THE EFFECTS OF A TECHNOLOGY SHOCK IN THE
EURO AREA

Nuno Alves
José Brandão de Brito
Sandra Gomes
João Sousa

Please address correspondence to Banco de Portugal, Economic Research Department
Av. Almirante Reis, no. 71 1150–012 Lisboa, Portugal;
João Sousa, tel: # 351-21-3130265, email: jmsousa@bportugal.pt

The Effects of a Technology Shock in the Euro Area[*]

Nuno Alves[†], José Brandão de Brito[†], Sandra Gomes[†], and João Sousa[†]

[†]Banco de Portugal

December 2005

Abstract

The aim of this paper is to estimate the effects of a technology shock in the euro area within a structural VAR framework. Since the impact of these shocks on labor use is a controversial issue in the related literature, we give particular attention to it. Given that the estimated effects of a technology shock are quite sensible to the low-frequency properties of the labor input measure, we resort to an extensive statistical analysis to investigate whether hours worked are better characterized as stationary or difference-stationary. We conduct a battery of classical unit root and stationary tests, analyze the small-sample properties of some of the tests-statistics, explore encompassing tests and Bayesian odds ratios to ascertain if the more appropriate VAR model is the one in which hours per capita enter in levels or first-differences. The evidence gathered is in support of hours being stationary, which leads to the conclusion that per capita hours worked rise after a technology shock in the euro area. As for the responses of the remaining variables, our results are in line with the bulk of the literature.

JEL Classification: E32; E24

Keywords: productivity, long-run restrictions, encompassing, hours worked

1 Introduction

Since Galí (1999) argued that a technology shock causes a decline in hours worked in the US, a new strand of the literature dedicated to the estimation of the impact of technology shocks on the labor input has emerged. The relevance of this topic stems from the real business cycle (RBC) paradigm prediction that technology shocks are able to generate the stylized positive co-movement of output, hours worked and productivity that characterize business cycles. Given

---

the strong pro-ciclycality of output and hours found in the actual data, the conjecture that hours fall after a technology shock would lead to the conclusion that technology shocks cannot be an important source of business cycles and, as a corollary, to the empirical irrelevance of the standard RBC model. This strong implication of Galí's finding, needless to say, spawned a large number of contributions aimed at scrutinizing the validity of his results. Against this background, the aim of this paper is to contribute to the debate on the empirical effects of technology shocks, particularly on the labor input, by looking at evidence for the euro area within a structural VAR framework. More specifically, we provide a new measure of hours worked in the euro area and apply a battery of statistical tests in quest for the more appropriate specification for our VAR with which the technology shocks and their dynamic effects can be estimated.

Many of the contributions in this strand of the literature use time series models combined with a minimal set of identifying restrictions supported by a wide array of theoretical models to identify technology shocks and subsequently estimate their impact on a set of macroeconomic variables. Within this category, the most common practice in estimating the effects of the technology shock seems to be to adopt a structural VAR approach in which the identifying restriction amounts to assume that labor productivity is solely driven by technology shocks in the long run[1]. Other contributions, such as Basu, Fernald and Kimball (2004) identify the technology shocks through an augmented-growth accounting exercise.

In spite of the numerous contributions to the literature, the debate seems far from settled. For the US alone, the evidenced collected by different authors oscillates between a clear 'Yes' and a clear 'No' to the question of whether technology shocks cause a decline in hours worked. On the 'Yes' camp, Galí (1999) and Galí and Rabanal (2005) using structural VARs on aggregate data find that in the short-run hours worked drop in response to a technology shock. Using a related approach, Francis and Ramey (2003a) reach an analogous result. Basu et al. (2004) pursue an alternative approach by calculating a direct measure of technology rather than estimating it from a VAR as the sole quantity that affects labor productivity in the long run. An immediate advantage of this methodology is that it isolates technology shocks from other factors that affect labor productivity in the long run, such as distortionary taxes on capital, variable capacity utilization, among others. They also conclude that hours fall in the short run after the occurrence of a technology shock. On the 'No' camp features prominently the work of Christiano et al. (2004b,a) and Altig et al. (2002), Altig, Christiano, Eichenbaum and Linde (2004). Using a structural VAR approach with data and identifying restrictions in all similar to Gali's contributions, these authors reach the conclusion that hours rise, instead of falling,

---

[1]Examples of this approach can be found in Galí (1999, 2004, 2005), Galí, López-Salido and Vallés (2002), Galí and Rabanal (2005), Altig, Christiano, Eichenbaum and Linde (2002), Christiano, Eichenbaum and Vigfusson (2004b,a). A close variant that identifies technology shocks as those with permanent effects on real wages rather than on labor productivity can be found in Francis and Ramey (2003a).

in the wake of a technology shock. Fisher (2002) distinguishes between investment-specific and neutral technology shocks, and lays out a methodology for estimating both within a structural VAR framework. Using aggregate data, Fisher finds that hours rise in the aftermath of a either investment-specific or neutral shocks.

Given that the above mentioned contributions use similar methodologies and data, with the exception of Basu et al. (2004), the relevant question is to understand what lies behind such disparate conclusions on the response of hours worked to technology shocks. The culprit seems to be the treatment of the low-frequency properties of the hours series. In fact, the contributions grouped into the 'Yes' camp all treat per capita hours as difference stationary, giving rise to a VAR specification henceforth referred as DSVAR, while those grouped in the 'No' camp argue in favor of hours being stationary in levels, and use a VAR specification henceforth referred as LSVAR. With this evidence, it seems that, at least for the contributions based on the structural VAR approach, the whole debate on this issue is centered around the assumption regarding the stationarity of hours. In this context, Christiano et al. (2004b) employ small-sample inference analysis on classical econometric techniques and Bayesian methods to argue that the hypothesis that per capita hours are stationary in levels is more appealing than the alternative of difference-stationarity for the case of the US.

The evidence available for non-US economies is less abundant and generally points to a decline of the labor input after a technology shock. That is the baseline conclusion reached in Galí (2005) for the G7 countries (with the exception of Japan), and Francis and Ramey (2003b) for the UK. In what concerns the euro area, which is the focus of our paper, Galí (2004) uses total employment as the measure of the labor input to also find that this variable drops after a positive technology shock. Once again, the labor input measure is treated as difference-stationary, which in face of the evidence for the US just described, might be the driving force behind the result of falling hours. Using a different methodology, one that employs sign restrictions as the means for identification of the VAR, Peersman and Straub (2004) challenge the results of Galí (2004) by finding that hours rise after a technology shock irrespective of hours being included in the VAR in levels or first-differences. In this context, this paper aims at contributing to the debate by taking the methodology laid down in Christiano et al. (2004b) and applying it to euro area data with two objectives in mind. First, to test which of the two specifications, the LSVAR and the DSVAR, is more appropriate and, second, to estimate the impact of a technology shock on the labor input measure as well as on a wider set of macroeconomic variables.

In what refers to the database, we use the Fagan, Henry and Mestre (2001) database, albeit an updated version, to backdate the official series for the earlier periods[2]. Also, we choose to measure the labor input with hours worked instead of the total employment measure used by Galí (2004) (see data appendix). Since a series for hours worked was not available for the euro

---

[2]Galí (2004) uses the original Fagan et al. (2001) database.

area, we constructed it. We reckon this as an improvement because per capita hours and not total number of employees is the relevant variable in most the general equilibrium models. In terms of the methodology, we employ the same structural VAR approach as Galí (2004), but a richer VAR specification. To anticipate the main results, we find that treating hours per capita as stationary in levels is more appropriate than as difference-stationary. Bearing on this result we estimate a LSVAR and find that hours per capita worked rise rather than fall after a technology shock. It turns out that, as for the US case, our results rest on the choice of the LSVAR as the 'correct' specification, since the DSVAR delivers the same qualitative results as Galí (2004).

The paper is structured as follows. In the following two sections, we briefly describe the identification strategy and the data. In section 4, we outline the main effects of a technology shock using alternatively the DSVAR and the LSVAR specifications. In section 5, we look in some detail to the low-frequency properties of the per capita hours series. In particular, we run some widely used unit root (and stationarity) tests and conduct a small-sample inference analysis to evaluate the statistical significance of the tests results. In section 6, we use bootstrapping and Bayesian techniques to carry out some encompassing tests pertaining to the DSVAR and the LSVAR specifications. Section 7 concludes.

2 Identification of the Technology Shock

In identifying the technology shock, we follow much of related literature by imposing the restriction that only technology shocks can affect labor productivity in the long run. In implementing it, we pursue the methodology advocated by Shapiro and Watson (1988).

The analysis is based on the following reduced-form VAR,

$$Y_t = \eta + B\left(L\right)Y_{t-1} + u_t, \; Eu_tu_t^{'} = V, \; u_t = Ce_t, \; Ee_te_t^{'} = I \tag{1}$$

where $B(L)$ is a polynomial of order q in the lag operator, L, u_t is the vector of the one-step-ahead forecast errors to Y_t, and e_t the vector of structural shocks. Notice that we are assuming that the one-step-ahead forecast errors are a linear combination of the structural shocks. The Y_t vector is defined as,

$$\underbrace{Y_t}_{9\times 1} = \begin{pmatrix} \Delta \ln\left(GDP_t/\text{Hours}_t\right) \\ \ln\left(\text{Hours}_t\right) \\ \Delta \ln\left(GDP\ \text{deflator}_t\right) \\ \ln\left(C_t/GDP_t\right) \\ \ln\left(I_t/GDP_t\right) \\ \text{Capacity Utilisation}_t \\ \ln\left(GDP_t/\text{Hours}_t\right) - \ln\left(W_t/P_t\right) \\ \text{Interest Rate}_t \\ \ln\left(GDP\ \text{deflator}_t\right) + \ln\left(GDP_t\right) - \ln\left(M1_t\right) \end{pmatrix}$$

$$\equiv \begin{pmatrix} \underbrace{\Delta y_t}_{1\times 1} \\ \underbrace{h_t}_{1\times 1} \\ \underbrace{X_t}_{7\times 1} \end{pmatrix} \tag{2}$$

The Shapiro and Watson (1988) procedure starts off with the following structural relationship:

$$\Delta y_t = \mu + \beta(L)\Delta y_{t-1} + \tilde{\alpha}_1(L)h_t + \tilde{\alpha}_2(L)X_t + \varepsilon_t^z \tag{3}$$

where Δ stands for the difference operator so that Δy_t is the change in the log of average labor productivity, which we assume covariance-stationary, $\beta(L)$, $\tilde{\alpha}_1(L)$ and $\tilde{\alpha}_2(L)$ are polynomials in the lag operator of orders $q - 1$, q and q, respectively, and ε_t^z captures the technology shocks. The vector X_t collects all the endogenous variables of the VAR apart from labor productivity and per capita hours.

The identifying restriction by which only technology shocks affect labor productivity in the long run amounts to imposing the following restrictions[3]:

$$\tilde{\alpha}_1(L) = \alpha_1(L)(1 - L); \ \ \tilde{\alpha}_2(L) = \alpha_2(L)(1 - L) \tag{4}$$

We introduce an additional restriction on $\tilde{\alpha}_2(L)$ so to ensure that the variables, interest rate and money velocity, appear with lags 1 to q rather than 0 to q. This additional restriction reflects our timing assumption that labor productivity reacts with a lag to shocks in both the interest rate and money velocity.

Substituting (4) into (3) yields the restricted structural equation, which is the object of interest when it comes to the estimation of the technology shocks, ε_t^z, and their effects on the variables of the VAR,

---

[3]For a more detailed exposition of the implementation of the Shapiro and Watson (1988) procedure see, for example, Christiano et al. (2004b) and Fisher (2002).

$$\Delta y_t = \mu + \beta(L)\Delta y_{t-1} + \alpha_1(L)\Delta h_t + \alpha_2(L)\Delta X_t + \varepsilon_t^z \tag{5}$$

To estimate equation (5) one cannot resort to OLS because in general X_t and ε_t^z will be correlated. We therefore employ instrumental variable estimation, using as instruments a constant, Δy_{t-s}, h_{t-s}, and X_{t-s}, for $s = 1, 2, \ldots, q$. With the estimated ε_t^z available, we only need the first column of the matrix C in (1) to compute the impulse responses entailed by a technology shock. To accomplish this we need to estimate the residuals, u_t, of the reduced-form VAR in (1). For this purpose, we choose $q = 4$. After having the u_t, the first column of C is obtained by regressing the u_t on the ε_t^z by OLS.

3   Data

3.1   Description

The variables used in this VAR analysis build on the original data series described in the data appendix and cover the period 1970:1-2004:3[4]. The variables are: productivity growth, per capita hours worked, inflation, consumption to output ratio, investment to output ratio, capacity utilization, labor productivity to real wage ratio, short term interest rate, money velocity. With the exception of the short term interest rate, all variables are expressed in natural logarithms.

One important novelty in this paper is the use of a quarterly series for hours worked in the euro area. As there is no official series of average hours per employee in the euro area we had to construct a new series. Per capita hours were thus obtained by multiplying average hours per persons in employment by total employment and then dividing by working age population. The remaining variables were constructed as follows. Productivity growth was obtained by taking first-differences of the log of the ratio between real GDP and per capita hours. Inflation was computed as the first-difference of the log of the GDP deflator. The consumption, investment, capacity utilization, and the short term interest rate variables were defined as in the original series in the Fagan et al. (2001) database. The nominal hourly wage was computed by dividing the total compensation per employee by the average hours per employee. The real wage resulted from deflating the nominal hourly wage with the GDP deflator. Money velocity was calculated using real GDP, the GDP deflator and the M1 monetary aggregate[5]. These series were then used to compute the variables that actually enter the VAR in (1). The plots are displayed in figure 1.

---

[4]The estimation of the VAR drops the first four observations to account for the fours lags used.

[5]For a detailed description of the series used to proxy the variables, refer to the data appendix.

3.2 Low-Frequency Properties of the Variables

A proper identification of the technology shock requires all variables in the VAR to be stationary. In order to ascertain the low-frequency properties of our data, we start by subjecting each variable in the VAR to the Augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) unit root tests[6]. The results of these tests are summarized in table 1[7].

Table 1: Unit Root Tests

Variable	Test	
	ADF	PP
Average Labor Productivity growth	-9,35***	-9,88***
Inflation	-0,99	-2,40
Consumption/GDP	-2,78*	-2,78*
Investment/GDP	-2,50	-1,71
Capacity Utilization	-3,98***	-3,40**
Average Labor Productivity/Hourly Real Wage	-0,77	0,45
Interest Rate	-1,66	-1,26
Money Velocity	-1,11	-1,38

'*', '**', '***', denote rejection at the 10%, 5%, 1% significance level, respectively.

On a first passage, we found productivity growth, the consumption to output ratio and capacity utilization to be stationary in levels. For the remaining variables, the outcome of the ADF and PP tests are in favor of nonstationarity of the series. The next step is to search for the presence of significant deterministic components.

Since graphical inspection and formal testing suggest that none, among the group of non-stationary variables, have a linear trend along the whole sample, we resort to the procedure proposed by Zivot and Andrews (1992) to test the null of a unit root against the alternative of stationarity about a broken linear trend. This procedure estimates a (sole) break date that maximizes the evidence against the null of nonstationarity, computes an ADF-type test-statistic and provides the relevant critical values[8]. The outcomes of the application of the Zivot and Andrews test to the group of nonstationary variables are summarized in table 2.

---

[6]The unit root tests results pertaining to per capita hours are not reported here, since the low-frequency properties of this series are looked at in detail in section 5.

[7]The reported tests included a constant, but no time trend.

[8]As suggested by Zivot and Andrews (1992), the break date is searched between the 15th and the 85th percentile of the sample.

Table 2: Tests of Unit Root v Broken Trend Stationarity

Variable	Zivot-Andrews Test	
	ADF-statistic	Break Date
Inflation	-6,23***	1997Q1
Investment/GDP	-4,25*	1981Q4
Average Labor Productivity/Hourly Real Wage	-4,72**	1974Q4
Interest Rate	-4,17*	1980Q1
Money Velocity	-2,07	1990Q3

'*', '**', '***', denote rejection at the 10%, 5%, 1% significance level, respectively.

Based on table 2, we take the variables, inflation, investment to output ratio, productivity to real wage ratio, and the interest rate, to be stationary about a broken linear trend. It follows that the broken trends have to be estimated and then removed before the relevant variables are included in the VAR. As regards money velocity, it seems that a broken linear trend is not enough to induce stationarity in this variable. A cursory look at the plot of the series suggests that a quadratic trend might be a better characterization of the deterministic components at play. In fact, the inclusion of a quadratic trend is found to be highly significant and to make detrended money velocity stationary[9].

To sum up, to estimate the VAR we use productivity growth, consumption to output ratio, capacity utilization as in their original series. For the variables, inflation, investment to output ratio, productivity to real wage ratio, and the interest rate, broken linear trends must be estimated and removed, whilst for money velocity a quadratic trend must be extracted. For the time being, we take an agnostic view as for the stationarity of per capita hours because the low-frequency properties of this variable is, as we will see, at the center of the controversy surrounding the effects of a technology shock on the labor input. Therefore, we start by running two versions of the VAR; one with per capita hours in levels and another with per capita hours in first-differences. Only in section 5 will we address the low-frequency properties of hours in a more definitive way.

4   Benchmark Results

In this section, we report the effects of a neutral technology shock in two alternative specifications of our structural VAR model. One in which per capita hours enter the VAR in levels

---

[9]The ADF test-statistic for money velocity after removing the estimated quadratic trend is -1.95.

(LSVAR), and another in which hours are first-differenced (DSVAR). The response of our set of variables after the occurrence of a technology shock for the DSVAR and LSVAR specifications are displayed in figures 2 and 3, respectively. In both figures, the solid lines depict our point estimates and the gray areas, their respective 95% confidence bands[10]. The responses of all variables are measured in percentage, except the interest rate's, which is measured in basis points. In the following description of the results we concentrate on the first 20 quarters after the shock.

We begin by looking at the results pertaining to the DSVAR. In figure 2, the effect of a one-standard deviation positive technology shock is to generate a steady increase in output that reaches roughly 0.8% after 20 quarters. Per capita hours endure what seems to be a permanent fall, a result similar to the one reported by Galí (2004) for the euro area[11]. There is also a positive permanent effect on real wages, consumption and investment, as expected. It is still worth noting that the impact on inflation, capacity utilization, money growth and the interest rate, is not statistically different from zero throughout the time horizon.

Turning to figure 3, we immediately conclude that the effects of a technology shock differ substantially when the specification adopted is the LSVAR. First, hours worked rise instead of declining. Second, inflation falls on impact and converges only gradually towards zero. Third, capacity utilization shows a hump-shaped positive response that is statistically significant in the first 10 quarters. The trajectories of the remaining variables are not too different from those obtained with the DSVAR.

To conclude, the choice between the DSVAR and the LSVAR specifications is not a simple matter of detail, since the predictions of the effects of a technology shock, particularly on per capita hours, inflation, and capacity utilization are completely different across the two alternative specifications. This implies that Gali's (2004) result that the labor input falls after a technology shock in the euro area only holds if the 'correct' specification for the structural VAR turns out to be the one that includes per capita hours in first-differences rather than in levels. Assessing this issue is what the remaining of the paper is devoted to.

5   Low-Frequency Properties of Hours Per Capita

As we have seen in the previous section, the effect of a technology shock on hours per capita hinges critically on the chosen VAR specification. For if we use the LSVAR, hours per capita rise after a technology shock, whereas the opposite result emerges when the DSVAR is adopted. To unveil which of the two specifications is more correct we need to look at the relative statistical appropriateness of each of them. In this context, it becomes crucial to properly ascertain the low-frequency properties of hours per capita.

---

[10]These bands were computed by bootstrap simulation using 2.000 draws.

[11]This is so in spite of Galí (2004) using employment rather than hours as the measure of labor input.

We start by looking at three popular univariate unit root testing procedures: the ADF, the PP, and the generalized least squares Dickey-Fuller test proposed by Elliot, Rothenberg and Stock (1996) (hereafter ERS), to test the null hypothesis of unit root in hours per capita. We reject this hypothesis at the 5% and 1% significance levels, for the ADF and PP tests, respectively. Interestingly, the outcome of the ERS test is opposite to the two previous, since we fail to reject the null at the 10% significance level[12]. This a rather puzzling result, since the ERS test is generally more powerful than the ADF or the PP tests, especially for highly persistent, albeit stationary, series. In face of this contradicting evidence we investigate this issue further[13].

As suggested by Christiano et al. (2004b), we consider a strand of unit root tests that includes stationary variables that are correlated with the subject of the test in the unit root testing equation. Hansen (1995) and Elliot and Jansson (2003) propose two different but related tests that include covariates in the unit root testing equation and show that this results in potentially high power gains, the more so the more relevant the covariates turn out to be[14]. Although the test proposed by Elliot and Jansson (2003) is more powerful than the covariate augmented Dickey-Fuller (CADF) test proposed by Hansen, we use the latter for the same two reasons that were laid out in Christiano et al. (2004b). The first is that the CADF test is better sized than the Elliot and Jansson's, the second is that, in our particular context, the CADF test corresponds to the test for weak instruments that is relevant for our instrumental variable approach to VAR identification.

As suggested by Elliot and Jansson (2003), in the context where the data-generating process (DGP) is a VAR, it is natural to use the endogenous variables of the VAR as covariates for the unit root test of any specific variable. Thus, to implement the CADF unit root test on hours, we use the standard ADF equation augmented with the endogenous variables of our VAR. Of the two alternative specifications, we drew the variables from the DSVAR because that is the one where hours are treated conformably with the unit root null hypothesis. Bearing in mind the notation of equation (2), to implement the test we run a regression of Δh_t on a constant, h_{t-1}, and the predetermined variables of the VAR instrumental variables regression, Δh_{t-s}, for $s = 1\ldots 3$, Δy_{t-s}, X_{t-s}, $s = 1\ldots 4$. The test statistic for the CADF test is just the t-ratio on the coefficient of h_{t-1}. The null hypothesis is then that this test statistic is equal to zero.

---

[12]For the ADF test with a constant and 2 lags, the test statistic yields -2.98, and the correspondent critical value to a 5% significance level is equal to -2.88. For the PP test, the test statistic, with a bandwidth of 7, yields -3.60, and the correspondent critical value to a 1% significance level is equal to -3.48. For the ERS test, the test statistic, with 2 lags, yields 0.35, and the correspondent critical value to a 10% significance level is equal to -1.61.

[13]We also ran the Zivot and Andrews (1992) test on per capita hours and found that we could not reject the null of unit root around a broken linear trend at the 10 % significance level.

[14]Small-sample Monte Carlo simulation presented in Elliot and Jansson (2003) show that the inclusion of relevant covariates yields large power gains in both these tests relative to the ADF test. Relative to the ERS test, significant power gains are uniformly achieved by the Elliot-Jansson test, whereas for the Hansen test gains are only secured for sufficiently relevant covariates.

To study the small-sample significance of the t-ratio we resort to bootstrapping. In particular, we simulate 5,000 artificial data sets, using the DSVAR as the DGP. For each data set, we compute the t-ratio on the coefficient of h_{t-1} in the testing equation underlying the CADF test. From the empirical sampling distribution of the t-ratios, we compute the p-value associated with the test statistic pertaining to the actual data[15]. It turns out that the test statistic of the actual data has a *p-value* of 0.02%, so that we have no doubt in rejecting the null hypothesis that hours per capita contain a unit root.

To complement the low-frequency analysis of per capita hours, we use the Kwiatkowski, Phillips, Schmidt and Shin (1992) (hereafter, KPSS), to test the null hypothesis of stationarity against the alternative of unit root. Using the asymptotic critical values provided by KPSS, we reject the null of stationarity at the 1% significance level[16]. These results conflict with the inference stemming from the ADF, PP, and CADF tests. However, as shown by Caner and Kilian (2001) the use of the KPSS asymptotic critical values to highly persistent variables in finite samples may cause extreme size distortions, which results in rejecting the null hypothesis too often. In face of this, the use of size-corrected critical values can be an important robustness check. As illustrated in Caner and Kilian (2001), using size-corrected critical values might impart significantly on the outcome of the KPSS test. Typically, the strategy used to compute the size-corrected critical values consists of using a DGP deemed appropriate for the variable under analysis[17]. Since in the overall exercise we are assuming that the model economy is characterized by a VAR, we should use this as our DGP when constructing the small-sample, size-corrected critical values for the KPSS test. As for the VAR specification, we use the LSVAR because this is the one specification that conforms with the null hypothesis of stationarity entailed by the KPSS test.

In order to obtain the size-corrected critical values in our framework, we generate 5,000 artificial data sets by bootstrapping and for each data set compute the KPSS test-statistic, using a lag truncation of 9 for the Newey-West Bartlett kernel. Using the resulting empirical sampling distribution, we compute the 90th, 95th, and 99th percentiles to get (simulated) critical values for our test. This yields (simulated) critical values for the 10%, 5%, and 1% significance levels, of 1.36, 1.39, and 1.42, respectively. Since the test-statistic for the actual data is found to be 1.25, we clearly fail to reject the null of stationary hours per capita. Notice that, as remarked by Caner and Kilian (2001), the small-sample size-correction has the consequence of

---

[15]The test statistic for the actual data is -4.82. For reference, the simulated critical values for the first and fifth percentile were found to be -3.45 and -2.62, respectively.

[16]The test statistic, with a bandwidth of 9, yields 1.25, which compares to asymptotic critical values of 0.35, 0.46, and 0.74, for 10%, 5%, and 1% significance level, respectively.

[17]Caner and Kilian (2001), working with an AR(1) with autocorrelation coefficient of 0.985 as DGP, found that correcting the critical value corresponding to the KPSS test for the 5% significance level for size distortion, would require changing from the asymptotic level of 0.46 to the size-corrected level of 0.75, which is much closer to the value of the test statistic found for the actual data on hours per capita.

taking away a significant amount of power to the KPSS test. This warrants some caution in taking conclusions based on this test alone.

On the whole, our analysis on the low-frequency properties of hours per capita based on classical econometric methods points to the conclusion that this variable is stationary. In fact, both the ADF and PP tests reject the null hypothesis of the presence of a unit root in hours per capita. The fact that the ERS test delivered the opposite conclusion was downplayed by the result of the CADF test, which being more powerful than the ERS for sufficiently relevant covariates, leads to rejection of the null of unit root. Finally, a small-sample analysis of the KPSS test reversed the outcome of the test based on the asymptotic critical values, re-enforcing the conjecture that hours per capita are indeed stationary. The overall conclusion that hours per capita ought to be stationary has the implication that the LSVAR is the more appropriate specification for our VAR analysis. We proceed with the investigation by shifting our attention away from the hours series and towards the relative performance of the two models in contention: LSVAR and DSVAR.

6   Encompassing Tests

Christiano et al. (2004b) put forward the encompassing criterion upon which both VAR specifications should be judged. The idea is simple. A good model is one that is able to replicate, or encompass, the results of the opposing model. In our context, this means that for the LSVAR to be convincing as the model that underlies the data under analysis, it should be able to generate data that when applied to the DSVAR would yield the same qualitative results as the DSVAR in the actual data. Similarly for the DSVAR. In what follows, we pretty much apply the procedure for the encompassing tests developed in Christiano et al. (2004b) to our specific problem.

6.1   Theoretical Aspects to Encompassing

Before going into the quantitative analysis it is worth noticing that, as shown in Christiano et al. (2004b), on a priori grounds the LSVAR should do better on the encompassing test. In fact, if the LSVAR is the true DGP, then using first-differences of hours per capita amounts to over-differencing, which entails a specification error[18]. If, otherwise, the DSVAR is right, then using hours in levels corresponds to not imposing a true restriction, which by itself does not entail specification error because that restriction is not outside the feasible parameter space. However, once finite-sample considerations are taken on board, the picture is less clear.

Another issue has to do with sampling uncertainty. If the DSVAR is correct, an analyst that mistakenly picks the LSVAR incurs in a weak instruments type of problem. That is because, in estimating the instrumental variables regression (5) in the context of the LSVAR, one takes

---

[18]Specification error here means that the true parameters can not be recovered, i.e. are outside the feasible parameter space.

lagged levels of hours to instrument the first differences of hours. If hours have a unit root, lagged levels of hours carry very little information for the first-differences of hours, i.e. we have a weak instruments problem. As we know, weak instruments entail not only large sampling uncertainty, but also bias. Both effects might, only if by accident, favor the DSVAR in the encompassing 'race'. As described below, this relation between the presence of a unit root in hours and the weak instruments problem constitutes a bridge between the classical econometric analysis pursued in the previous section and the encompassing approach.

6.2 Encompassing Results

Here, we evaluate the encompassing performance of each VAR specification in turn. In particular, we analyze whether each of the specifications, taken as the true DGP, is able to replicate the alternative's specification pattern of the responses to a technology shock for each variable included in the VAR.

6.2.1 Taking the LSVAR as the True DGP

To assess the ability of the LSVAR in encompassing the response of the VAR variables to the technology shock entailed by the DSVAR, we proceed as follows. We first generate 5,000 artificial data sets with the length of our actual sample data using the LSVAR parameterization, estimated from the actual data. For each artificial data set, we 'incorrectly' take the DSVAR as correct, and so first-difference each of the per capita hours artificial series. With it, we estimate the responses of all the variables to the technology shock implied by the DSVAR and take averages for the responses of each variable for each of the 20 quarters after the shock. These averages represent the point estimates of the responses pertaining to the 'incorrect' DSVAR specification when the the data is generated by the 'true' LSVAR specification.

In figure 4, the line with circles denote the average responses of the DSVAR with simulated data, whereas the thick line corresponds to the responses of the DSVAR in the actual data. The line with triangles represents the responses of the LSVAR in the actual data. The 95% confidence bands apply to the line with circles. Focusing, for now, on the response of hours, it is clear from figure 4 that the line with circles is very close to the thick line. This means that the mean simulated impulse responses of the DSVAR are very similar to the responses estimated from the actual data. In other words, the data generated by the LSVAR when applied to the DSVAR yields the same kind of response as the actual data. Crucially, hours are estimated to fall after a technology shock using the DSVAR when they are predicted to rise in the 'true' model. Moreover, for most of the remaining variables, the average estimated responses (lines with circles) move close together with the ones obtained for the actual data (thick lines), and for all of them, the actual data responses lie within the confidence bands of the simulated data.

Overall, it seems that the LSVAR does a good job in encompassing the DSVAR, since it

is able to generate data that when applied to the DSVAR yields results very similar to those obtained in the actual data.

6.2.2 Taking the DSVAR as the True DGP

In this section we employ a procedure similar to the one above, the only difference being that we take the DSVAR to be the 'true' model and the LSVAR the 'alternative' model. This means that we generate the same 5,000 artificial data sets by simulating the previously estimated DSVAR. We then accumulate the resulting growth rate of hours to obtain the corresponding level series that enters the LSVAR.

Figure 5 displays the impulse responses of this exercise. The line with circles corresponds to the average responses pertaining to the LSVAR model obtained from the simulated data sets. The line with triangles denotes the responses of the 'true' DSVAR model in the actual data, and the thick line represents the responses of the 'alternative' model in the actual data. The confidence bands apply to the line with circles. Focusing first on the response of hours, the average response of the LSVAR obtained with the simulated data is significantly apart from the one obtained from the actual data. In particular, the prediction based on the simulated data implies a fall in hours worked after a technology shock, when hours are estimated to rise in the actual data. On its own, this result implies that the DSVAR does a poor job in encompassing the LSVAR. However, the responses pertaining to the LSVAR in the actual data (thick line) fall within the confidence region of the one based on the simulated data. This only happens because the confidence bands are very wide. Nevertheless, strictly speaking it cannot be said that the DSVAR model does not encompass the LSVAR model. This large sampling uncertainty stems from treating hours as containing a unit root, which leads to a weak instruments problem. This problem will surface regardless of the appropriateness of the DSVAR specification.

One way to assess whether this weak instruments problem is an intrinsic feature of the underlying population model or is coming from incorrectly imposing the DSVAR as DGP, is to run a weak instrument test on the actual data. Failure in rejecting the weak instruments hypothesis is evidence that the sampling uncertainty is being driven by the use of the LSVAR when the DSVAR is the correct specification. Conversely, rejection of the weak instruments hypothesis means that the sampling uncertainty follows directly from the DSVAR prediction that working with the LSVAR unavoidably leads to a weak instruments problem. Recall that with the LSVAR, we use lags of the levels to instrument the contemporaneous values of the first differences. In our case, we are particularly interested in knowing if the incremental information in the level of hours lagged one period about the first-difference of hours is close to zero, i.e. if we have a weak instruments problem. That ought to be the case if the DSVAR were correct. As noted by Christiano et al. (2004a) the weak instrument test coincides with Hansen's CADF unit root test, which means that rejecting the null of lagged level of hours being a weak instrument

for differenced hours amounts to reject the null that hours contain a unit root and vice versa.

The setup of the weak instrument test is exactly the same of the CADF unit root test described in section 5. The only difference is that the weak instruments test is an F-test, rather than a t-test, on the coefficient of h_{t-1}. The test statistic for the actual data was found to be 23.3. In order to analyze the significance of this value, we once again employ a bootstrap procedure that consists of using the DSVAR to generate 5,000 artificial data with the length of our actual data. From the resulting empirical sampling distribution we compute the *p-value* of the test statistic of the actual data to be 0.02%, so that we clearly reject the null hypothesis that lagged levels of hours are a weak instrument for the first difference of hours[19]. It thus seems fair to argue that the ability of the DSVAR in encompassing the predictions of the LSVAR, albeit minute, rests on a type of the weak instruments problem that is purely artificial.

From all the evidence amassed so far, we take the view that the LSVAR is a better description of the problem in hand. The next issue is to evaluate statistically by how much.

6.2.3   Model Comparison

As we have seen, in a strict sense, the two alternative specifications encompass each other. That is because the responses of the variables to a technology shock in the actual data fall always within the confidence region of the responses entailed by the simulated data. We also concluded that the LSVAR tracks much closer the results of the DSVAR in the actual data than the contrary. As in Christiano et al. (2004*b*), we take a Bayesian approach to model comparison, by calculating the posterior odds ratio between the two specifications to get an idea of the relative plausibility of the models.

The model that best describes the underlying reality is the one that is able to generate data that, when applied to each of the specifications, is able to mimic the corresponding impulse response functions. In particular, in the wake of a technology shock, hours must rise when the LSVAR is adopted, and fall for the DSVAR. The aim is to assess through simulation how frequently does each of the alternative DGP lead to the event that the hours response to the technology shock is positive when applied to the LSVAR and negative when applied to the DSVAR. One way to operationalize this test is to consider the average response of hours in the first six periods for the LSVAR (denoted by μ_h) and for the DSVAR (denoted by $\mu_{\Delta h}$). Thus, the event we want both models to explain is the joint occurrence of a positive such average for the LSVAR and a negative one for the DSVAR. Let that event be denoted by Q, so that $Q = (\mu_h > 0, \mu_{\Delta h} < 0)$.

To study the relative performance of each model taken as DGP to predict Q, we generate 5,000 artificial data sets using each of the alternative specifications and count the frequency of

---

[19]Using the 99th percentile of the empirical sampling distribution, we found the critical value pertaining to the 1% significance level to be 12.3.

15

correct predictions of Q. The results can be written as:

$$P(Q|L) = 0.916$$
$$P(Q|D) = 0.271$$

where L and D denote the LSVAR and the DSVAR, respectively, $P(Q|L)$ denotes the marginal likelihood of Q given that the DGP is L, and similarly for D. Assuming equal priors for both models, i.e. $P(L) = P(D) = 0.5$, we can compute the posterior odds in favor of model L as:

$$\frac{P(L|Q)}{P(D|Q)} = \frac{P(Q|L)P(L)}{P(Q|D)P(D)} = 3.374$$

This means that the LSVAR is 3.4 times more likely to be the true DGP than the DSVAR.

7   Final Remarks

Our main results clearly favor the LSVAR as the most appropriate specification, something that is plainly illustrated by the 3.4 Bayes odds ratio in favor of the LSVAR. That in turn suggests that hours worked rise in the wake of a technology shock in the euro area, which constitute a challenge to the results reported in Galí (2004) and also Smets and Wouters (2003). Although these two contributions use a different measure for labor input, that is unlikely to be the main driving force behind the disparity between the two sets of results. More plausibly, as in the debate for the US, the main source of the disagreement is on the treatment given to the low-frequency properties of the labor input measure.

Given the sensibility of the estimated effects of a technology shock to the low-frequency properties of the labor input measure, we resort to a thorough statistical procedure to investigate the issue of whether hours are better characterized as stationary or difference-stationary. We conduct a battery of classical unit root and stationary tests, analyze the small-sample properties of some of the tests-statistics, explore encompassing tests and Bayesian odds ratios to ascertain if the more appropriate VAR model is the one in which hours per capita enter in levels or first-differences. The evidence gathered is in support of hours being better described as stationary in levels, which leads to the conclusion that per capita hours worked rise after a technology shock in the euro area. This result might become important when it comes to the calibration or estimation of macro models for the euro area.

References

Altig, D., Christiano, L., Eichenbaum, M. and Linde, J. (2002), Technology shocks and aggregate fluctuations. mimeo.

Altig, D., Christiano, L., Eichenbaum, M. and Linde, J. (2004), Firm-specific capital, nominal rigidities and the business cycle. mimeo.

Basu, S., Fernald, J. and Kimball, M. (2004), 'Are technology improvements contractionary?', *NBER working paper* **10592**.

Caner, M. and Kilian, L. (2001), 'Size distortions of tests of the null hypothesis of stationarity: Evidence and implications for the PPP debate', *Journal of International Money and Finance* **20**, 639–657.

Christiano, L., Eichenbaum, M. and Vigfusson, R. (2004*a*), 'The response of hours to a technology shock: Evidence based on direct measures of technology', *Journal of the European Economic Association* **2**, 381–395.

Christiano, L., Eichenbaum, M. and Vigfusson, R. (2004*b*), What happens after a technology shock? mimeo.

Elliot, G. and Jansson, M. (2003), 'Testing for unit roots with stationary covariates', *Journal of Econometrics* **115**, 75–89.

Elliot, G., Rothenberg, T. J. and Stock, J. H. (1996), 'Efficient tests for an autoregressive unit root', *Econometrica* **64**, 813–836.

Fagan, G., Henry, J. and Mestre, R. (2001), An area wide (AWM) model for the euro area, Working Paper 42, European Central Bank.

Fisher, J. (2002), 'Technology shocks matter', *Federal Reserve Bank of Chicago working paper* **14**.

Francis, N. and Ramey, V. (2003*a*), Is the technology-driven real business cycle hypothesis dead? shocks and aggregate fluctuations revisited. mimeo.

Francis, N. and Ramey, V. (2003*b*), The source of historical economic fluctuations: An analysis using long-run restrictions. mimeo.

Galí, J. (1999), 'Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations?', *American Economic Review* **89**, 249–271.

Galí, J. (2004), 'On the role of technology shocks as a source of business cycles: Some new evidence', *Journal of the European Economic Association* **2**, 372–380.

Galí, J. (2005), Trends in hours, balanced growth and the role of technology in the business cycle,. mimeo.

Galí, J., López-Salido, D. and Vallés, J. (2002), 'Technology shocks and monetary policy: Assessing the Fed's performance', *NBER working paper* **8768**.

Galí, J. and Rabanal, P. (2005), 'Technology shocks and aggregate fluctuations: How wll does the RBC model fit post war U.S. data?', *NBER Macroeconomics Annual* .

Hansen, B. (1995), 'Rethinking the univariate approach to unit root testing: Using covariates to increase power', *Econometric Theory* **11**, 1148–1171.

Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992), 'Testing the null hypothesis of stationarity against the alternative of a unit root', *Journal of Econometrics* **54**, 159–178.

Peersman, G. and Straub, R. (2004), Technology shocks and robust sign restrictions in a euro area SVAR, Working Paper 373, European Central Bank.

Shapiro, M. and Watson, M. (1988), 'Sources of business cycle fluctuations', *NBER Macroeconomics Annual* pp. 111–148.

Smets, F. and Wouters, R. (2003), 'An estimated dynamic stochastic general equilibtium model of the euro area', *Journal of the European Economic Association* **1**, 1123–1175.

Zivot, E. and Andrews, D. W. K. (1992), 'Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis', *Journal of Business and Economic Statistics* **10**(3), 251–270.

Data Appendix

The data used in the paper refers to the current twelve euro area member states. For periods prior to 1999, data are an aggregation of the available country series. This advises some caution with data in beginning of sample, not only due to methodological considerations but also because country data availability becomes scarcer as we move back in time. Data covers the period from the first quarter of 1970 to the third quarter of 2004.

As far as possible we used official statistical sources, such as the Eurostat, the ECB, the European Commission and the OECD. However, euro area series at a quarterly frequency are often available for only a relatively short time-span and we had to backdate a number of series. To do this we relied mostly on the database by Fagan *et al.* (2001) (hereafter Area-Wide Model (AWM) database). This was the case of the Eurostat national account series in volume, which only start in 1991 and therefore had to be chain-linked backwards. Regarding national accounts deflators, the Eurostat series were chain linked with ECB data, which corrects for exchange rate variations among member countries in the period prior to 1999. Inflation is measured as the year-on-year rate of change of the GDP deflator. Data on compensation per employee are published by the ECB. The series start in the first quarter of 1991, and were chain linked with data from the AWM database. Euro area capacity utilization series is published by the European Commission and is available since 1985. For the previous period we constructed a *proxy* for the euro area aggregate based on available data for member countries. The monetary aggregates series are published by the ECB. The short term interest rate series used is the three-month Euribor provided by *Bloomberg* and for periods before 1999 we used data from the AWM database.

Since there is no official series of hours worked in the euro area we had to construct a new series. To do this we used country data on average hours worked per person in employment published by the OECD. However, and as mentioned by the OECD, there are significant differences in the sources and coverage of national data, implying that comparisons of the level of average hours worked across countries are probably not suitable. Therefore, we aggregated the quarterly rates of change of country series based on the euro area structure of employment (across countries) to get an index of average hours worked in the euro area. The behavior of the constructed series is reasonable, namely when compared with the behavior of the ECB's estimate of euro area average hours worked that was used on a box published in the October 2004 Monthly Bulletin[20], in terms of its annual rate of change.

---

[20]The ECB used annual data from the European Labor Force Survey, which is available only at annual frequency and for a relatively short time-span.

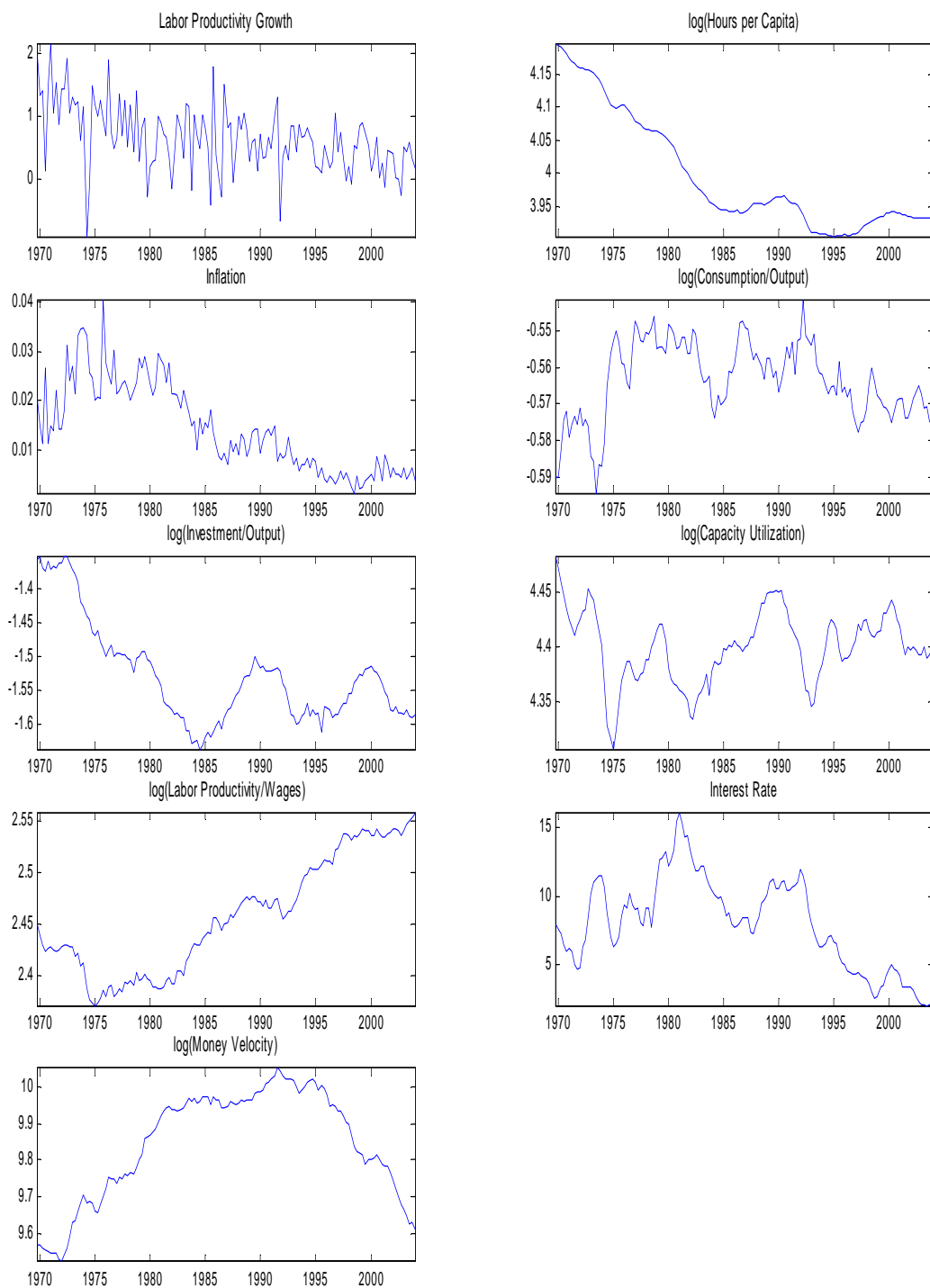Figure 1: Raw data on the variables used in the VAR

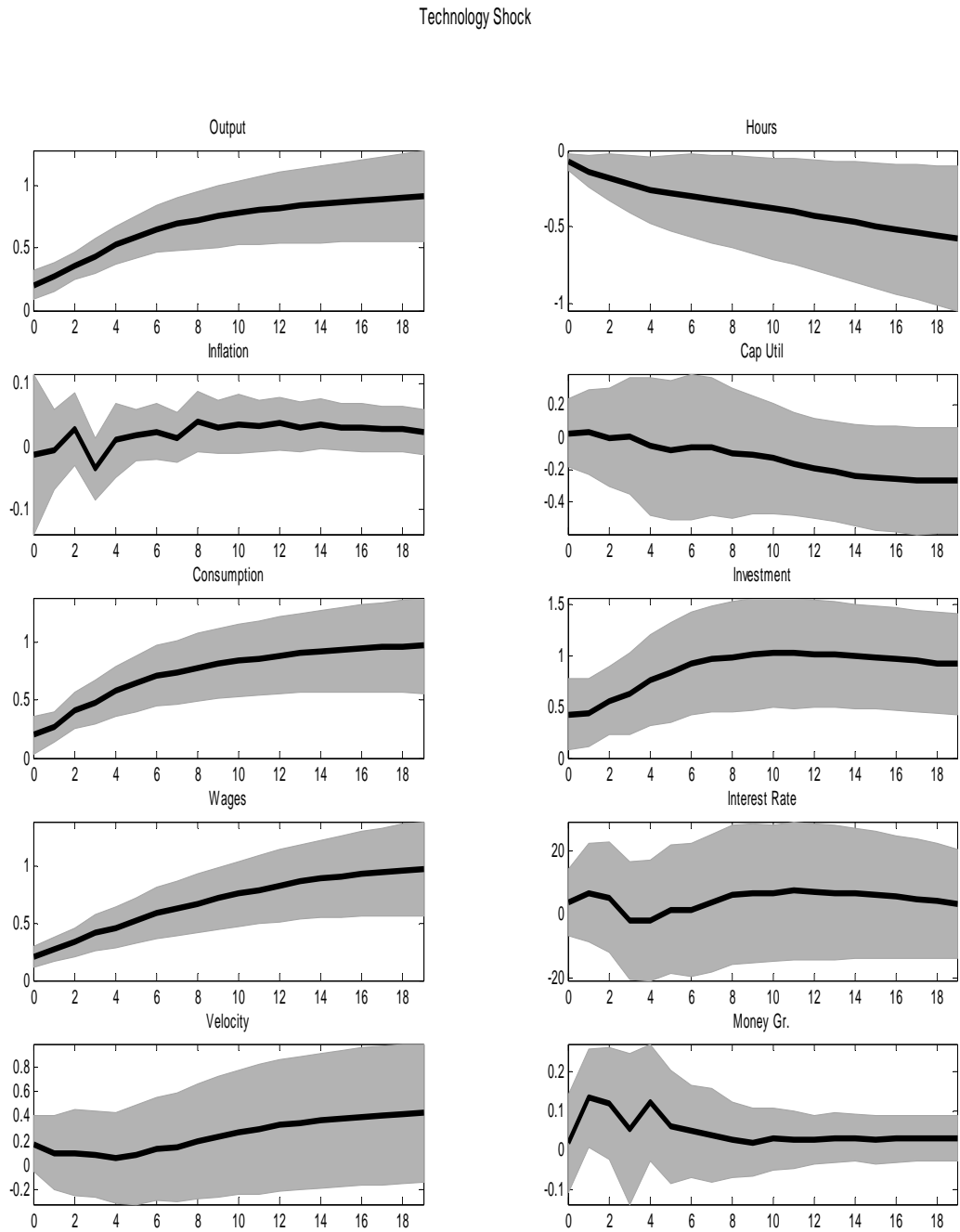Figure 2: Impulse responses of a technology shock in the DSVAR

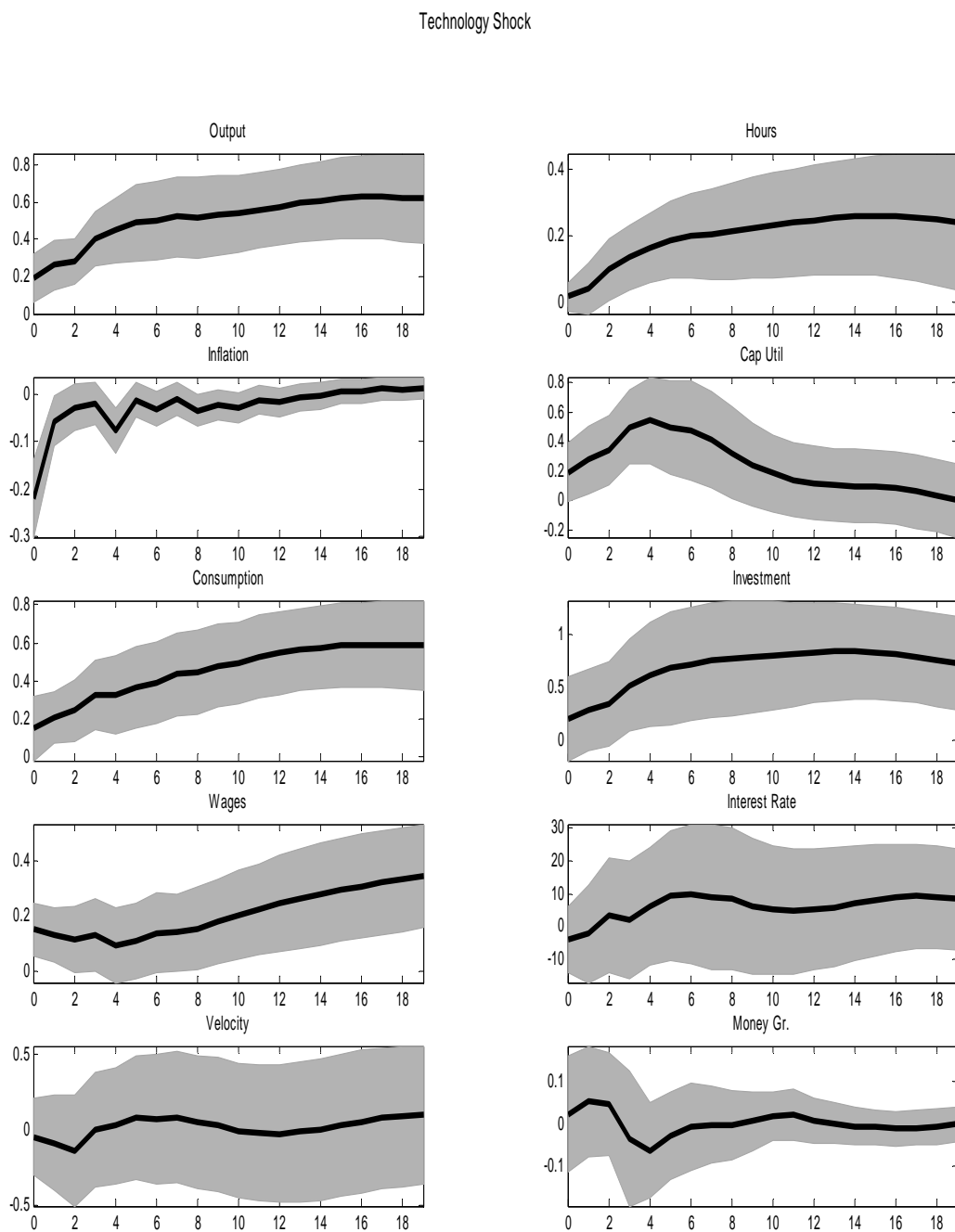Figure 3: Impulse responses of a technology shock in the LSVAR

Technology Shock

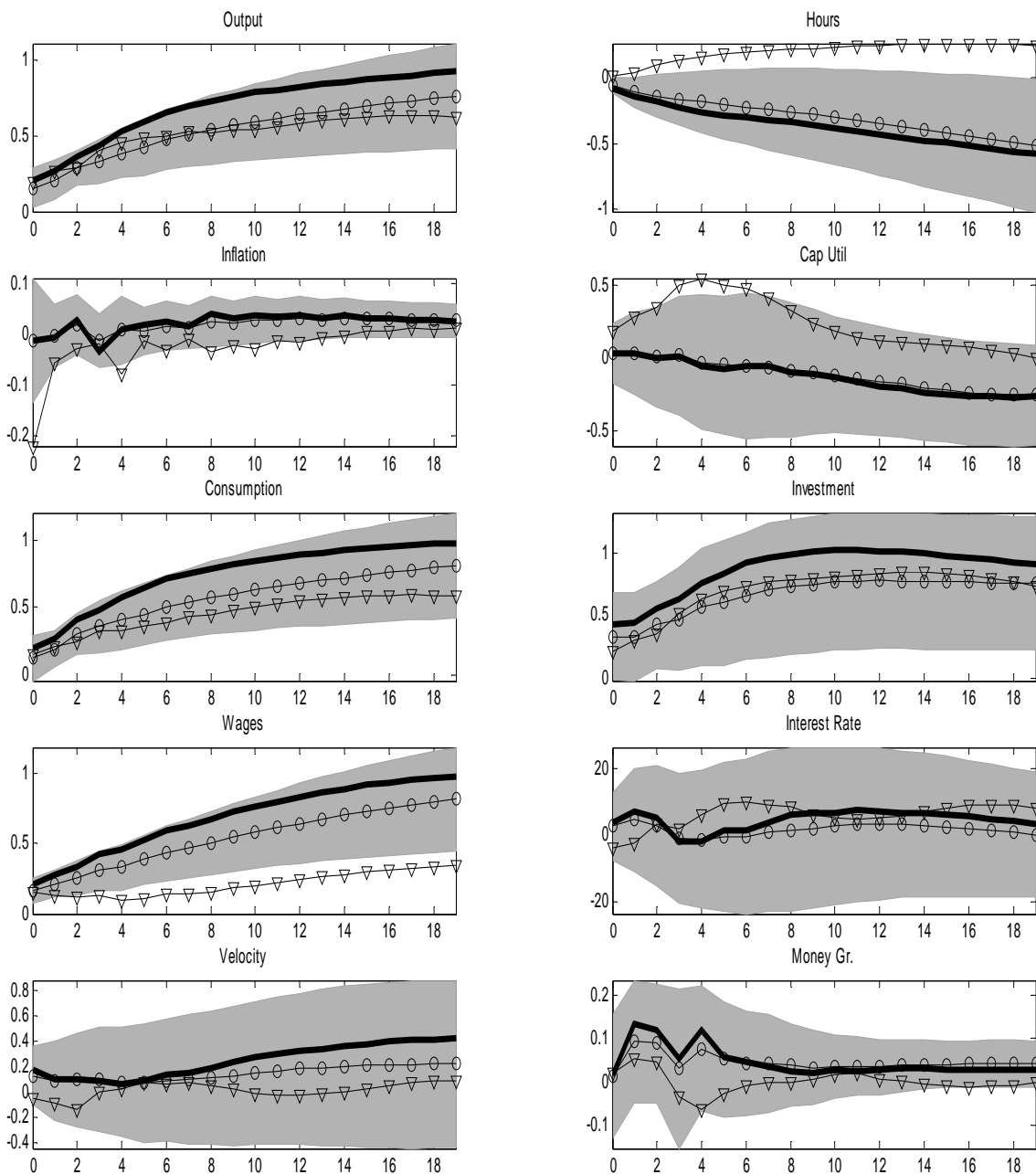Figure 4: Encompassing Test with LSVAR as DGP
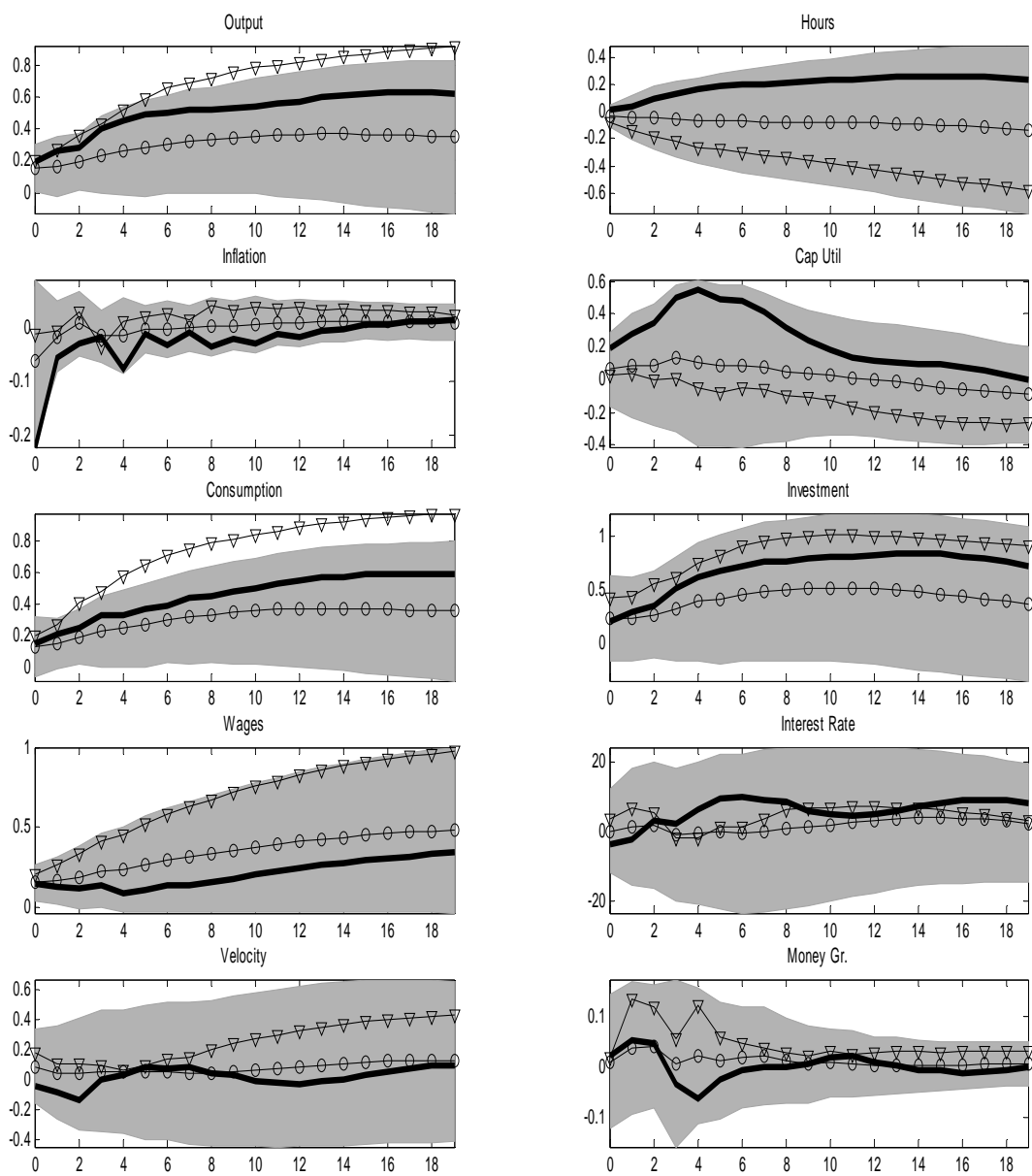
Technology Shock

Thick Line: Impulse responses of DSVAR with the actual data.
Line with circles: Impulse responses of DSVAR with data simulated from LSVAR.
Line with triangles: Impulse responses of LSVAR with the actual data.
Gray area: Confidence bands relative to the lines with circles.

23

Figure 5: Encompassing Test with DSVAR as DGP

Technology Shock

Thick Line: Impulse responses of LSVAR with the actual data.
Line with circles: Impulse responses of LSVAR with data simulated from DSVAR.
Line with triangles: Impulse responses of DSVAR with the actual data.
Gray area: Confidence bands relative to the lines with circles.