade-off can be investigated without possessing any detailed knowledge of the underlying structural model of the economy.[16]

2.5 Why it is so difficult to find the 'right' data to estimate the slope of the long-run Phillips curve

The fact that, for a researcher to be able to econometrically identify the slope of the long-run Phillips curve, inflation must contain a unit root, highlights why it is so difficult to find the 'right' data to explore this issue. As it has been extensively documented, e.g., in Barsky (1987) and Benati (2008), under metallic standards (that is, before WWI) inflation had consistently been indistinguishable from white noise, and, if anything, it had been weakly negatively serially correlated. By the same token, during the period between the outbreak of WWI and the early part of the post-WWII period inflation had exhibited some persistence, but it had still been most likely stationary. Further, although for sample periods which are dominated by the Great Inflation episode evidence of a unit root in inflation is sometimes (but, as I discuss in Section 4 below, not always) strong, following the disinflations of the early 1980s inflation persistence has dramatically decreased, so that for sample periods starting in mid-1980s the null of a unit root can once again be strongly rejected at conventional significance levels. Even more ominously for the possibility of identifying the slope of the long-run Phillips curve, under inflation targeting regimes, EMU, and the Swiss 'new monetary policy concept' inflation is, today, close to white noise. This means that, historically, *only* sample periods dominated by the Great Inflation episode can legitimately be used to investigate the slope of the long-run Phillips curve, simply because such sample are the only ones containing the relevant 'experiment' mentioned by King and Watson (1994), whereas data generated under either metallic standards or inflation targeting regimes are exactly the kind of data that cannot be used for this purpose.

2.6 The long-run Phillips trade-off in a low-inflation environment

A conceptually related issue pertains to the slope of the long-run Phillips curve in a low-inflation environment. Several papers—from Akerlof, Dickens, and Perry (1996),

[16] This point is also made by Fisher and Seater (1993):

> '[...] absent knowledge of the underlying structure, the consequences of an event cannot be inferred if the event has not occurred. In order for inferences regarding [long-run neutrality (long-run superneutrality)] to be drawn from a reduced form, the data must contain permanent stochastic changes in the level (growth rate) of the money supply.'.

to Benigno and Ricci (2011)—have argued that, due to the presence of downward nominal wage rigidity, the long-run Phillips trade-off may become quite significantly flat within a low-inflation environment, even if it is very steep, or even vertical, at higher inflation rates. Unfortunately, a proper econometric investigation of this hypothesis is, and will most likely remain, impossible.

First, for the reasons I discussed in section 2.4, extracting the long-run Phillips trade-off from an estimated structural (DSGE) macroeconomic model is (at least, currently) not a feasible option.

Second, for a pure time-series approach to be feasible, inflation should exhibit, over the sample period, two characteristics which, historically, have been mutually exclusive: it should be 'low' (which, in practice, means that it should be bounded between, say, minus two and two per cent), and it should contain a unit root. The problem is that although it is easy to find periods during which either (i) inflation has been low, stable, and the null of a unit root can be strongly rejected (e.g., inflation targeting regimes), or (ii) inflation has been high, volatile, and the null of a unit root cannot be rejected (for several countries, the Great Inflation episode and the surrounding years), historically there has not been a single period during which inflation has been low, stable, and it has exhibited a unit root behaviour.

As a result, the notion that the long-run Phillips curve may become flat within a low-inflation environment will most likely remain a theoretically compelling, but empirically unproven conjecture.

3 Choosing the Sample Periods

Since the null hypothesis of a unit root in inflation can be strongly rejected for sample periods which are not dominated by the Great Inflation episode, in what follows I consider the following sample periods: for the Euro area, the period 1970Q1-1998Q4;[17] for the United Kingdom, the period 1972Q2-1992Q3;[18] for Canada, the period 1961Q2-1990Q4;[19] for Sweden the period 1970Q1-1992Q4;[20] for Australia, the period 1969Q3-1994Q2;[21] and for Japan the period following the collapse of Bret-

[17] EMU started in January 1999, whereas Euro area data are only available starting from 1970Q1.

[18] June 23, 1972 marks the floating of the pound *vis-à-vis* the U.S. dollar, whereas inflation-targeting was introduced on October 8, 1992. As shown by Benati (2008), before the June 1972 floating of the pound U.K. inflation exhibited quite significantly lower persistence.

[19] Canada introduced inflation targeting in February 1991. 1961Q1 is when Canadian national account data first become available.

[20] The unemployment rate is only available since 1970Q1, whereas inflation targeting was introduced in January 1993.

[21] The short rate is available since 1969Q3. In dating the start of Australia's inflation targeting regime, for which, different from other countries, there never was a clear announcement on the part of the government, we follow Bernanke, Laubach, Mishkin, and Posen (1999).

ton Woods.[22] Finally, for the United States, as I discuss below, I consider several alternative sample periods.

4 Results from Unit Root Tests

Table 1 reports, for either country, bootstrapped p-values for augmented Dickey-Fuller (henceforth, ADF) tests for inflation, the unemployment rate, and a short-term interest rate. For either series, p-values have been computed by bootstrapping 10,000 times estimated ARIMA(p,1,0) processes. In all cases, the bootstrapped processes are of length equal to the series under investigation.[23] As for the lag order, since, as it is well known, results from unit root tests may be sensitive to the specific lag order which is being used, for reasons of robustness I consider three alternative lag orders, one, two, and four.[24]

Starting from inflation, the null of a unit root cannot be rejected based on either the GDP deflator or the CPI for the Euro area, the United Kingdom, and Canada. For Australia it cannot be rejected based on the CPI, but it can be strongly rejected based on the GDP deflator. By the same token, evidence for Japan is not strong, with two p-values just above 5 per cent based on the GDP deflator, and it is likewise weak for Sweden based on the CPI.[25] Because of such a comparatively weak evidence of a unit root in inflation for either Japan, Sweden, or Australia, in what follows I exclude the three countries from the analysis, and I uniquely focus on the remaining four. Finally, for the United States evidence of a unit root based on the entire sample period since January 1959 is weak, with the p-values being consistently below 5 per cent based on the CPI, and being below 5 per cent for p=1 and p=2 based on the GDP deflator,[26] whereas for the period since August 1979, when Paul Volcker became Federal Reserve Chairman, a unit root in inflation can be overwhelmingly rejected based on either price index.[27] This suggests that results for the U.S. based on the full

[22] After August 1971, the Japanese government has not explicitly introduced any new monetary regime, and has instead relied, most of the time, on a generic committment to price stability.

[23] To be precise, letting T be the length of the series under investigation, we bootstrap an artificial series of length T+100, and we then discard the first 100 observations in order to eliminate dependence on initial conditions.

[24] For the United States, for which I use monthly data, the corresponding lag orders are three, six, and twelve.

[25] The sample period for the GDP deflator, starting in 1980Q1, is too short to allow us to make any reliable inference, and I therefore just decided to ignore it.

[26] Whereas the monthly interpolated U.S. GDP deflator from Mark Watson's website is available only starting from January 1959, the CPI is available since January 1947. Bootstrapped p-values for ADF tests based on three, six, and twelve lags for CPI inflation since January 1947 are equal to 0.000, 0.000, and 0.001 respectively, thus pointing towards an overwhelming rejection of the null of a unit root.

[27] This is consistent with a large body of research documenting a decrease in the persistence of U.S. inflation during the period following the Great Inflation episode. Stock and Watson (2007), for example, identify a hump-shaped pattern in the evolution of the innovation variance of the

sample period should be viewed with suspicion, and in what follows I will therefore focus on the pre-Volcker sample period, for which evidence of a unit root in inflation is instead strong based on either the CPI or the GDP deflator.[28]

Evidence of a unit root in the unemployment rate is strong for either the Euro area, the United Kingdom, or Canada. For the United States results from the ADF tests are compatible with the notion that the unemployment rate contains a unit root for both sub-periods, whereas evidence is weak for the full sample starting in January 1959. Further, for the entire post-WWII sample period (that is, since January 1948), bootstrapped p-values for ADF tests with three, six, and twelve lags are equal to 0.051, 0.001, and 0.031 respectively, thus pointing towards a quite strong rejection of the null of a unit root. Once again, since I want to give the U.S. long-run Phillips trade-off its best chance of manifesting itself, in what follows I focus on the pre-Volcker sample period.

As previously mentioned, one possible limitation of King and Watson's (1994) analysis was its bivariate nature, and its eschewing of information on the monetary policy stance and the state of the business cycle. Beyond the first differences of inflation and the unemployment rate, in what follows I therefore also include, in the VAR, an output gap measure, the consumption/GDP ratio, a long-short spread, and either the level or the first difference of a short-term interest rate. As for the short rate, evidence of a unit root is strong for all countries, and based on either lag order, with the single exception of the United Kingdom for which p-values range between 0.018 and 0.024. In what follows, the short rate will therefore enter the VAR in levels for the United Kingdom, and in first differences for all other countries.

5 Results Based on VARs Featuring a Single Permanent Inflation Shock

I start by considering either Classical or Bayesian VARs featuring a single permanent inflation shock. For either country I estimate the VAR(p) model

$$Y_t = B_0 + B_1 Y_{t-1} + ... + B_p Y_{t-p} + u_t \tag{9}$$

where $Y_t \equiv [y_t, \Delta\pi_t, \Delta U_t, \Delta R_t, S_t]'$ for the Euro area and Canada, and $Y_t \equiv [y_t, \Delta\pi_t, \Delta U_t, R_t, CY_t, S_t]'$ for the United Kingdom, with y_t, π_t, U_t, R_t, CY_t, and S_t being the output gap, inflation, the unemployment rate, the short rate, the consumption/GDP ratio, and the long-short spread, respectively. These variables effectively

permanent component of U.S. inflation, with a peak around the time of the Great Inflation, and a significant fall over subsequent years.

[28] My choice of focusing on the pre-Volcker sample period is motivated by a sizeable body of literature—see, in particular, Clarida, Gali, and Gertler (2000) and Lubik and Schorfheide (2004)—documenting how, before Volcker, U.S. monetary policy was not sufficiently aggressive in tackling inflation, thus resulting in comparatively large, and very highly persistent inflation fluctuations.

expand the informational content of King and Watson's original bivariate VAR along several dimensions. The short rate and, to a lesser extent, the long-short spread contain information about the monetary policy stance. As recently shown by Kurmann and Otrok (2013), the long-short spread possesses a remarkably strong informational content for future movements in technology. Finally, the consumption/GDP ratio and the output gap estimate can be regarded as two 'noisy estimates' of the state of the business cycle.[29]

As I discuss more extensively in Section 7 below, the United States is the only country for which Johansen's trace test produces evidence of cointegration between the short rate and inflation, with the bootstrapped, bias-corrected estimate of the second element of the cointegrating vector being equal to -1.099, and with a p-value of 0.0872 for rejecting the null hypothesis that it is equal to -1. For the United States I therefore set $Y_t \equiv [y_t,\ \Delta\pi_t,\ \Delta U_t,\ CR_t,\ CY_t,\ S_t]'$, where CR_t is the cointegrating residual produced by Johansen's procedure (that is, $CR_t = R_t$ - $1.0989{\cdot}\pi_t$). *All* of the SVAR-based results for the United States contained in this paper, however, are robust to the inclusion, in the VARs, of the *ex post* real rate R_t-π_t,[30] instead of the cointegrating residual CR_t, in the specific sense that the results produced by the two alternative specifications are qualitatively the same, and numerically very close.[31] For the Euro area, Canada, and the United Kingdom I set the lag order to p=4. For the United States, for which I work at the monthly frequency, I set it to p=12. I identify the permanent inflation shock based on the restriction that it is the only shock impacting upon inflation in the infinite long run.[32] This is tantamount to imposing that in the second row of the long-run impact matrix $C_\infty \equiv [I_N - B(1)]^{-1}A_0$, where A_0 is the structural shocks' impact matrix, all of the elements except the first one are equal to zero, so that the first element of ϵ_t=$A_0^{-1}u_t$ is the permanent inflation shock.[33]

[29] As for the consumption/GDP ratio, see the extended discussion in Cochrane (1994).

[30] Bootstrapped p-values for ADF tests with 3, 6, and 12 lags for the the *ex post* real rate are equal to 0.000, 0.006, and 0.082, respectively, thus pointing towards its stationarity at the 10 per cent level.

[31] This alternative set of results is available upon request. In fact, given that the p-value for rejecting the null hypothesis that the second element of the cointegrating vector is -1 is equal to 0.0872, it is not entirely clear which of the two specifications should be preferred.

[32] Another possibility would be, in the spirit of Barsky and Sims (2011), to identify the single shock which explains the largest fraction of inflation variance at (say) the 10-year horizon. I leave this alternative identification strategy to future work.

[33] Roberts (1993) estimated a VAR for the U.S. for the first differences of inflation and the unemployment rate, and the logarithm of M_2 velocity, and identified two permanent shocks to inflation and the unemployment rate, respectively, by imposing the restrictions that they are the only shocks impacting either variable in the long run. A key difference between Roberts (1993) and the present work is that he imposed orthogonality between the two shocks (that is, he imposed a vertical long-run Phillips curve). See also Bullard and Keating (1995), who, working with bivariate VARs for the first difference of inflation and either output growth, or its first difference, use the same restriction used herein in order to identify permanent inflation shocks.

5.1 Classical estimation

Working within a Classical context, I estimate (9) *via* OLS based on the standard formulas found in Luetkepohl (1991). Concerning the estimation of the impact matrix of the structural shocks, as extensively discussed by Christiano, Eichenbaum, and Vigfusson (2006), reliably estimating C_∞ requires a good estimate of the spectral density of X_t at the frequency zero, $S_X(0) = C_\infty C_\infty'$, an object which VARs, given their focus on fitting the short-run dynamics of the data, should not in general be expected to capture well. Following Christiano *et al.* (2006) I therefore consider, beyond the standard estimator of $S_X(0)$ produced by the VAR—that is, $\hat{S}_X^{VAR}(0) = [I_N - \hat{B}(1)]^{-1}\hat{V}\{[I_N - \hat{B}(1)]^{-1}\}'$, where $\hat{V}$ is the estimated covariance matrix of the VAR's reduced-form innovations—the Bartlett estimator of the spectral density matrix,[34]

$$\hat{S}_X(\omega) = \frac{1}{2\pi}\left[\hat{\Gamma}_0 + \sum_{k=1}^{q}\left(1\text{-}\frac{k}{q+1}\right)\left(\hat{\Gamma}_k e^{-i\omega k} + \hat{\Gamma}_k' e^{i\omega k}\right)\right] \tag{10}$$

where $\hat{\Gamma}_k$ is the k-th autocovariance matrix of X_t.

I use standard bootstrapping techniques in order to both bias-correct the estimated long-run impacts of the permanent inflation shock as in Kilian (1998),[35] and characterise the extent of uncertainty around the bias-corrected estimates. Specifically, I bootstrap the estimated reduced-form VAR 10,000 times, and based on each bootstrap replication I estimate a VAR(p); I impose the same identification scheme I imposed on the VAR estimated based on the actual data; and I compute the implied long-run impact on the unemployment rate of the permanent inflation shock. Then, I use such bootstrapped distributions, first, to bias-correct the simple estimate of the long-run impact; and second, to characterise the extent of uncertainty surrounding such bias-corrected estimates, by simply rescaling the original bootstrapped distribution in such a way that its median be equal to the bias-corrected estimate.[36]

5.2 Bayesian estimation

Working within a Bayesian context, I estimate (9) as in Uhlig (1998) and Uhlig (2005). Specifically, I exactly follow Uhlig (1998, 2005) in terms of both distributional assumptions—the distributions for the VAR's coefficients and their 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

[34] See Hamilton (1994).

[35] So, to be clear, the only difference between Kilian's (1998) paper and the present wiork is that he dealt with the bias-correction of IRFs, whereas I am here bias-correcting the long-run impacts of the structural shocks.

[36] In general, however, the extent of the bias is very small, so that, in the end, bias-correcting does not make a material difference to the results.

Appendix B of Uhlig (2005). Finally, for each estimated VAR I consider 10,000 draws from the posterior distribution of the VAR's coefficients and covariance matrix of innovations (the draws are computed exactly as in Uhlig (1998, 2005)).

5.3 Evidence

Figure 2 shows, for either country, the distributions of the long-run impacts on the unemployment rate of a one per cent permanent shock to inflation,[37] together with the distributions of the fraction of the permanent component of the unemployment rate which is explained by the permanent inflation shock. Table 2 reports the modes, the medians, and the 90%-coverage percentiles of the distributions of the long-run impact on the unemployment rate of a one per cent permanent shock to inflation, together with the fractions of the mass of the distributions for which the impact is estimated to be negative.

For either country, results are very close both across methodology (Classical *versus* Bayesian) and, within the Classical approach, for the two alternative estimators of the spectral density matrix of the VAR at the frequency zero, and for neither country they allow to reject the 'natural' null hypothesis of a vertical long-run Phillips curve. Specifically,

first, for either country, both the modes and the medians of the distributions of the estimated long-run impacts on the unemployment rate of a one per cent permanent inflation shock are quite close to zero, with the modes ranging between -0.105 and -0.216 for the United States, -0.040 and 0.056 for the Euro area, -0.104 and -0.072 for the U.K., and -0.121 and -0.088 for Canada. Taking, just for the sake of the argument, the modes of the distributions of the long-run impacts produced by the Bartlett estimator of the spectral density matrix as the 'benchmark' estimates, the numbers reported in Table 1 imply that a permanent increase in inflation by 10 percentage points would permanently decrease the unemployment rate by just *1.05 per cent* in the United States, by 0.88 per cent in the United Kingdom, and by 1.21 per cent in Canada, whereas in the Euro area it would increase it by 0.56 per cent. Even if one is willing to believe that these point estimates truly capture the authentic unemployment-inflation trade-offs out there, these are hardly trade-offs which might induce policymakers to 'try to play the Phillips curve' ...

Second, for all countries, all of the 90 per cent confidence intervals contain zero.

Third, the fractions of the mass of the distributions for which the impact is estimated to be negative are uniformly greater than standard levels of statistical significance, thus highlighting how the null hypothesis of a vertical long-run Phillips curve cannot be rejected at conventional levels.

Fourth, for either country, and based on either methodology, the distributions of

[37] Following Evans (1994), throughout the entire paper we focus on the long-run impact on the unemployment rate of a one per cent permanent shock to inflation—that is, the inverse of the slope of the long-run Phillips curve—rather than the slope itself.

the fraction of the permanent component of the unemployment rate which is explained by the permanent inflation shock are clustered towards zero, thus providing additional evidence in favor of the notion that the long-run Phillips curve is vertical.[38]

6 Results Based on Bayesian VARs Allowing for Four Permanent Inflation Shocks

Since, in principle, the results discussed in the previous section are not incompatible with the notion that *some* of the shocks exerting a permanent impact on inflation may induce a non-zero long-run Phillips trade-off, working within a Bayesian context I then proceed to disentangle permanent inflation shocks into demand- and supply-side ones, by imposing Canova and Paustian's (2011) DSGE-based 'robust sign restrictions' on their impact on the endogenous variables at t=0.[39]

6.1 Identification

My identification strategy is based on a combination of long-run and sign restrictions. I start by separating the VAR's structural shocks into two *sets*, depending on the fact that they do, or they do not have a permanent impact on inflation. Let the structural VAR(p) model be given by

$$Y_t = B_0 + B_1 Y_{t-1} + ... + B_p Y_{t-p} + A_0 \epsilon_t \tag{11}$$

where Y_t defined as before; A_0 being the impact matrix of the structural shocks at $t = 0$; and $\epsilon_t \equiv [\epsilon_t^{TE}, \epsilon_t^{MO}, \epsilon_t^{TA}, \epsilon_t^{MK}, \epsilon_t^{TR\prime}]'$ being the structural shocks, which, as standard practice, are assumed to be unit-variance and orthogonal to one another, with ϵ_t^{TE}, ϵ_t^{MO}, ϵ_t^{TA}, ϵ_t^{MK} being Canova and Paustian's (2011) 'technology', 'monetary policy', 'taste', and 'markup' shocks (to be discussed shortly), which are here allowed to exert a permanent impact on inflation, and ϵ_t^{TR} being instead a 2×1 vector of shocks which, by construction, have a transitory impact on inflation. The second row of the matrix of long-run impacts of the structural shocks, $[I_N - B(1)]^{-1} A_0$—i.e., the row

[38] By definition, a vertical long-run Phillips curve implies that the fraction of the permanent component of the unemployment rate which is explained by the permanent inflation shock is equal to zero.

[39] A subtle but important point here is the following. Since Canova and Paustian's 'robust sign restrictions' have originally been derived based on a New Keynesian model log-linearized around a zero-inflation steady-state, and in which inflation is *stationary*, it is in principle an open question whether such restrictions would also hold in the case in which, within the same model, inflation has a unit root. In fact, this is indeed the case, with the results for the case in which inflation is I(1) being numerically very close to the 'benchmark' results one obtains based on the specification reported in their Table 1 (all of these results are available upon request).

corresponding to inflation—is therefore postulated to have the following structure,

$$\text{Long-run impacts of the structural shocks on } \pi_t\text{:} \quad \begin{array}{ccccc} \epsilon_t^{TE} & \epsilon_t^{MO} & \epsilon_t^{TA} & \epsilon_t^{MK} & \epsilon_t^{TR\prime} \\ \end{array} \left[\begin{array}{ccccc} x & x & x & x & 0_{1\times 2} \end{array} \right] \tag{12}$$

—where a '0' means that the corresponding long-run impact has been restricted to zero, whereas an 'x' means that it has been left unrestricted—thus implying that ϵ_t^{TE}, ϵ_t^{MO}, ϵ_t^{TA}, and ϵ_t^{MK} may have a permanent impact on inflation, whereas ϵ_t^{TR} does not. The restriction that the two shocks in ϵ_t^{TR} are the only shocks which do not have a permanent impact on inflation is sufficient to disentangle them from the other four shocks. As for separating ϵ_t^{TE}, ϵ_t^{MO}, ϵ_t^{TA}, and ϵ_t^{MK} from one another, I achieve that by imposing the following set of sign restrictions on impact:

	Shock:			
Variable:	ϵ_t^{TE}	ϵ_t^{MO}	ϵ_t^{TA}	ϵ_t^{MK}
Output gap	+	-	+	-
Inflation	-	-	+	+
Unemployment rate	+	+	-	+
Short rate	-	+	+	+

where '+' means 'greater than, or equal to zero', and '-' means 'smaller than, or equal to zero'. In words, a technology shock causes the output gap and the unemployment rate not to decrease, and inflation and the short rate not to increase; a monetary shock causes the output gap and inflation not to increase, and unemployment and the short rate not to decrease; a taste shock causes unemployment not to increase, and all other variables not to decrease; and a markup shock causes the output gap not to increase, and all other variables not to decrease. These restrictions are the same as the 'robust sign restrictions' reported by Canova and Paustian (2011) in their Table 2 for their benchmark DSGE model featuring sticky prices, sticky wages, and several standard frictions (see the column labelled there as 'M'), with the only obvious difference that, since their model features employment, instead of the unemployment rate, the signs I am here imposing on the unemployment rate are the opposite of those reported by Canova and Paustian for employment.

In what follows I impose these sign restrictions only on impact. The reason for doing this is that, as stressed by Canova and Paustian (2011), whereas sign restrictions on impact are, in general, robust—in the specific sense that they hold for the vast majority of sub-classes within a specific class of DSGE models, and for the vast majority of plausible parameters' configurations—restrictions at longer horizons are instead, as they put it, 'whimsical', meaning that they are hard to pin down, and in general, they are not robust across sub-classes of models, and for alternative plausible parameters' configurations.[40] It is to be noticed that Canova and Paustian (2011) reached this conclusion based on a *quarterly* DSGE model. Since for the United States

[40] One obvious limitation of imposing the sign restrictions only on impact is that we are here using

I am here working at the monthly frequency, for reasons of consistency I impose the sign restrictions both on impact, and for the two months after the impact.

6.2 Computing the structural impact matrix A_0

For each draw from the posterior distribution, I compute the structural impact matrix, A_0 by combining the procedure proposed by Rubio-Ramirez, Waggoner, and Zha (2005) for imposing sign restrictions[41] with the imposition of the previously discussed zero restrictions on the matrix of long-run impacts of the structural shocks, $[I_N - B(1)]^{-1}A_0$, by means of a deterministic rotation matrix.[42] Specifically, for draw j from the posterior distribution of the VAR's estimates, for j = 1, 2, 3, ..., 10,000, let $P_j D_j P_j'$ be the eigenvalue-eigenvector decomposition of the VAR's covariance matrix, Ω_j, and let $\tilde{A}_{0,j} \equiv P_j D_j^{\frac{1}{2}}$. I draw 2,000 $N \times N$ matrices K_z, from the $N(0, 1)$ distribution, and for each of them I take its QR decomposition (that is, I compute matrices Q_z and R_z such that $K_z = Q_z \times R_z$) and I compute the 'starting estimate' of the structural impact matrix as $\bar{A}_{0,j}^z = \tilde{A}_{0,j} \cdot Q_z'$. I then impose the zero restrictions in the second row of the matrix of the long-run impacts of the structural shocks *via* an appropriate Householder matrix[43] H. If the resulting structural impact matrix $A_{0,j}^z = \bar{A}_{0,j}^z H$, for z = 1, 2, 3, ..., 2,000, satisfies the sign restrictions I store it, otherwise I discard it. In the end, for each draw j from the posterior distribution of the VAR's estimates, I end up with Z randomly drawn impact matrices satisfying both the sign and the long-run restrictions. If Z=0, I simply ignore draw j from the posterior; if Z=1, I just keep the single successful impact matrix; and if Z>1 I randomly draw and keep one, among the Z successful impact matrices. In this way, I end up, for each of the 10,000 draws from the posterior distribution, with either zero or one randomly drawn structural impact matrices satisfying both the long-run and the sign restrictions. In what follows the inference is based on the draws for which I got one successful structural impact matrix.

---

a comparatively limited amount of information in order to achieve identification. As a consequence, our results necessarily end up being less sharp than they could have been had we been reasonably confident about imposing a specific pattern of sign restrictions at horizons greater than zero. This compounds a well-known limitation of sign restrictions which has been extensively discussed by Fry and Pagan (2007): as these authors stress, sign restrictions are intrinsically 'weak information', since they are based on the notion of uniquely imposing a specific pattern of signs on the IRFs. The rationale behind our decision of imposing sign restrictions only on impact is that it is better to impose a limited amount of information about which we can be reasonably confident than a greater amount of information about which we have limited confidence.

[41] See at http://home.earthlink.net/~tzha02/ProgramCode/SRestrictRWZalg.m.

[42] I wish to thank Dan Waggoner for useful suggestions on the implementation of the identification strategy adopted herein.

[43] I compute the Householder matrix *via* Algorithm 5.5.1 of Golub and VanLoan (1996).

6.3 Evidence

Figure 3 shows, for either country, the posterior distributions of the long-run impacts on the unemployment rate of a one per cent permanent shock to inflation, whereas Figures 4 and 5 show the posterior distributions of the fractions of the permanent component of inflation and the unemployment rate, respectively, explained by either of the permanent inflation shocks. Table 3 reports the modes, the medians, and the 90%-coverage percentiles of the posterior distributions of the long-run impact on the unemployment rate of a one per cent permanent shock to inflation, together with the fractions of draws for which the impacts are estimated to be negative. Finally, Tables 4 and 5 report, for inflation and the unemployment rate, respectively, the medians and the 90%-coverage percentiles of the posterior distributions of the fractions of the permanent component of inflation and the unemployment rate, respectively, explained by either of the permanent inflation shocks, together with the fractions of the mass of the posterior distribution which are below three selected 'cut-off points', 0.1, 0.05, and 0.01 (this is in order to provide a numerical measure of how strongly clustered towards zero such distributions are).

The main findings can be summarised as follows.

First, the extent of uncertainty associated with estimated long-run trade-offs is uniformly substantial. This originates from the fact that the feature of the data we are here estimating pertains to the infinite long-run, and, as it is well known—see, first and foremost, Faust and Leeper (1997)—this inevitably produces imprecise estimates, unless the researcher is willing to impose upon the data very strong informational assumptions (which, in general, is not advisable to do). Within the present context, this problem is compounded by our use of sign restrictions, which, as stressed by Fry and Pagan (2007), are intrinsically 'weak information', and should therefore not be expected to produce strong inference. On the other hand, it has to be stressed that, for the reasons discussed in the Section 1.1, the approach to identification adopted herein is, most likely, the most credible (or, to be more precise, the least incredible) one given the current state-of-the-art in structural VAR econometrics. To put it differently, given the lack of reliability of inertial restrictions extensively documented, e.g., by Canova and Pina (2005),[44] and the unattractiveness of 'calibrating' elements of the VAR's structural impact matrix, it is not clear *what* kind of short-run restrictions can be combined with long-run ones in order to disentangle the various shocks exerting a permanent impact on inflation, other than sign restrictions. In a sense, this is the best we can do given the current state of knowledge, and if the price to be paid for using the least incredible identification strategy is a significant extent of uncertainty, we ought to live with it ...

Second, consistent with the results based on VARs featuring a single permanent inflation shock, for either country, neither of the four structural shocks which are here allowed to have a permanent impact on the inflation rate is estimated to have

[44] See also Carlstrom, Fuerst, and Paustian (2009).

had a long-run impact on the unemployment rate significantly different from zero at conventional levels. For the United States, for example, the fractions of draws for which the long-run impact on the unemployment rate of a one per cent permanent shock to inflation is estimated to have been negative range between 44.4 per cent for the taste shock to 76.4 per cent for the monetary shock, whereas for the Euro area they range between 40.9 per cent for the taste shock to 65.0 per cent for the technology shock. This shows that even disentangling the overall set of structural shocks inducing a permanent impact on inflation into shocks with different characteristics does not allow us to detect a specific shock which has generated a non-zero long-run trade-off between inflation and unemployment. An important point to stress, however, is that, as pointed out in the previous paragraph, the overall extent of uncertainty is substantial, so that the results reported in Table 3 are, in fact, compatible with a comparatively wide range of possible slopes. This means that, although we regard the null hypothesis of a vertical long-run trade-off (that is, within the present context, no long-run impact on unemployment of permanent inflation shocks) as the natural one, a researcher with a different view of the world, and therefore different priors about what the natural benchmark should be, might not see her views about the slope of the long-run Phillips curve falsified.

Third, at the same time, it has to be stressed that either modal or median estimates of the long-run impact on the unemployment rate of a one per cent permanent shock on inflation are typically quite small, and point towards weak-to-non-existent trade-offs. Focusing on modal estimates, for example, they range between -0.369 and 0.034 for the United States; between -0.170 and 0.050 for the United Kingdom; and between -0.156 and 0.060 for Canada. For the Euro area, they range between -0.033 and 0.033 with the single exception of the technology shock, for which the modal estimate is equal to -0.501.

Fourth, the evidence reported in Figure 4 suggests that, for either country, mark-up shocks have played an important role in introducing a unit root in inflation. In order to interpret these results, it is important to keep in mind that, within the present context, mark-up shocks should be logically expected to capture disturbances to food and energy prices, which played an important role around the time of the Great Inflation. For either the United Kingdom or Canada, monetary policy shocks also played a comparatively important role, whereas for the Euro area this has been the case for taste shocks.

Fifth, turning to the fractions of the permanent component of the unemployment rate explained by individual shocks, for either country, none of the four shocks especially stands out, with the possible exception of the technology shock for the Euro area.

Summing up, evidence for the United States suggests that, even if one really wants to give the long-run Phillips curve its best chance of manifesting itself, and therefore decides to narrowly focus on the pre-Volcker period, for which evidence of a unit root in inflation is stronger, still, it is not possible to detect strong evidence of a non-

vertical long-run trade-off conditional on any shock. The same holds true for either of the other three countries. The extent of uncertainty, however, is substantial, so that a researcher holding a very strong prior that the long-run Phillips curve is very steep, but not vertical, will not find her view of the world falsified by these results.

7 What About Cointegration?

Up until now I have completely ignored the issue of cointegration. On logical/conceptual grounds, the notion of cointegration between inflation and the unemployment rate—which implies that *all* of the shocks exerting a permanent impact on inflation have an equi-proportionate permanent impact on the unemployment rate[45]—might appear (at least, to me ...) as bizarre. Why should the ratio between the permanent impacts on unemployment and inflation induced by (say) a technology shock be *identical* to the ratio between the permanent impacts on the two variables induced by a monetary policy shock? From a statistical point of view, however, it is important to ascertain what the data actually tell us, and in this section I therefore proceed to perform either bivariate or trivariate cointegration tests between inflation, the unemployment rate or the short rate based on the procedure proposed by Johansen.

Table 6 reports bootstrapped[46] p-values for Johansen's trace test of the null of no cointegration between either inflation and unemployment, inflation and the short rate (with the single exception of the United Kingdom, for which results from the ADF tests reported in Table 1 strongly reject the null of a unit root in the short rate), or inflation, the unemployment rate, and the short rate. p-values have been computed by bootstrapping the VAR estimated for the first difference of the relevant vector of series. To be clear, this means that, given (e.g.) the vector $Y_t = [\pi_t, U_t]'$, we start by selecting the lag order for cointegration tests as the maximum between the lag orders selected based on the Schwartz and the Hannan-Quinn criteria,[47] and we perform Johansen's trace test of the null of no cointegration. Then, we estimate the VAR[48]

[45] To be clear, this implies that, if (e.g.) the permanent impact on the unemployment rate of a one per cent permanent increase in the inflation rate induced by monetary shocks is equal to -2, it ought to be the case that the same holds true for *all* of the other shocks: for each of them, the ratio between the two permanent impacts on the unemployment rate and the inflation rate ought to be, likewise, equal to -2.

[46] Since I am here bootstrapping critical and p-values, my results are robust to (i) the possible non-normality of the shocks, and (ii) small-sample problems. As for the former issue, under very general conditions the residuals of the VAR estimated under the null of no cointegration are consistent estimates of the true underlying shocks, in the sense that they do converge to such shocks in probability limit. As a result, the underlying shocks may have, in population, any non-degenerate distribution we can think of, because bootstrapping the residuals automatically takes care of that. As for the latter issue, as I discuss in Section 8.1.1 below, bootstrapping effectively takes care of that, too.

[47] See Luetkepohl (1991).

[48] The VAR we are estimating here is not a cointegrated VAR, that is, it is equal to the VECM representation without the error-correction term.

for ΔY_t, we bootstrap it 10,000 times, thus generating bootstrapped artificial series $\Delta\tilde{Y}_t^j$, based on each of them we compute corresponding bootstrapped artificial series $\tilde{Y}_t^j$—that is, those for the levels of the series—and based on each of them we perform the same trace test we previously computed based on the actual data, thus building up the empirical distribution of the trace statistic under the null of no cointegration. Then, based on this distribution, we compute critical values (not reported here) and p-values. For the United States we detect no evidence of cointegration between inflation and the unempoyment rate, which implies that, under this respect, the previously discussed SVAR-based results are not subject to any *caveat*. On the other hand, we detect strong evidence of cointegration between inflation and the Federal Funds rate, with a p-value equal to 1.0e-4. This is consistent with Barsky's (1987) findings on the appearance, for the first time, of a Fisher effect in U.S. data for the post-WWII sample period comprising the Great Inflation episode.[49] The median estimate of the second element of the cointegrating vector is -1.1011, whereas the p-value for rejecting the null hypothesis that the cointegration vector is [1; -1] is 0.0872. For the Euro area and Canada, on the other hand, bivariate cointegration tests do not detect any evidence of cointegration between inflation and the short rate. Although at first sight puzzling—taken at face value, these results imply a rejection of the Fisher hypothesis that permanent shifts in inflation should map one-to-one into corresponding shifts in interest rates—it is important to stress that empirical evidence on the violation of the Fisher hypothesis is widespread,[50] so that these results should not be seen as surprising at all. Finally, for either the United States, the Euro area or, Canada we detect strong evidence of cointegration between inflation, unemployment, and the short rate, whereas for the United Kingdom evidence points, at the 10 per cent level, towards cointegration between inflation and the unemployment rate.

For the United States, the Euro area and Canada, for which we detected evidence of cointegration between the three series, we also report bootstrapped p-values for testing the null hypothesis of one single cointegrating vector, versus the alternative of two cointegrating vectors. Details of the bootstrapping procedure are the same as before, with the only difference that, instead of bootstrapping the estimated VAR for ΔY_t under the null of no cointegration, we bootstrap the VECM estimated conditional on there being one single cointegrating vector. Whereas for Canada the p-value, at 0.668, is very far from being signifiant at any conventional significant level, for the Euro area, at 0.012, it suggests the presence of an additional cointegrating vector. Finally, for the United States, the p-value, equal to 0.074, points towards the presence of an additional cointegrating vector at the 10 per cent level.

Figure 7 reports, for the Euro area and the United Kingdom, for which we detected evidence of cointegration in the bivariate representation for $Y_t = [\pi_t, U_t]'$ at the 10

[49] Working in the middle of the Great Inflation episode, and therefore with less information that Barsky, Fama (1975) produced very similar evidence for the period 1953-1971.

[50] After all, the key reason why Fama's (1975) paper had such a resonance was precisely because it produced, for the first time, decisive evidence in favor of the Fisher effect.

per cent level, the bootstrapped distribution of the ratio between the permanent impacts of the common shock on the unemployment rate and inflation, respectively, and of the elements of the loading vector of the cointegrating residual in the VECM representation. Results are completely different from those we previously discussed based on the structural VARs. The long-run impact on the unemployment rate induced by the common shock, normalized in such a way as to induce a one per cent permanent increase in inflation, is here not only highly statistically significantly different from zero—as implied by the results from the trace test, and as testified by the fact that mass of the bootstrapped distribution is pretty much away from zero—but it is also sizeable, with modal estimates around -1.1 for the Euro area, and -1.9 for the United Kingdom.

7.1 Monte Carlo evidence on the performance of Johansen's procedure

How should we interpret these results? One possibility is that they are just a statistical fluke due to the bad performance of Johansen's procedure in small-samples.

7.1.1 The case of two independent random walks

As Table 7 shows, this is actually not the case. The table reports results from four sets of 10,000 Monte Carlo simulations of lengths equal to the actual sample lengths I am working with for the four countries.[51] For each simulation, I randomly generate two independent random walks, and I apply *exactly* the same procedure I previously applied to the actual data, computing the p-values by bootstrapping the estimated VAR for the first differences of the two random walks. Ideally, out of the 10,000 simulations, the fraction of bootstrapped p-values below x per cent should be equal to x per cent. As the results reported in the table show, the bootstrapped Johansen procedure I am using herein gets quite remarkably close to this ideal: the fraction of simulations for which the bootstrapped trace statistic incorrectly rejects the null of no cointegration between the two independent random walks at a given significance level ranges between 11.3 and 11.9 per cent at the 10 per cent level; between 5.5 and 6.0 per cent at the 5 per cent level; and between 1.1 and 1.3 per cent at the 1 per cent level. Quite remarkably, the performance for the United Kingdom, for which we have just 81 quarterly observations, is not dramatically different from that for Canada, for which we have instead 118 observations. This testifies to the power of bootstrapping, which can effectively take into account of the specific characteristics of the data the researcher is working with.

[51] Concerning the simulations pertaining to the United States, I set T=82—that is, the number of quarters in the sample period Januanry 1959-July 1979—rather than T=246, which would be the corresponding number of months.

7.1.2 Taking the estimated structural VARs as data-generation processes

Although the bootstrapped Johansen procedure used herein performs remarkably well conditional on a data-generation process (henceforth, DGP) in which the series of interest are independent random walks, it is an open question how well such a procedure performs conditional on DGPs such as the previously estimated structural VARs. To put it differently, suppose that the structural VARs we previously estimated—in which, by construction, there is *no* cointegration whatsoever between any series—are, for either country, the true model of the economy: how often would the Johansen procedure incorrectly reject the null of no cointegration? Table 8 reports evidence on this, by showing the fraction of simulations for which Johansen's bootstrapped trace statistic incorrectly rejects the null of no cointegration at a given significance level, based on taking the estimated structural VARs as the DGPs. As the tables shows, Johansen's procedure tends to spuriously detect cointegration a non-negligible fraction of the times. Taking the estimated Bayesian structural VARs featuring four permanent inflation shocks as DGPs, for example, for the Euro area and the United Kingdom, the fractions of bootstrapped p-values for Johansen's trace statistic for testing the null of no cointegration between inflation and the unemployment rate which are smaller than 10 per cent are equal to 26.9 and 23.6 per cent respectively. This means that the trace test would incorrectly reject the null of no cointegration between inflation and unemployment at the 10 per cent level about one-fourth of the times. As for the trace test of cointegration between inflation, unemployment, and the short rate, the fraction of bootstrapped p-values smaller than 10 per cent range between 45.4 and 72.1 per cent, thus essentially pointing towards the unreliability of such tests conditional on data generation processes such as thise we estimated in Section 6. Results based on taking the Classical VARs featuring a single permanent inflation shock as the DGP[52] are in line with those discussed so far for the trace test of cointegration between inflation, unemployment, and the short rate, and are instead significantly better for the test of cointegration between inflation and unemployment.

Overall, the results reported in Table 8 suggest that evidence such as that reported in Table 6 should be discounted, as those results might as well be due to the limitations of cointegration tests conditional on these specific DGPs.

7.2 The adjustment dynamics implied by the estimated loading vectors

Even ignoring the results reported in Table 8, the evidence shown in the last two columns of Figure 8 is hardly compatible with the notion of an *exploitable* long-run Phillips trade-off, that is, a trade-off which a policymaker might be able to purpose-

[52] Specifically, we are here considering the standard, VAR-based estimator of the spectral density matrix at the frequency zero.

fully use in order to permanently decrease the unemployment rate by engineering a higher equilibrium inflation rate. As pointed out by Evans (1994) in his comment on King and Watson (1994),

> '[...] for a trade-off to be viewed as exploitable, a decision-maker must have confidence that unemployment can be reduced by engineering a higher rate of inflation [...] It would be of little comfort to many politicians if the Phillips curve trade-off simply implied that if unemployment instead turned higher, at least inflation would be lower.'

The bootstrapped distributions of the estimates of the elements of the loading vector of the error-correction term in the VECM reported in Figure 8, however, clearly point towards the second case, with the estimated loading for the first difference of the unemployment rate being uniformly very small, and, in the case of the United Kingdom, not being significantly different from zero at conventional levels. This implies that, following a shock to the common stochastic trend, the bulk of the dynamic adjustment to disequilibrium takes place through movements in inflation, rather than *via* movements in unemployment. This is difficult to reconcile with the notion of an exploitable Phillips trade-off, and might instead be compatible (e.g.) with the notion that, in both countries, policymakers engineered sharp recessions to end the Great Inflation. In turn, such recessions caused permanent increases in the unemployment rate *via* hysteresis effects, and finally, inflation came down gradually.

8 Conclusions

In this paper I have used both Classical and Bayesian structural VARs identified based on either long-run restrictions, or a combination of long-run and sign restrictions, to investigate the long-run trade-off between inflation and the unemployment rate in the U.S., the Euro area, the U.K., and Canada over the post-WWII period. Results based on Classical VARs featuring a single permanent inflation shock do not allow to reject the null hypothesis of a vertical long-run Phillips curve for either country, with both the modes and medians of the bootstrapped distributions of the long-run impact on unemployment of a one per cent permanent shock to inflation being close to zero. Results based on Bayesian VARs allowing for four permanent inflation shocks, which are sorted out from one another by means of DSGE-based robust sign restrictions, produce a very similar picture. The overall extent of uncertainty is however substantial, especially in the latter case, thus suggesting that the data are compatible with a comparatively wide range of possible slopes of the long-run trade-off. For all countries, Johansen's cointegration tests point towards the presence of cointegration between either inflation and unemployment, or inflation, unemployment, and a short-term interest rate, with the long-run Phillips trade-off implied by the estimated cointegrating vectors being negative and sizeable. I argue however that this evidence

should be discounted, as, conditional on the estimated structural VARs—which, by construction, do not feature cointegration between any variable—Johansen's procedure tends to spuriously detect cointegration a non-negligible, and sometimes large, fraction of the times.

References

Akerlof, G., W. T. Dickens, and G. Perry (1996): "The Macroeconomics of Low Inflation," *Brookings Papers on Economic Activity*, 1996(1), 1–76.

Barsky, R. (1987): "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19(1), 3–24.

Barsky, R., and E. Sims (2011): "News shocks and business cycles," *Journal of Monetary Economics*, 58(3), 273–289.

Benati, L. (2008): "Investigating Inflation Persistence Across Monetary Regimes," *Quarterly Journal of Economics*, 123(3), 1005–1060.

Benigno, G., and L. A. Ricci (2011): "The Inflation-Output Trade-Off with Downward Wage Rigidities," *American Economic Review*, 101(4), 1436–66.

Bernanke, B., M. Gertler, and M. Watson (1997): "Systematic Monetary Policy and the Effects of Oil Price Shocks," *Brookings Papers on Economic Activity*, 1997(1), 91–157.

Bernanke, B. S., T. Laubach, F. S. Mishkin, and A. Posen (1999): *Inflation Targeting: Lessons from the International Experience*. Princeton University Press.

Bullard, J., and J. W. Keating (1995): "The Long-Run Relationship Between Inflation and Output in Postwar Economies," *Journal of Monetary Economics*, 36, 477–496.

Canova, F., and G. de Nicolo (2002): "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, 49(6), 1131Ű–1159.

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

Canova, F., and J. Pina (2005): "What VAR Tell Us About DSGE Models," in Diebolt, C. and Kyrtsou, C. New Trends In Macroeconomic, Springer Verlag.

Carlstrom, C., T. Fuerst, and M. Paustian (2009): "Monetary Policy Shocks, Choleski Identification, and DNK Models," *Journal of Monetary Economics*, 56(7), 1014–1021.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2006): "Assessing Structural VARS," *in D. Acemoglu, K. Rogoff and M. Woodford, eds. (2007), NBER Macroeconomics Annuals 2006.*

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, CXV(1), 147–180.

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

Evans, C. (1994): "Comment on: The Post-War U.S. Phillips Curve, A Revisionist Econometric History," *Carnegie-Rochester Conference Series on Public Policy*, 41, 221–230.

Fama, E. (1975): "Short-Term Interest Rates as Predictors of Inflation," *American Economic Review*, 65(3), 269–282.

Faust, J. (1998): "The Robustness of Identified VAR Conclusions About Money," *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.

Faust, J., and E. Leeper (1997): "When Do Long-Run Identifying Restrictions Give Reliable Results?," *Journal of Business and Economic Statistics*, 15(3), 345–353.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2008): "How Structural Are the Structural Parameters?," *NBER Macroeconomics Annuals*, The MIT Press, 83–137.

Fisher, M. E., and J. J. Seater (1993): "Long Run Neutrality and Superneutrality in an ARIMA Framework," *American Economic Review*, 83(3), 402–415.

Fry, R., and A. Pagan (2007): "Some Issues in Using Sign Restrictions for Identifying Structural VARs," NCER Working Paper n. 14, April 2007.

Golub, G., and C. VanLoan (1996): *Matrix Computations, III Edition*. Johns Hopkins University Press, Baltimore.

Hamilton, J. (1994): *Time Series Analysis*. Princeton, NJ, Princeton University Press.

Hansen, B. (1999): "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics*, 81(4), 594–607.

Kilian, L. (1998): "Small-Sample Confidence Intervals for Impulse-Response Functions," *Review of Economics and Statistics*, pp. 218–230.

King, R., and M. Watson (1994): "The Post-War U.S. Phillips Curve: A Revisionist Econometric History," *Carnegie-Rochester Conference Series on Public Policy*, 41, 157–219.

Kurmann, A., and C. Otrok (2013): "News Shocks and the Slope of the Term Structure of Interest Rates," *American Economic Review*.

Lubik, T., and F. Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94(1), 190–217.

Lucas, R. E. (1972a): "Econometric Testing of the Natural Rate Hypothesis," *in Eckstein, O., ed. The Econometrics of Price Determination, Washington, D.C., Board of Governors of the Federal Reserve System*, pp. 50–59.

——— (1972b): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4(April 1972), 103–24.

——— (1973): "Some International Evidence on Output-Inflation Trade-Offs," *American Economic Review*, 63(3), 326–334.

Lucas, R. E. (1976): "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy*, 1, 19–46.

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

Pybus, T. (2011): "Estimating the UKŠs Historical Output Gap," *Office for Budget Responsibility Working paper No. 1, November 2011.*

Roberts, J. M. (1993): "The Sources of Business Cycles: A Monetarist Interpretation," *International Economic Review*, 34(4), 923–934.

Rubio-Ramirez, J., D. Waggoner, and T. Zha (2005): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, 77(2), 665–696.

Sargent, T. J. (1971): "A Note on the Accelerationist Controversy," *Journal of Money, Credit and Banking*, 3(3), 721–725.

Sargent, T. J. (1987): *Macroeconomic Theory, II Edition.* Orlando, Academic Press.

Solow, R. (1968): "Recent Controversy on the Theory of Inflation: An Eclectic View," *in Proceedings of a Symposium on Inflation: Its Causes, Consequences, and Control, S. Rousseaus (ed.), New York: New York University.*

Stock, J. H., and M. W. Watson (2007): "Why Has U.S. Inflation Become Harder to Forecast?," *Journal of Money, Credit, and Banking*, 39(1), 3–33.

Svensson, L. E. (2013): "The Possible Unemployment Cost of Inflation Below a Credible Target," paper presented at the NBER Monetary Economics Meeting, March 8, 2013.

Tobin, J. (1968): "Discussion," *in Proceedings of a Symposium on Inflation: Its Causes, Consequences, and Control, S. Rousseaus (ed.), New York: New York University.*

Uhlig, H. (1998): "Comment On: The Robustness of Identified VAR Conclusions About Money," *Carnegie-Rochester Conference Series on Public Policy*, 49, 245–263.

——— (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52(2), 381–419.

A The Data

Here follows a detailed description of the dataset.

A.1 United States

A monthly seasonally adjusted series for the civilian unemployment rate ('UNRATE: Civilian Unemployment Rate, Seasonally Adjusted, Percent') is from the U.S. the Department of Labor, Bureau of Labor Statistics, and it has been converted to the quarterly frequency by taking averages within the quarter. A monthly seasonally unadjusted series for the 3-month Treasury Bill ('TB3MS: 3-Month Treasury Bill: Secondary Market Rate') is from Board of Governors of the Federal Reserve System, and it has been converted to the quarterly frequency by taking averages within the quarter. By the same token, monthly seasonally unadjusted series for the 1-, 3-, 5-, and 10-year Treasury constant maturity rates (acronyms are GS1, GS3, GS5, and GS10, respectively), and for the effective Federal Funds rate (acronyms is FEDFUNDS) are all from the Board of Governors of the Federal Reserve System. A monthly seasonally adjusted series for the consumer price index ('CPIAUCSL: Consumer Price Index, Seasonally Adjusted, Monthly, Index 1982-84=100') is from the U.S. the Department of Labor, Bureau of Labor Statistics, and it has been converted to the quarterly frequency by taking averages within the quarter. Quarterly seasonally adjusted series for real GDP ('GDPC96: Real Gross Domestic Product, 3 Decimal, Seasonally Adjusted Annual Rate, Billions of Chained 2005 Dollars') and the GDP deflator ('GDPCTPI: Gross Domestic Product: Chain-type Price Index, Seasonally Adjusted') are from the U.S. Department of Commerce: Bureau of Economic Analysis. The quarterly series for potential GDP ('GDPPOT: Real Potential Gross Domestic Product, Quarterly, Billions of Chained 2005 Dollars') is from the U.S. Congressional Budget Office, Budget and Economic Outlook. Real consumption expenditure of non-durable goods and services has been computed based on Table 1.1.6 of the U.S. National Income and Product Accounts. The output gap has been computed as the percentage deviation of real GDP from the Congressional Budget Office estimate of potential real GDP. It is important to stress that, since both series are expressed in the same unit (billions of chained 2005 dollars), such a computation is indeed perfectly legitimate. Monthly interpolated seasonally adjusted series for real GDP, the GDP deflator, and real consumption expenditure are from Mark Watson's website, and are available from January 1959 to June 2010. The quarterly estimate of the output gap implied by the Congressional Budget Office estimate of potential output has been interpolated to the monthly frequency as in Bernanke, Gertler, and Watson (1997), using, as interpolators, the monthly ratios between investment in non-residential structures and GDP, between investment in residential structures and GDP, and between the change in inventories and GDP. In turn, these ratios have been constructed based on the monthly interpolated national accounts data found at

Mark Watson's website.

A.2 Euro area

Quarterly, seasonally adjusted series for real GDP, the GDP deflator, the CPI, real consumption expenditure, real gross fixed capital formation, the unemployment rate, the output gap, and a short and a long rate are from the European Central Bank's database.

A.3 United Kingdom

The output gap estimate is from Pybus (2011). I wish to thank Tom Pybus, of the U.K. *Office for Budget Responsibility*, for kindly sharing his output gap series with me. Quarterly, seasonally adjusted series for real GDP (acronym is 'ABMI'), real final consumption by households ('ABJR'), real total gross fixed capital formation ('NPQT'), nominal GDP ('YBHA'), and real change in inventories ('CAEX') are from the U.K. Office for National Statistics. A quarterly, seasonally adjusted series for the GDP deflator has been computed as the ratio between YBHA and ABMI. Monthly, seasonally adjusted series for the unemployment rate based on the claimant count (acronym is 'BCJE') and the retail price index are from the U.K. Office for National Statistics, and they have been converted to the quarterly frequency by taking averages within the quarter. Monthly, seasonally unadjusted series for the discount rate and a long-term rate ('Interest Rates, Government Securities, Government Bonds') are from the International Monetary Fund's International Financial Statistics, and they have been converted to the quarterly frequency by taking averages within the quarter. Since, over the sample period, the consumption/GDP ratio exhibits some trend (which, according to the permanent income hypothesis should obviously not be there), before entering the ratio into the VARs I quadratically detrend it.

A.4 Canada

Quarterly, seasonally adjusted series for real GDP (acronym is 'CANRGDPQDSNAQ'), the harmonized unemployment rate for all persons ('CANURHARMQDSMEI'), real private final consumption expenditure ('CANPFCEQDSNAQ'), real gross fixed capital formation ('CANGFCFQDSNAQ'), the GDP implicit price deflator ('CANGDPDEFQISMEI'), and the CPI are from the Organisation for Economic Co-operation and Development (either from the Quarterly National Accounts, or from the Main Economic Indicators). Since, over the sample period, the consumption/GDP ratio exhibits some trend (which, according to the permanent income hypothesis should obviously not be there), before entering the ratio into the VARs I quadratically detrend it. A monthly, seasonally unadjusted series for the short rate, treasury bill auction, average yields, is from the Bank of Canada. A monthly, seasonally unadjusted series

for a long rate, is from the International Monetary Fund's International Financial Statistics. Both monthly series have been converted to the quarterly frequency by taking averages within the quarter. As for the output gap, an 'official' series computed by the Bank of Canada, and available from its website, starts in 1981Q1. Since our sample period starts in 1961Q1, in our analysis we use instead HP-filtered log real GDP, which, since 1981Q1, has exhibited a remarkably close co-movement witrh the Bank of Canada's output gap estimate. Specifically, (i) the contemporaneous correlation between the two series has been equal to 0.949, whereas (ii) the fraction of variance explained by the first principal component of the panel composed of the two standardized series has been equal to 0.974.

A.5 Japan

Monthly, seasonally adjusted series for the unemployment rate and the consumer price indices are from the International Monetary Fund's International Financial Statistics. Both monthly series have been converted to the quarterly frequency by taking averages within the quarter.

A.6 Australia

Monthly, seasonally adjusted series for the unemployment rate and the consumer price indices are from the International Monetary Fund's International Financial Statistics. Both monthly series have been converted to the quarterly frequency by taking averages within the quarter.

A.7 Sweden

Monthly, seasonally adjusted series for the unemployment rate and the consumer price indices are from the International Monetary Fund's International Financial Statistics. Both monthly series have been converted to the quarterly frequency by taking averages within the quarter.

Table 1 Bootstrapped p-values[a] for augmented Dickey-Fuller tests without trend

	GDP deflator inflation			CPI inflation			Unemployment rate			Short rate		
	p=1	p=2	p=4	p=1	p=2	p=4	p=1	p=2	p=4	p=1	p=2	p=4
United States[b]												
full sample (Jan. 1959-Jun. 2010)	0.000	0.045	0.311	0.000	0.002	0.041	0.280	0.026	0.043			
pre-Volcker (Jan. 1959-Jul. 1979)	0.074	0.398	0.640	0.082	0.686	0.737	0.382	0.254	0.352	0.259	0.414	0.168
Volcker, Greenspan, and Bernanke (Aug. 1979-Jun. 2010)	0.001	0.018	0.008	0.000	0.000	0.001	0.585	0.103	0.113			
Euro area (1970Q1-1998Q4)	0.263	0.515	0.744	0.400	0.693	0.663	0.150	0.238	0.258	0.590	0.328	0.362
Japan (1966Q4-2011Q4)	0.014	0.052	0.054	0.002	0.031	0.110	0.586	0.575	0.572			
United Kingdom (1972Q1-1992Q3)	0.039	0.222	0.199	0.073	0.136	0.272	0.468	0.169	0.329	0.020	0.018	0.024
Canada (1961Q1-1990Q4)	0.030	0.136	0.193	0.096	0.150	0.229	0.570	0.527	0.633	0.316	0.333	0.350
Sweden (1970Q1-1992Q4)				0.000	0.015	0.175	0.981	0.805	0.440			
Australia (1969Q3-1994Q2)	0.000	0.000	0.001	0.014	0.123	0.304	0.408	0.286	0.307			

[a] Based on 10,000 bootstrap replications of estimated ARIMA processes.

[b] Based on monthly data, lag orders are equal to 3, 6, and 12 months.

Table 2 Results based on VARs allowing for a single permanent inflation shock: bootstrapped and posterior distributions of the long-run impact on the unemployment rate of a one per cent permanent shock to inflation, and fractions of the mass of the distribution for which the impact is negative

	Mode, median, and 90%-coverage percentiles			Fraction of the mass of the distribution for which the impact is negative
United States, pre-Volcker (January 1959-July 1979)				
Classical				
Bartlett estimator of the spectral density matrix	-0.105	-0.200	[-0.620; 0.185]	0.820
VAR-based estimator of the spectral density matrix	-0.216	-0.191	[-0.613; 0.185]	0.811
Bayesian	-0.200	-0.200	[-0.869; 0.404]	0.750
Euro area (1970Q1-1998Q4)				
Classical				
Bartlett estimator of the spectral density matrix	0.056	-0.011	[-0.426; 0.453]	0.518
VAR-based estimator of the spectral density matrix	-0.040	0.003	[-0.432; 0.452]	0.495
Bayesian	-0.008	0.012	[-0.767; 0.758]	0.485
United Kingdom (1972Q1-1992Q3)				
Classical				
Bartlett estimator of the spectral density matrix	-0.088	-0.081	[-0.308; 0.158]	0.734
VAR-based estimator of the spectral density matrix	-0.104	-0.080	[-0.302; 0.157]	0.736
Bayesian	-0.072	-0.065	[-0.411; 0.287]	0.680
Canada (1961Q1-1990Q4)				
Classical				
Bartlett estimator of the spectral density matrix	-0.121	-0.096	[-0.271; 0.070]	0.838
VAR-based estimator of the spectral density matrix	-0.121	-0.096	[-0.270; 0.070]	0.837
Bayesian	-0.088	-0.100	[-0.373; 0.134]	0.796

Table 3 Results based on Bayesian VARs allowing for four permanent inflation shocks: posterior distributions of the long-run impact on the unemployment rate of a one per cent permanent shock to inflation, and fractions of draws for which the impact is negative

	Mode, median, and 90%-coverage percentiles			Fraction of draws for which impact is negative
United States, pre-Volcker (January 1959-July 1979)				
technology shock	-0.235	-0.235	[-3.306; 2.591]	0.644
monetary shock	-0.302	-0.379	[-2.854; 2.047]	0.764
taste shock	0.034	0.063	[-1.385; 1.925]	0.444
mark-up shock	-0.369	-0.312	[-1.471; 0.857]	0.726
Euro area (1970Q1-1998Q4)				
technology shock	-0.501	-0.399	[-5.021; 5.290]	0.650
monetary shock	-0.033	0.090	[-2.671; 2.880]	0.440
taste shock	0.033	0.083	[-1.034; 1.370]	0.409
mark-up shock	0.033	0.061	[-1.019; 1.164]	0.438
United Kingdom (1972Q1-1992Q3)				
technology shock	-0.070	-0.065	[-1.601; 1.342]	0.587
monetary shock	-0.170	-0.172	[-0.875; 0.440]	0.813
taste shock	-0.070	-0.035	[-1.734; 1.625]	0.543
mark-up shock	0.050	0.029	[-0.615; 0.618]	0.449
Canada (1961Q1-1990Q4)				
technology shock	-0.156	-0.177	[-2.144; 1.587]	0.697
monetary shock	-0.156	-0.225	[-1.049; 0.203]	0.870
taste shock	-0.132	-0.107	[-1.892; 1.506]	0.625
mark-up shock	0.060	0.025	[-0.479; 0.458]	0.447

Table 4 Results based on Bayesian VARs allowing for four permanent inflation shocks: fractions of the permanent component of inflation explained by individual shocks

	Median, and 90%-coverage percentiles		Fraction of the mass of the posterior distribution below:		
			0.1	0.05	0.01
United States, pre-Volcker (January 1959-July 1979)					
technology shock	0.067	[0.001; 0.443]	0.599	0.440	0.202
monetary shock	0.138	[0.001; 0.646]	0.426	0.303	0.138
taste shock	0.187	[0.002; 0.656]	0.363	0.250	0.107
mark-up shock	0.398	[0.066; 0.856]	0.088	0.031	0.002
Euro area (1970Q1-1998Q4)					
technology shock	0.109	[0.001; 0.605]	0.481	0.348	0.156
monetary shock	0.087	[0.001; 0.490]	0.533	0.385	0.177
taste shock	0.328	[0.012; 0.755]	0.185	0.111	0.045
mark-up shock	0.299	[0.051; 0.715]	0.129	0.048	0.005
United Kingdom (1972Q1-1992Q3)					
technology shock	0.092	[0.001; 0.536]	0.518	0.376	0.172
monetary shock	0.287	[0.007; 0.770]	0.246	0.159	0.065
taste shock	0.073	[0.001; 0.496]	0.565	0.421	0.193
mark-up shock	0.356	[0.062; 0.793]	0.099	0.036	0.004
Canada (1961Q1-1990Q4)					
technology shock	0.076	[0.001; 0.521]	0.564	0.412	0.194
monetary shock	0.263	[0.007; 0.707]	0.245	0.155	0.061
taste shock	0.093	[0.001; 0.585]	0.516	0.380	0.176
mark-up shock	0.373	[0.065; 0.794]	0.091	0.036	0.004

Table 5 Results based on Bayesian VARs allowing for four permanent inflation shocks: fractions of the permanent component of the unemployment rate explained by individual shocks

	Median, and 90%-coverage percentiles		Fraction of the mass of the posterior distribution below:		
			0.1	0.05	0.01
United States, pre-Volcker (January 1959-July 1979)					
technology shock	0.064	[0.001; 0.474]	0.602	0.452	0.210
monetary shock	0.130	[0.001; 0.616]	0.434	0.312	0.140
taste shock	0.053	[0.001; 0.360]	0.656	0.485	0.214
mark-up shock	0.271	[0.004; 0.803]	0.290	0.194	0.087
Euro area (1970Q1-1998Q4)					
technology shock	0.308	[0.008; 0.776]	0.224	0.144	0.057
monetary shock	0.058	[0.001; 0.407]	0.623	0.472	0.221
taste shock	0.069	[0.001; 0.461]	0.587	0.429	0.198
mark-up shock	0.089	[0.001; 0.559]	0.525	0.382	0.178
United Kingdom (1972Q1-1992Q3)					
technology shock	0.073	[0.001; 0.500]	0.576	0.418	0.198
monetary shock	0.163	[0.002; 0.677]	0.383	0.265	0.114
taste shock	0.072	[0.001; 0.477]	0.577	0.423	0.193
mark-up shock	0.160	[0.002; 0.696]	0.401	0.286	0.128
Canada (1961Q1-1990Q4)					
technology shock	0.096	[0.001; 0.539]	0.511	0.362	0.160
monetary shock	0.176	[0.002; 0.677]	0.374	0.264	0.114
taste shock	0.082	[0.001; 0.559]	0.545	0.396	0.184
mark-up shock	0.074	[0.001; 0.539]	0.566	0.423	0.190

Table 6 Bootstrapped *p*-values[a] for Johansen's cointegration tests

	Trace test of the null of no cointegration between:		
	π_t and U_t	π_t and R_t	π_t, U_t and R_t
United States, pre-Volcker (January 1959-July 1979)	0.141	1.0e-4	0.000
Euro area (1970Q1-1998Q4)	0.023	0.568	0.005
United Kingdom (1972Q1-1992Q3)	0.094	NA[b]	NA[b]
Canada (1961Q1-1990Q4)	0.509	0.493	0.010
	Test of the null of one cointegrating vector, versus the alternative of two, for π_t, U_t and R_t:		
United States, pre-Volcker (January 1959-July 1979)	0.074		
Euro area (1970Q1-1998Q4)	0.012		
Canada (1961Q1-1990Q4)	0.668		

[a] Based on 10,000 bootstrap replications. [b] For the United Kingdom, results from the ADF tests reported in Table 1 strongly reject the null of a unit root in the short rate.

Table 7 Monte Carlo evidence on the performance of the Johansen procedure: fraction of simulations for which the bootstrapped trace statistic incorrectly rejects the null of no cointegration between two independent random walks at a given significance level[a]

Sample size (in quarters)	Fractions of bootstrapped p-values for Johansen's trace statistic below:		
	0.1	0.05	0.01
$T = 82$ *(U.S., pre-Volcker: 1959Q1-1979Q3)*	0.119	0.059	0.013
$T = 115$ *(Euro area: 1970Q1-1998Q4)*	0.118	0.060	0.012
$T = 81$ *(United Kingdom: 1972Q1-1992Q3)*	0.118	0.059	0.011
$T = 118$ *(Canada: 1961Q1-1990Q4)*	0.113	0.055	0.011

[a] Based on 10,000 simulations.

Table 8 Fraction of simulations for which Johansen's trace test incorrectly rejects the null of no cointegration between inflation and unemployment at a given significance level, based on estimated structural VARs

	Based on Classical SVARs with a single permanent inflation shock			*Based on Bayesian SVARs with four permanent inflation shocks*		
	Fractions of bootstrapped p-values for Johansen's trace statistic below:					
	0.1	0.05	0.01	0.1	0.05	0.01
	I: testing cointegration between π_t and U_t					
Euro area (1970Q1-1998Q4)	0.170	0.101	0.029	0.269	0.183	0.083
United Kingdom (1972Q1-1992Q3)	0.117	0.054	0.017	0.236	0.155	0.060
	II: testing cointegration between π_t, U_t, and R_t					
United States, pre-Volcker (January 1959-July 1979)	0.671	0.548	0.191	0.721	0.605	0.256
Euro area (1970Q1-1998Q4)	0.341	0.221	0.079	0.454	0.349	0.185
Canada (1961Q1-1990Q4)	0.409	0.318	0.104	0.514	0.402	0.222

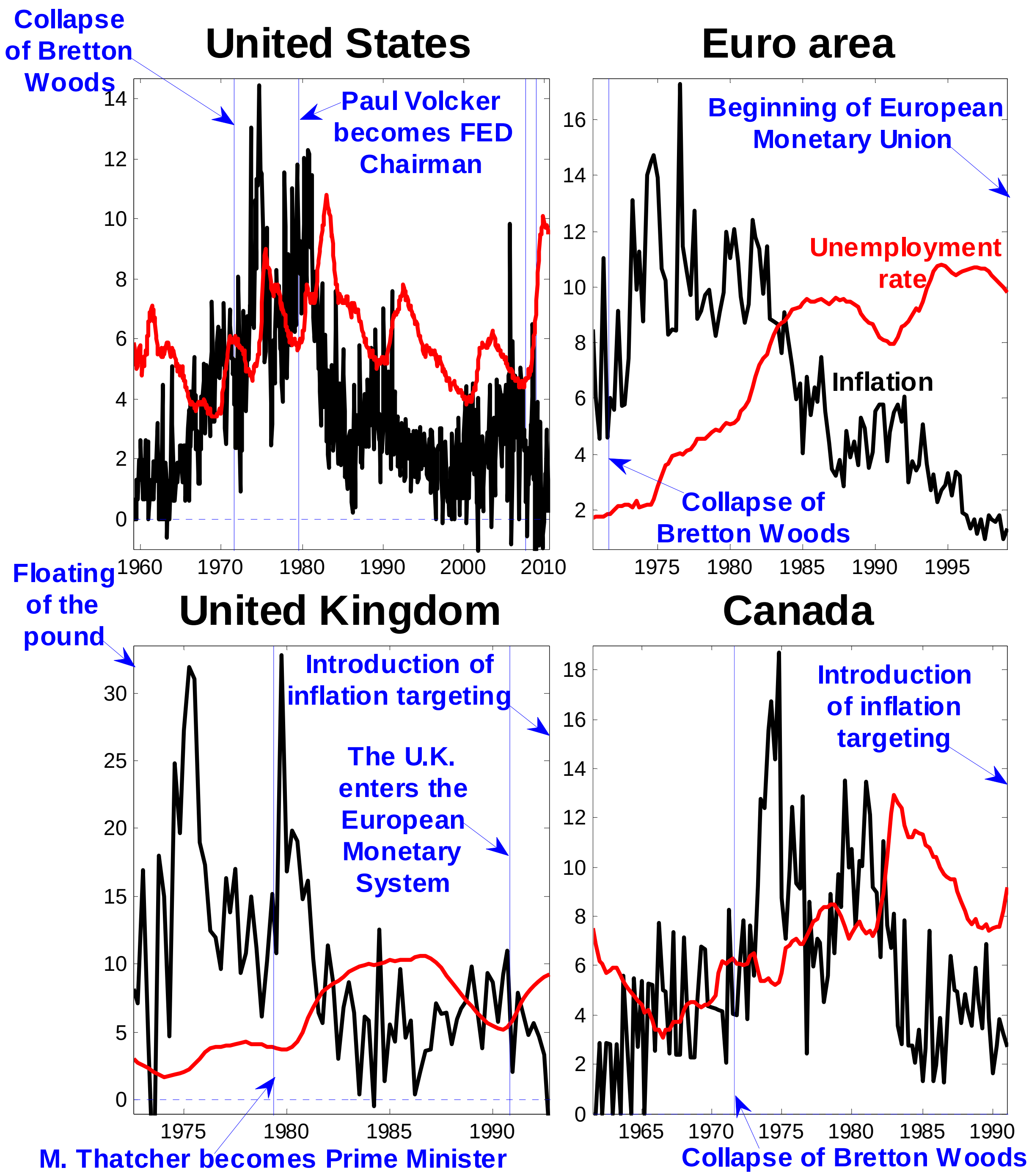

Figure 1 GDP deflator inflation and the unemployment rate

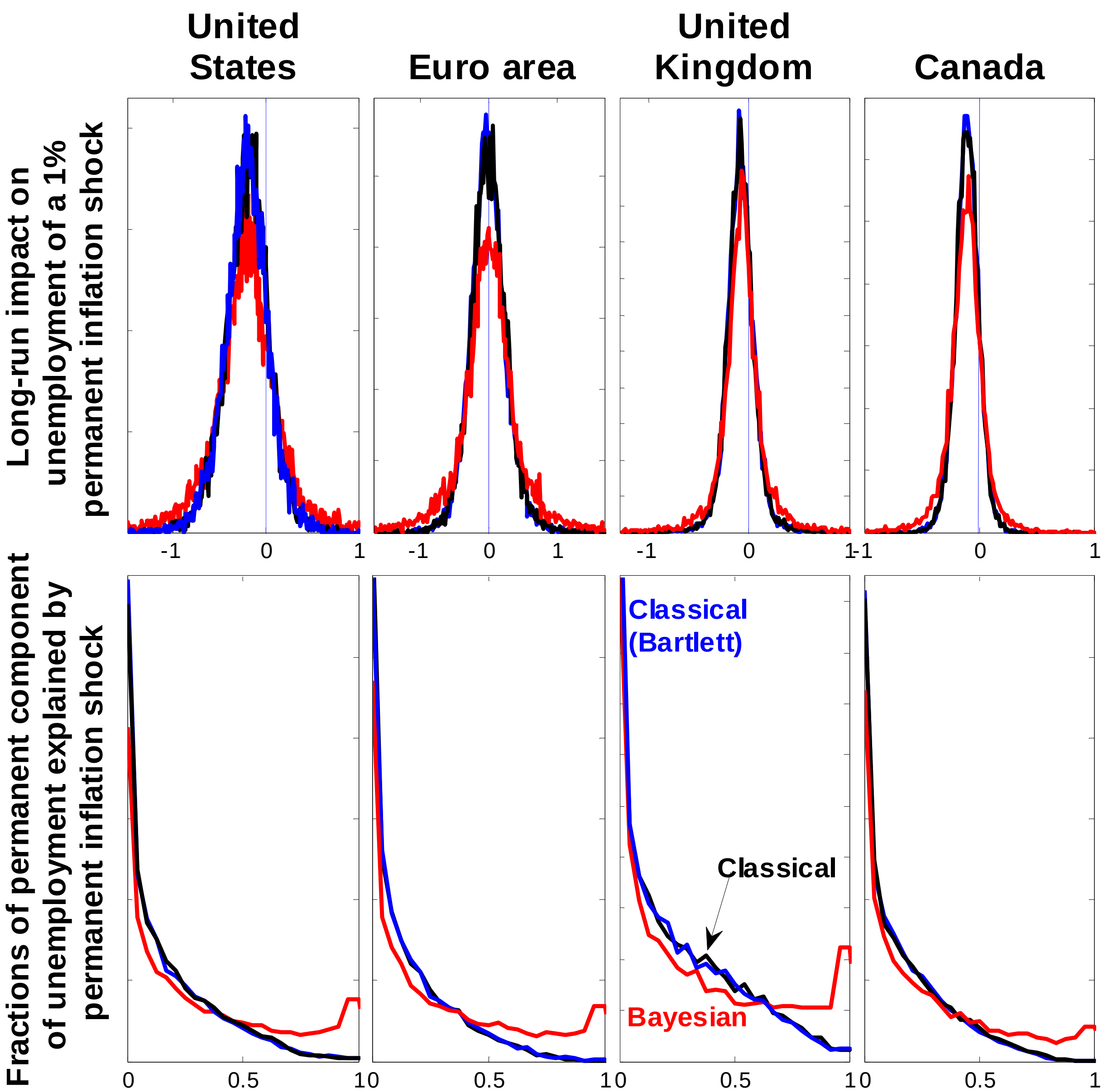

Figure 2 Results based on VARs allowing for a single permanent inflation shock: bootstrapped or posterior distributions of (*i*) the long-run impacts on the unemployment rate of a one per cent permanent shock to inflation, and (*ii*) of the fractions of the permanent component of unemployment explained by the permanent inflation shock

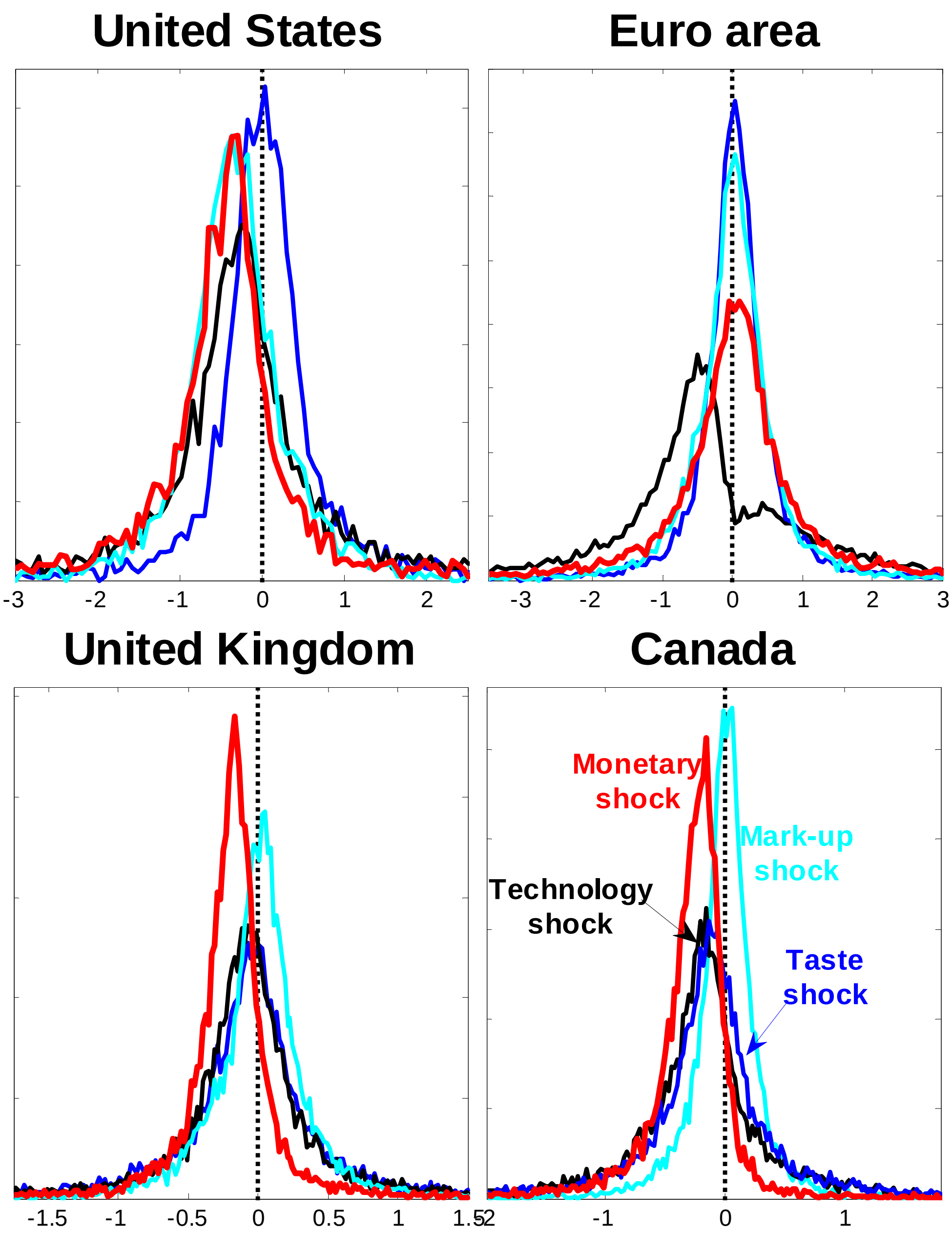

Figure 3 Results based on Bayesian VARs allowing for four permanent inflation shocks: posterior distributions of the long-run impacts on the unemployment rate of a one per cent permanent shock to inflation

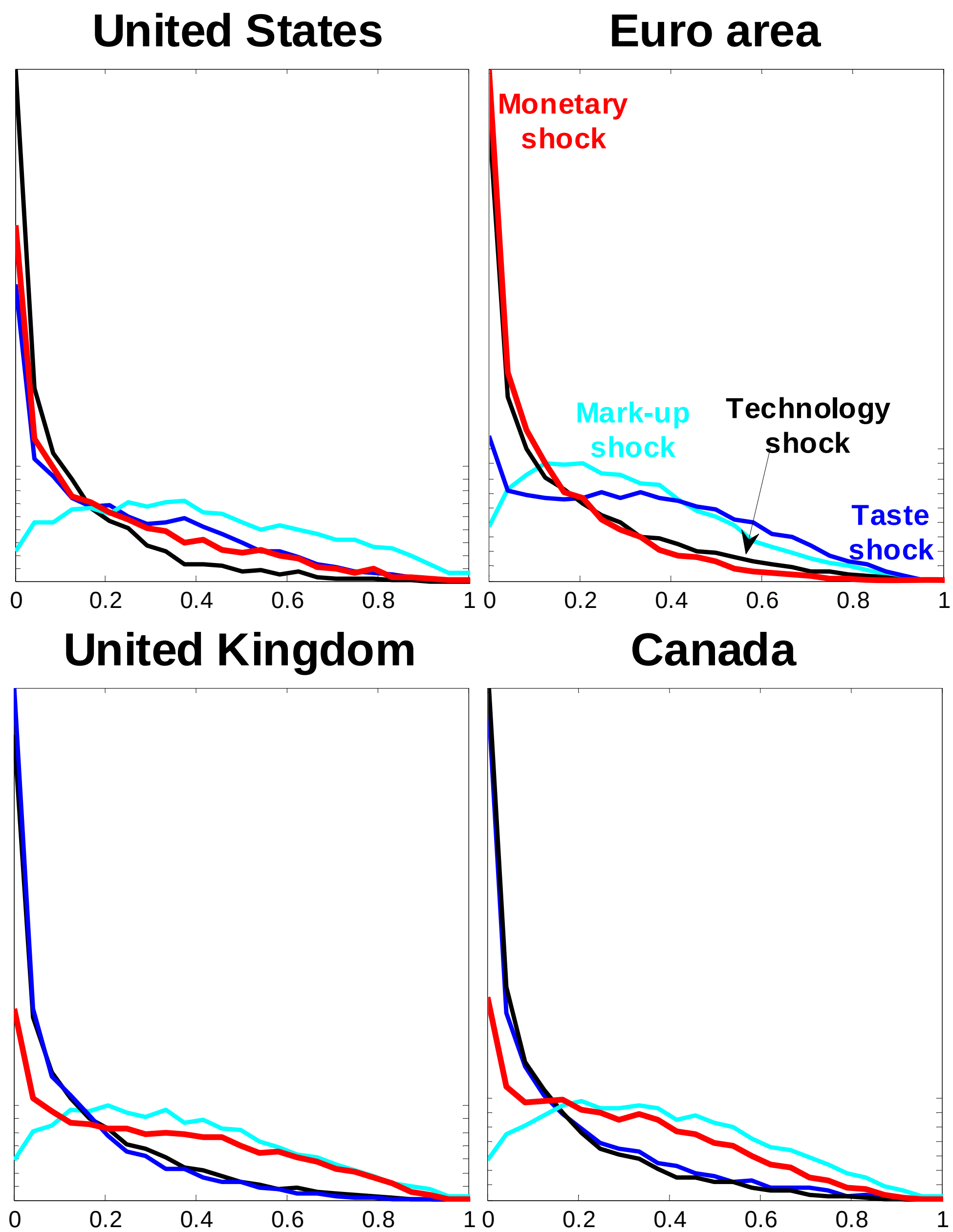

Figure 4 Results based on Bayesian VARs allowing for four permanent inflation shocks: posterior distributions of the fractions of the permanent component of inflation explained by individual shocks

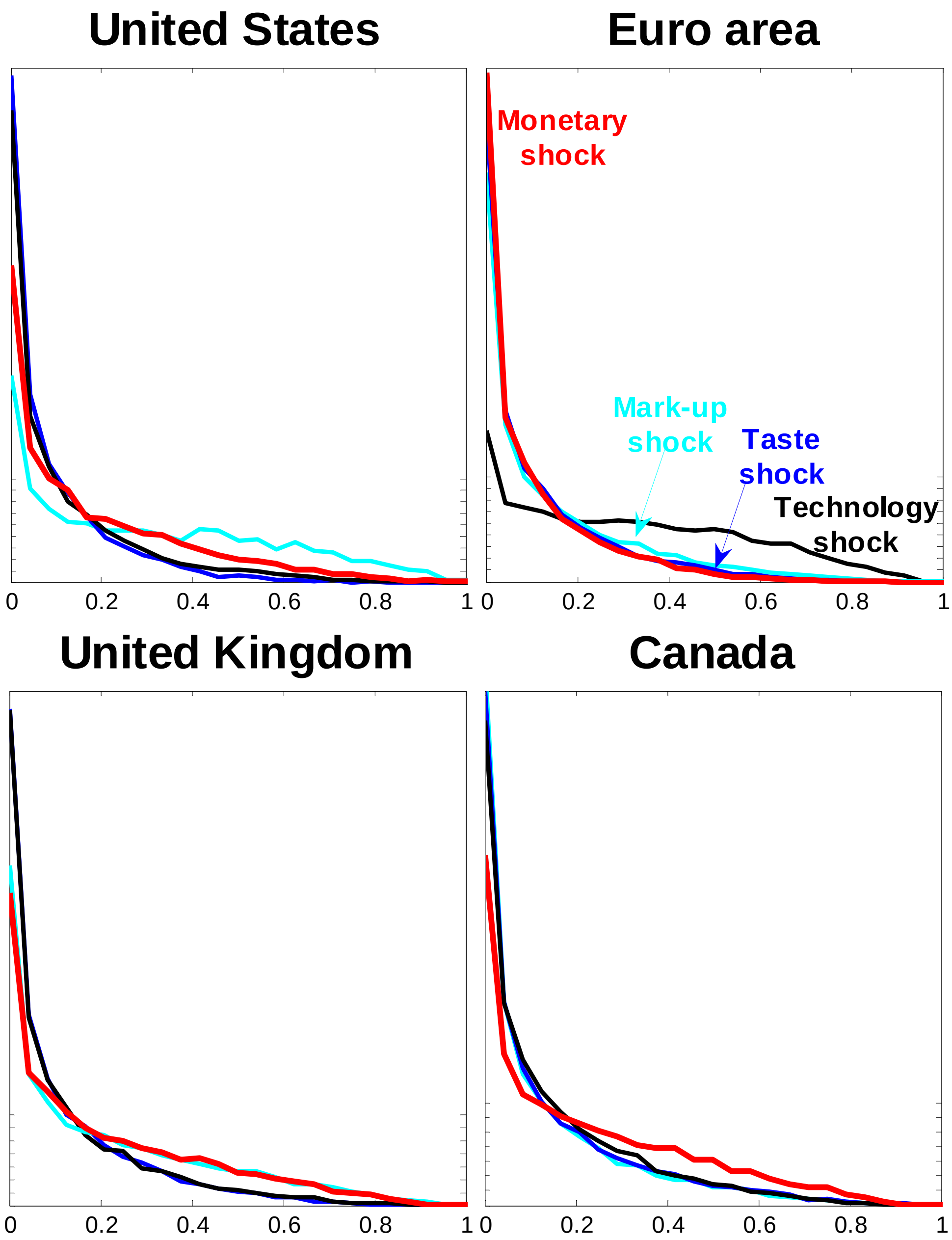

Figure 5 Results based on Bayesian VARs allowing for four permanent inflation shocks: posterior distributions of the fractions of the permanent component of the unemployment rate explained by individual shocks

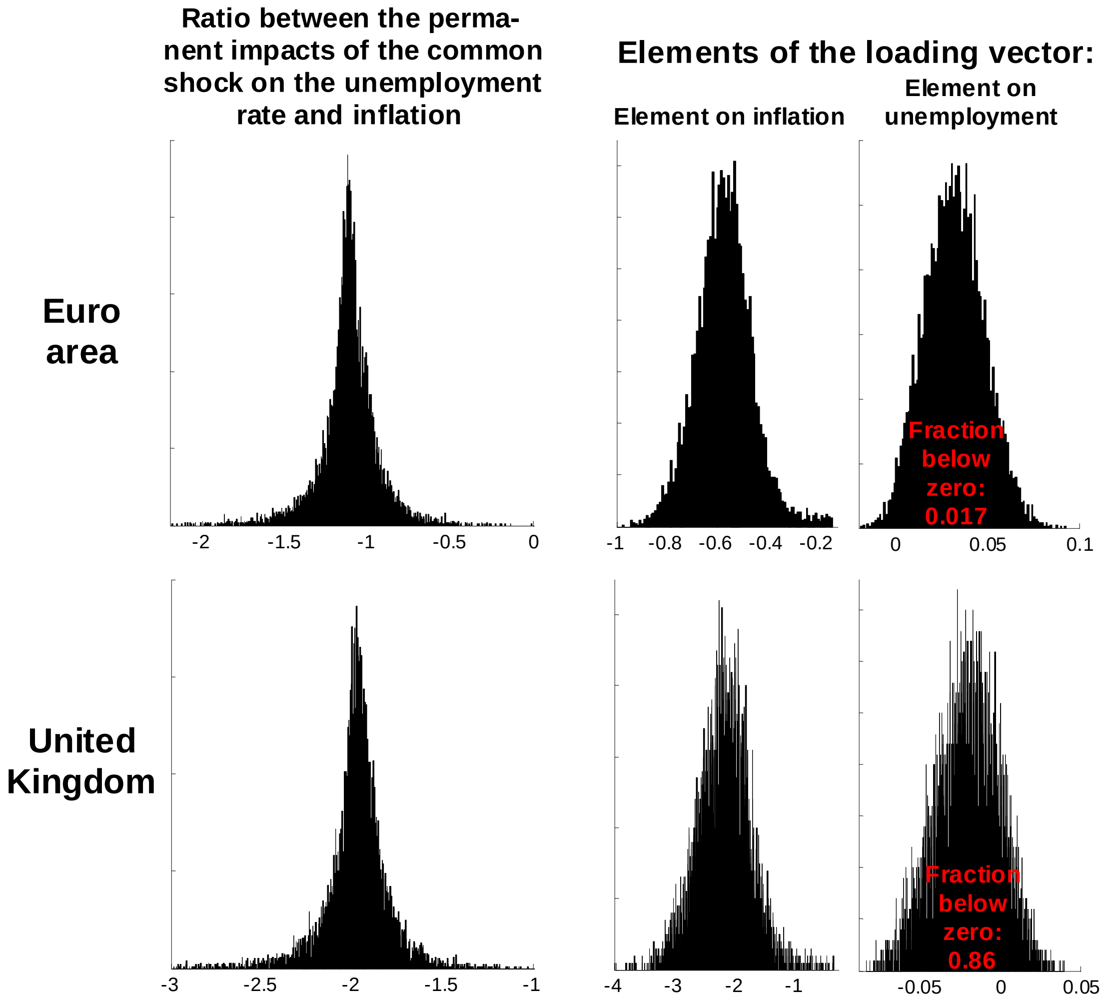

Figure 6 Results based on the Johansen procedure: bootstrapped distributions of the ratio between the permanent impacts of the common shock on the unemployment rate and inflation, and of the elements of the loading vector