Are unit root tests useful in the debate over the (non)stationarity of hours worked?

Amélie Charles, Olivier Darné, Fabien Tripier

Are Unit Root Tests Useful in the Debate over the (Non)Stationarity of Hours Worked?

Amélie CHARLES[*], Olivier DARNÉ[†], Fabien TRIPIER[‡]

June 14, 2011

Abstract

The performance of unit root tests on simulated series is compared, using the business-cycle model of Chang et al. (2007) to generate data. Overall, Monte Carlo simulations show that the efficient unit root tests of Ng and Perron (2001) are more powerful than the standard unit root tests. These efficient tests are frequnetly able (i) to reject the unit-root hypothesis on simulated series, using the best specification of the business-cycle model found by Chang et al. (2007), in which hours worked are stationary with adjustment costs, and (ii) to reduce the gap between the theoretical impulse response functions and those estimated with a Structural VAR model. The results of Monte Carlo simulations show that the hump-shaped behaviour of data can explain the divergence between unit root tests.

---

[*]Audencia Nantes, School of Management, 8 route de la Jonelière, 44312 Nantes Cedex 3. Email: acharles@audencia.com.

[†]LEMNA, University of Nantes, IEMN–IAE, Chemin de la Censive du Tertre, BP 52231, 44322 Nantes, France. Email: olivier.darne@univ-nantes.fr.

[‡]Corresponding author: LEMNA, University of Nantes, IEMN–IAE, Chemin de la Censive du Tertre, BP 52231, 44322 Nantes, France. Email: fabien.tripier@univ-nantes.fr.

1 Introduction

Economists use econometrics to identify key statistical properties of the data, which are afterwards incorporated into theoretical models. For an econometric tool to be useful for this purpose, it must pass a "natural economic test"[1]: it should be possible to re-identify the statistical properties of the data that was identified by this econometric tool when the theoretical model is used as the Data Generating Process (DGP hereafter). Theoretical models are widely used as DGP by researchers on the business cycle, to assess the performance of econometric methods, as in Erceg et al. (2004) and Chari et al. (2008) for structural VAR, in Lindé (2005) to compare the full information maximum likelihood approach and the generalized method of moments, in An and Schorfeide (2007) for Bayesian methods, in Canova and Sala (2009) for methods based on impulse response functions, and in Gorodnichenko and Ng (2010) for methods of moments. To the best of our knowledge, models of the business cycle have not yet been used to assess the performance of unit root tests.

Applications of unit root tests to financial and macroeconomics series have challenged conventional economic theory and stimulated the development of new theories in numerous fields, such as economic fluctuations (Nelson and Plosser, 1982).[2] The debate over the stationarity of hours worked was sparked by Gali (1999)'s results on the effects of technological shocks[3], which contradict the technology-driven business cycle theory. Gali (1999)'s results are based on a Structural VAR (SVAR hereafter) model *à la* Blanchard and Quah (1989) that uses the first difference in hours worked. Gali (1999) motivates this specification

---

[1] This expression is borrowed from Chari et al. (2008) who apply this "natural economic test" to the methodology of structural VAR with long-run restriction.

[2] For example, the detection of a unit root in output by Nelson and Plosser (1982) legitimated the development of business cycle models with very persistent or non-stationary shocks to factors' productivity. The first generation of Real Business Cycle models considered a very persistent autoregressive process for the technological shock; see Kydland and Prescott (1982), Hansen (1985) and Prescott (1986). The effects of technological shocks have been modelled as a random walk, generally in multiple-shocks models as in King et al. (1991) and Christiano and Eichenbaum (1992). See Hansen (1997) for a discussion of this issue.

[3] Gali (1999) concludes that technological shocks play a minor role in the business cycle and that a positive technological shock induces a decrease in the number of hours worked.

by appealing to the outcome of standard Augmented Dickey-Fuller (ADF hereafter) tests.[4] Among the responses to Gali (1999)'s findings[5], Christiano et al. (2004) obtained results opposite to those of Gali (1999) by using the level of hours worked, and not the first difference of the series in the SVAR. As Gali (1999), Christiano et al. (2004) motivate the specification of the SVAR by the outcome of a stationarity test. Whelan (2009) also obtained results that contradict those obtained by Gali (1999) using different tests and data. In response to these mixed results, one strand of the literature suggests abandoning the use of standard unit root tests.[6] These results can be explained by a well-known shortcoming of unit root tests, which is that the properties of unit root tests are generally weak for the sample size of typical macroeconomic time series (about 100-200 observations) (Haldrup and Jansson, 2006), such as the hours worked series.

Previous studies on the stationarity of hours worked suffers from two further drawbacks. Firstly, they consider few and relatively "old" standard unit root tests (especially ADF) and do not include the recent developments of efficient unit root tests, especially those of Elliott et al. (1996) (ERS hereafter) and Ng and Perron (2001) (NP hereafter). These efficient tests allow the elimination of the deterministic components that are included in the test regression of the standard unit root tests (a constant mean in the hours worked series) to bring a gain in efficiency to the unit root tests by increasing their power (Schmidt and Phillips, 1992). Secondly, when several tests are used, their performances are not compared in the business-cycle model framework. However, if observed data are viewed as one realization of an economic model, it is essential that the unit root tests used perform well when this economic model is used to generate data.[7] To show the usefulness of the unit root tests in the debate over (non)stationarity of hours worked, we compare

---

[4]Gali and Rabanal (2004) extend the set of tests to the KPSS test and confirm the findings of Gali (1999).

[5]For example, Francis and Ramey (2005) develop Real Business Cycle models consistent with a negative response of employment to a positive technological shock, whereas Chari et al. (2008) argue that SVAR are useless for developing business cycle theory.

[6]To overcome the choice between the first difference in hours worked or the level of the series, Fève and Guay (2009) suggest using a more clearly stationary variable in the SVAR instead of hours, namely the ratio of consumption to output, and show how to recover the responses of hours to shocks in a second step, independently of the specification of the series (in level or in first difference).

[7]For example, one issue with standard unit root tests used in Chari et al. (2008) is that they are unable to reject the hypothesis where the hours series has a unit root whereas the hours series in the model is highly persistent, but stationary.

herein the performance of several tests (ADF, ERS and NP) using a business cycle model to generate data.[8]

Herein, we adopt the model proposed by Chang et al. (2007), which has several important attractive features. It (i) allows for either stationary or non-stationary hours worked, (ii) considers whether or not there are adjustment costs of labor, and (iii) has been estimated with Bayesian methods to account for certain facts about the business cycle that pertain to output and labor. We use the four model specifications estimated by Chang et al. (2007) to assess the sensitivity of test performances to the choice of the DGP. For each specification, we simulate the model for various sample sizes (100, 200, 500, 1000) and evaluate the size and power properties of the various unit root tests.

We show that the performance of the unit root tests is very sensitive to the specification of the model, i.e. the structure of shocks as well as the existence of adjustment costs. Even if the ADF and NP tests give similar (incorrect) properties for the data generating process with stationary hours and no adjustment costs of labor, the NP tests strongly dominate the ADF test when the adjustment of labor is costly. This result indicates the need to assess the performance of tests rigorously before applying them to observed data. It also raises the issue of how to specify the model, given the effect that the specification can have on the evaluation of tests. In the model of Chang et al. (2007), adjustment costs are a powerful propagation mechanism that induce hump-shaped responses of hours worked to shocks with a quicker return to the steady state level. Monte Carlo simulations show a similar difference in performance between ADF and NP unit root tests for ARMA processes with hump-shaped behavior. Since adjustment costs are widely supported by quantitative macroeconomic studies, and notably by Chang et al. (2007), among others, these results lead us to prefer the model specification with adjustment costs and therefore to recommend the NP tests rather than the ADF test. Finally, we investigate the implications of specifying the model in this way for the SVAR methodology.

The SVAR methodology has been discussed extensively in the literature (e.g., Faust and Leeper, 1997;

---

[8]Others unit root tests have been developed to overcome the limitations of the standard unit root tests, such as the presence of structural breaks (e.g., Perron, 1989; Zivot and Andrews, 1992; Perron and Rodriguez, 2003; Lee and Strazicich, 2003) or the presence of nonlinearity (Enders and Granger, 1998; Caner and Hansen, 2001; Kapetanios et al., 2003; Kapetanios and Shin, 2006). We do not use these tests because the DGP do not show breaks and/or nonlinearity.

Cooley and Dwyer, 1998) and criticized for its inability to identify the correct Impulse Response Functions (IRFs hereafter) when a business cycle model is as DGP (e.g., Ercerg et al., 2004; Ravenna, 2007; Chari et al., 2008; Dupaigne et al., 2007). Chari et al. (2008) demonstrate that the bias in the estimated IRFs is larger when the VAR is specified with the first difference in hours worked rather than the level of the series. Our contribution is to improve the specification of the VAR in the SVAR methodology. To demonstrate our improvement in the specification, we simulate output and hours series with small sample size (200 quarters) for a specification of the model that uses stationary hours and labor adjustment costs. This specification is held to be more consistent with the empirical facts than other specifications (Chang et al., 2007). We apply unit root tests to series of hours worked. Then, and depending on the outcomes of tests, we specify an empirical VAR in first difference or in level. Finally, we estimate IRFs using the long-run restrictions. The NP test indicates more frequently that hours worked are stationary than the ADF test; hence, the empirical VAR is more frequently specified in level and the estimated bias of IRFs smaller when the NP test is used, rather than the ADF test. The reduction of the bias is important even if, unfortunately, confidence intervals remain large.

The remainder of the paper is organised as follows. Section 2 describes the methodology that we use. Section 3 presents the results, and discusses (i) the effect of the results on the persistence mechanism and hump-shaped behaviour, and (ii) the implications of the results for SVAR methodology. Section 4 concludes.

2   Methodology

This section presents the models that are used to generate data, the unit root tests, the SVAR methodology, and the Monte Carlo study.

2.1   Business-Cycle Models as Data Generating Processes

We now describe the model briefly and present the various specifications that are suggested by Chang et al. (2007) and are used to generate data. The model is real and perfectly competitive. Households consume, accumulate physical capital, and supply production factors (labor and physical capital) to firms. Households

maximize the expected intertemporal utility function

$$\mathrm{E}_t \left\{ \sum_{s=0}^{\infty} \beta^{t+s} \left(\ln C_{t+s} - \frac{(H_{t+s}/B_{t+s})^{1+1/\nu}}{1+1/\nu} \right) \right\}$$

(1)

where $0 < \beta < 1$ is the subjective discount factor, ν the Frisch elasticity of labor supply, C_t the household consumption, H_t the household hours worked, B_t a preference shock on the disutility of labour, and t the period. The representative household faces the budget constraint

$$W_t H_t + R_t K_t = C_t + K_{t+1} - (1-\delta) K_t$$

(2)

where δ is the rate at which physical capital depreciates, W_t the wage rate, R_t the rate at which physical capital is rented, and K_t the stock of physical capital held by the household. Firms combine physical capital and labour to produce the final good according to

$$Y_t = \left(A_t H_t^d \right)^{\alpha} \left(K_t^d \right)^{1-\alpha} \left[1 - \varphi \left(\frac{H_t^d}{H_{t-1}^d} - 1 \right)^2 \right]$$

(3)

where $0 < \alpha < 1$ is the elasticity parameter of the production function, A_t is the technological shock common to all firms, H_t^d and K_t^d the demand of inputs, and $\varphi \geq 0$ measures the size of the adjustment costs of labour. The model description is closed with the shock processes

$$\ln A_t = \gamma + \ln A_{t-1} + \varepsilon_{a,t}, \ \varepsilon_{a,t} \sim iid\,(0,\sigma_a)$$

(4)

$$\ln B_t = \rho_b \ln B_{t-1} + (1-\rho_b) \ln B_0 + \varepsilon_{b,t}, \ \varepsilon_{b,t} \sim iid\,(0,\sigma_b)$$

(5)

where $\gamma > 0$ is the deterministic component of the drift of technological shocks and $0 < \rho_b \leq 1$ denotes the persistence of shocks to the household's utility function.

2.2 The Unit Root Tests

Dickey and Fuller (1981) developed the ADF unit root test for testing the hypothesis that a univariate time series contains a unit root against the alternative hypothesis that it is level stationary or trend stationary. For our case of interest, i.e. a constant mean in the hours worked series (Gali, 1999; Whelan, 2009), the test regression is defined by

$$\Delta y_t = \alpha + \beta_0 y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + \varepsilon_t$$

(6)

6

where $\{\varepsilon_t\}$ is a sequence of independent normal random variables with mean zero and variance σ^2, i.e. $\varepsilon_t \sim IN(0; \sigma^2)$. The ADF t-test is performed by testing the null hypothesis $\beta_0 = 0$ against the alternative $\beta_0 < 0$.

Some studies show that the elimination of deterministic components (here the constant mean) may result in the unit root tests being more efficient by increasing their power. ERS develop a unit root test based on a quasi-difference detrending of the series. They suggest using the Dickey-Fuller generalized least squares (DF-GLS) test using the following regression

$$\Delta \tilde{y}_t = \beta_0 \tilde{y}_{t-1} + \sum_{j=1}^{k} \beta_j \Delta \tilde{y}_{t-j} + \varepsilon_t \tag{7}$$

where $\tilde{y}_t$ is the locally detrended series y_t. The DF-GLS t-test is performed by testing the null hypothesis $\beta_0 = 0$ against the alternative $\beta_0 < 0$. The local detrending series is defined by

$$\tilde{y}_t = y_t - \hat{\psi}' z_t$$

where z_t is equal to 1 for the constant mean case, and $\hat{\psi}'$ is the GLS estimator obtained by regressing $\tilde{y}$ on $\hat{z}$ where $\tilde{y} = (y_1, (1 - \bar{\alpha}B) y_2, ..., (1 - \bar{\alpha}B) y_T)$, $\bar{z} = (z_1, (1 - \bar{\alpha}B) z_2, ..., (1 - \bar{\alpha}B) z_T)'$, and $\bar{\alpha} = 1 + c/T$. ERS recommend using $\bar{c} = -7$ for the constant mean case. They also consider a point optimal test of the unit root null hypothesis $\alpha = 1$ against the alternative $\alpha = \bar{\alpha}$ (see Appendix A).

NP also propose efficient unit root tests based on the regression (7). Their tests, called M-GLS tests (see Appendix A), are modifications of the Phillips and Perron (1988) test, which is a non-parametric approach to correct residual autocorrelation by modifying the Dickey-Fuller test statistics: first, to correct the size distortions (as suggested by Perron and Ng, 1996), and second, to improve the power (as suggested by Elliott et al., 1996).

2.3 The SVAR methdolology

Chari et al. (2007), among others, provide a general description of the SVAR methodology with long-run restrictions together with a collection of Matlab programs. Stationary data X_t are described by the following

empirical VAR with p lag:

$$X_t = B_1 X_{t-1} + B_2 X_{t-2} + \ldots B_p X_{t-p} + v_t \tag{8}$$

v_t are the canonical innovations, with $E v_t v_t' = \Omega$, and B_i matrices of autoregressive coefficients for $i = 1, ..., p$. Equation (8) is inverted to get the Wold decomposition

$$X_t = v_t + C_1 v_{t-1} + C_2 v_{t-2} + \ldots \tag{9}$$

where the C's satisfy $I = (I - B_1 L - B_2 L^2 \ldots - B_p L^p)(I + C_1 L + C_2 L^2 + \ldots)$ for all values of L. The model with structural innovations is defined as follows:

$$X_t = A_0 \epsilon_t + A_1 \epsilon_{t-1} + A_2 \epsilon_{t-2} + \ldots \tag{10}$$

with $A_0 \epsilon_t = v_t$ and $A_j = C_j A_0$, $j \geq 1$. The identifying restrictions for SVAR are $E \epsilon_t \epsilon_t' = I$ and the (1,1) element of $\sum_{j=0}^{\infty} A_j$, or equivalently $[\sum_{j=0}^{\infty} C_j] A_0$, is equal to 0. This gives a system of four equations and four unknowns.

2.4 The Monte Carlo design

The model is calibrated using the outcome of the estimations of Chang et al. (Table 2, p. 1366, 2007) for the four specifications given in Table 1. The model is simulated using the programs provided by the authors.[9] All experiments are based on 30,000 replications. We consider separately each specification of the model that is used to generate data.

1. The specifications of the DSGE described in Table 1 are used to generate simulated macroeconomic data of length T. The sample sizes considered are $T = 100, 200, 500$ and 1000.

2. Unit root tests are applied to simulated data for hours worked to compute their properties. We base the choice of lag length on the sequential procedure proposed by Ng and Perron (1995) for the ADF test and we use the modified Akaike information criteria suggested by Ng and Perron (2001) for efficient unit root tests.[10] The observed unit-root test statistics are compared to their finite-sample 5% critical

---

[9]The required programs are dsge.g, dsgemod.src, and dsgesim.src.

[10]Ng and Perron (2001) show that the popular Akaike and Schwarz information criteria are not sufficiently flexible for unit root tests to select the appropriate number of lags in the regression (mainly when there are negative moving-average errors).

values given in (i) the original papers of the unit root tests, (ii) MacKinnon (1991) and Vougas (2007) for the small finite-sample, and (iii) our computations.

3. Simulated data from DSGE models for output and hours are used to estimate SVAR with long-run restrictions. If the test indicates that hours are stationary, the hours series is introduced in level in the SVAR; otherwise, the first difference in hours is introduced. For each test, we compute the population moments of IRFs.

3   Results

We now present the results for the performance of unit root tests and the SVAR predictions from the Monte-Carlo experiments, and discuss the role of persistence mechanisms in Section 3.2.

3.1   The performances of unit root tests

Table 3 displays the results for the DGP where hours worked are stationary. Table 4 reports the results for the DGPs where the hours worked are non-stationary, without (Panel A) and with (Panel B) adjustment costs. The power of unit root tests is given in Table 3 and the size of unit root tests is presented in Table 4. For the DGPs with non-stationary hours worked (Table 4), the unit root tests show good size, whatever the sample sizes, and without and with adjustment costs.

For the DGPs with stationary hours worked (Table 2), the major issue concerns stationary hours worked when the samples are small ($T = 100$ and 200). Such sample sizes are typical for macroeconomic series. In this case, significant differences appear between tests and interestingly also between model specifications. Overall, the NP tests (MZ_α, MZ_t and MPT) exhibit higher power than the other tests that we studied, particularly the ADF test, but with some difference according to the model specification. For the model without adjustment costs (Panel A) with $T = 200$, the NP tests reject the unit root hypothesis at a rate of 45% (especially for MZ_α and MZ_t) against 37% for the ADF test ($T = 200$) and 40% for ERS tests (DF-GLS and PT). This slight difference does not warrant a preference for the NP tests over the

standard unit root test. The conclusion is different for the model with adjustment costs (Panel B). In this case, the NP tests reject the unit root hypothesis at a rate of 72% and 50%, against 35% and 12% for the ADF test for $T = 200$ and $T = 100$, respectively. Note that the ERS tests are slightly less powerful than the NP tests.[11]

In light of the foregoing, it would seem that the efficient unit root tests, especially the NP tests, are more powerful than the standard unit root test. This indicates that the NP tests should be preferred to the ADF test in this framework, given the fact that the model with adjustment costs is more consistent with empirical facts than the model without adjustment costs, as shown by Chang et al. (2007).

3.2   The persistence mechanisms

How well a test performs, given the specification of the model, is a function of the amplification and propagation mechanisms of the model in question. Adjustment costs are well-known to propagate the effects of shocks in the economy. Agents smooth the adjustment of labour to reduce total costs. Given that adjustment costs increase the persistence of shocks in the economy, it is surprising that the NP tests reject more frequently the unit root hypothesis for the DGP with adjustment costs. This result can be explained by the fact that with adjustment costs, shocks on the household's utility function are less persistent ($\rho_b = 0.80$) than without adjustment costs ($\rho_b = 0.95$). If we consider simultaneously very persistent shocks on the household's utility function (i.e. $\rho_b = 0.95$) and the persistence induced by labour adjustment costs, the NP tests would fail to reject the unit root hypothesis.[12] However, Chang et al. (2007) show that labour adjustment costs result in a reduction in the persistence of shocks that are due to variations in the supply of labour (measured by ρ_b) in the model.

To clarify this point, we make a distinction between the *endogenous persistence*, associated with adjustment costs, and the *exogenous persistence*, associated with the persistence of the exogenous shocks

---

[11]Note that we also consider the stationarity test of Kwiatkowski et al. (1992) in the Monte Carlo experiments. This test shows strong size distortions in small samples, especially for $T = 100$. The results of the KPSS test are available upon request.

[12]If we impose $\rho_b = 0.95$ in the specification 3 of the model, the rates of rejection are 65% for the NP test and 86% for the ADF (T=200). The complete table is available upon request.

on the supply of labour. In their procedure for estimating the business cycle model, Chang et al. (2007) proposed that there is an inverse relation between the two forms of persistence. A high value for φ, which measures the size of adjustment costs, is associated with a low value of ρ_b, which measures the autocorrelation of the shocks due to variation in the supply of labour (see Table 1 and Chang et al., 2007, Fig. 2 p. 1367). Figure 1 shows the sharp contrast in the IRFs of hours worked between the two specifications (with and without adjustment costs). The model without adjustment costs generates monotonic responses of labour to a stationary supply shock, but which last for a very long time, whereas the model with adjustment costs generates hump-shaped responses of labour with a quicker return to the steady-state level. Hump-shaped behavior of series is a major issue in the literature on the business cycle.(see, e.g., Cogley and Nason, 1995). Further, Chang et al. (2007) conclude that the model with adjustment costs and stationary hours has the best fit among the four specifications. Given these findings, our results suggest that we should use the efficient unit root tests proposed by Ng and Perron (2001) because they are more powerful than the ADF test when simulated series are hump-shaped.

To confirm this intuition concerning the effect of the hump-shaped behaviour on the ADF test, we performed another Monte Carlo study. We simulated an ARMA(1,2) process with hump-shaped behavior, i.e. $y_t = 0.80y_{t-1} + \varepsilon_t + 0.65\varepsilon_{t-1} + 0.60\varepsilon_{t-2}$, with ε_t is i.i.d.. We also simulated an AR(1) model as a benchmark, defined as $y_t = \rho y_{t-1} + \varepsilon_t$, with $\rho = 0.85$. The AR model presents the same mean-reversion behaviour as that of the ARMA model. Table 4 gives the power of unit root tests and Figure 1 plots the IRFs. The sample sizes, the number of replications, and the choices of lag length for the unit root tests are based on the same procedures as were used in the previous Monte Carlo experiment.

For the AR(1) process (Panel A), all the unit root tests present good power, even for small sample sizes. Note that the ADF shows lower power than the efficient unit root tests when $T = 100$. For the ARMA(1,2) process with a hump-shaped behaviour (Panel B), all the unit root tests have high power for large sample sizes ($T = 1000$ and 500). When $T = 200$, the efficient unit root tests have good power, whereas the ADF test shows a loss of power (with a rate of rejection of 90% for the NP and ERS tests, against 50% for the ADF test). More interesting, the ADF test presents a low power for $T = 100$ with a rate of rejection of 17% against 60% for the efficient tests. These results show that the ADF test is affected by

the hump-shaped behaviour, whereas the NP tests show good power. Note that the ADF test is also more biased than the NP tests by the presence of a MA component in the ARMA model without hump-shaped behaviour (Panel C).

3.3 Ilustration with observed data

Our Monte Carlo experiments indicate that they are strong differences between the various unit root tests on the (non)stationarity of hours worked. It is crucial to see whether the results still differ when observed data are used instead of simulated data. To this end, we applied the ERS and NP efficient tests to the three data sets used in Chang et al. (2007). We obtained different results for the (non)stationarity of the hours worked than those obtained by the authors using ADF tests (see Table 5). For two of the three series, the unit root hypothesis is rejected by the efficient unit root tests, whereas this hypothesis is never rejected for the three series according to the ADF test.

3.4 Implications on SVAR

We now derive and present the implications of our results for the SVAR methodology with long-run restrictions. We restrict our attention to the model specification that has the greatest agreement with the empirical facts (Chang et al., 2007), i.e. with stationary hours and labour adjustment costs (specification 3), and with a small sample size ($N = 200$).

The SVAR has been criticized by Chari et al. (2008) for being unable to provide useful predictions for business cycle theory. To show the restrictive feature of this method, they simulate a DSGE model and apply the SVAR methodology to simulated series and show that for realistic sample sizes, the estimated IRFs are far from the actual IRFs. This leads the authors to conclude that the SVAR methodology has no practical application in business cycle theory.

We performed an exercise that is similar to that of Chari et al. (2008), with one major exception: we used the outcome unit root tests to choose the SVAR specification (in level or in first difference). If the unit root test indicated that hours are stationary, the SVAR was specified with the hours series in level.

12

Otherwise, if the test indicated that hours are not stationary, the SVAR was specified with the hours series in first difference. We compared the population moments of IRFs according to the unit root test used: NP or ADF. The results are reported in Figures 3, 4, and 5.

The medians of population IRFs show significant differences according to the unit root test that is used. Several features of the theoretical IRFs are replicated when the NP test is used instead of the ADF test. Firstly, the median IRF of output in response to the no-technological shock is hump-shaped for the NP test, as in the DSGE model, whereas it is monotonically decreasing for the ADF test. Secondly, the median IRF of hours worked in response to the no-technological shock is positive and hump-shaped for the NP test, as in the DSGE model, whereas it is negative for the ADF test. Thirdly, the median IRF of hours worked in response to the technological shock returns to zero in the long-run for the NP test, as in the DSGE model, whereas it is remains highly positive for the ADF test.

These findings show the advantages for the SVAR methodology of using efficient unit root tests, such as the NP test, rather than the standard ADF test. Nevertheless, there are also dimensions for which the results of the NP test are less satisfactory. For the NP test, the median IRF of hours worked in response to the technological shock is strongly overestimated for the first quarters after the shocks and the long-run median IRF of output in response to the technological shock is underestimated. This last point is the only one for which the median IRF is closest to the DSGE model IRF when the ADF test is used, rather than the NP test. Furthermore, it is worth mentioning that interval confidence remains very large for the two tests. On this point, unfortunately, our methodology does not rectify this well-known failure of SVAR; see Chari et al. (2008) and Christiano et al. (2004) for discussions of this issue.[13]

4   Concluding remark

The mixed results of the unit root tests on the (non)stationarity of hours worked casts doubt on how far they can be useful for developing business cycle theory. In the work reported herein, we attempted to improve

---

[13] A full treatment of this issue lies beyond the scope of this paper. Christiano et al. (2006), Kascha and Mertens (2009), Liu and Theodoridis (2010), Gust and Vigfusson (2009) propose alternative solutions to it.

the contribution of unit root tests to economic theory by linking the process by which the quality of the tests is assessed to economic theory. From Monte Carlo simulations using data generated by a well-specified business cycle model, namely the Chang et al. (2007) model with labour adjustment costs, we showed that the efficient unit root tests proposed by Ng and Perron (2001) are more powerful than the standard ADF unit root test. This result can be explained by the fact that the labour adjustment costs generate hump-shaped behaviour and reduce the persistence of shocks to the household's utility function. The effect of hump-shaped behaviour on the ADF test is confirmed from Monte Carlo experiments on ARMA models. This finding suggests that the Ng and Perron tests should be preferred in this framework. Furthermore, we found that using the NP tests, rather than the ADF test, to choose the SVAR specification (in level or in first-difference) for the hours worked narrows the gap between the theoretical impulse response functions and those estimated with a SVAR model. In light of these findings, macroeconomists may find it beneficial in their research on unit root tests to generate data using business cycle models.

Appendix A. Unit root tests

Point optimal test of Elliott et al. (1996)

Elliott et al. (1996) consider a point optimal test of the unit root null hypothesis $\alpha = 1$ against the alternative $\alpha = \bar{\alpha}$. given by

$$PT = \left[S(\bar{\alpha}) - \bar{\alpha} S(1) \right] / s_{ar}^2$$

where $S(a)$ is given by $(y_a - z_a \psi)'(y_a - z_a \psi)$, and s_{ar} is the autoregressive spectral density estimator of the long-term variance. The value of $\bar{c}$ is chosen such that the asymptotic power of test is 50% against the local alternative $(\bar{\alpha} = 1 + \bar{c}/T)$. ERS advise $\bar{c} = -7$ for the constant mean case.

M-GLS tests of Ng and Perron (2001)

The M-GLS tests proposed by Ng and Perron (2001) are defined as

$$
\begin{aligned}
MZ_t &= \left(T^{-1} \tilde{y}_T^2 - s_{ar}^2 \right) \left(4 s_{ar}^2 T^{-2} \sum_{t=1}^{T} \tilde{y}_{t-1}^2 \right)^{-1/2} \\
MZ_a &= \left(T^{-1} \tilde{y}_T^2 - s_{ar}^2 \right) \left(2 T^{-2} \sum_{t=1}^{T} \tilde{y}_{t-1}^2 \right)^{-1}
\end{aligned}
$$

where s_{ar} is the autoregressive spectral density estimator of the long-term variance. NP also consider a modified feasible point optimal test

$$MPT = \left[\bar{c}^2 T^{-2} \sum_{t=1}^{T} \tilde{y}_{t-1}^2 - \bar{c} T^{-1} \tilde{y}_T \right] / s_{ar}^2$$

References

An, S., Schorfheide, F. (2007). Bayesian Analysis of DSGE Models, Econometric Reviews, 26(2-4), 113-172.

Blanchard, O.J., Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. American Economic Review, 79, 655-73.

Canova, F., Sala, L. (2009). Back to square one: Identification issues in DSGE models, Journal of Monetary Economics, 56(4), 431-449.

Chang, Y., Doh, T., Schorfheide, F. (2007). Non-stationary Hours in a DSGE Model. Journal of Money, Credit and Banking, 39, 1357-1373.

Chari, V.V., Kehoe, P.J., McGrattan, E.R. (2007).

Chari, V.V., Kehoe, P.J., McGrattan, E.R. (2008). Are structural VARs with long-run restrictions useful in developing business cycle theory. Journal of Monetary Economics, 55, 1337-52.

Christiano, L.J., Eichenbaum, M. (1992). Current real business cycle theories and aggregate labor market fluctuations. American Economic Review, 82, 430-450.

Christiano, L.J., Eichenbaum, M., Vigfusson, R. (2004). What happens after a technology shock. Working Paper No 9819, NBER.

Christiano, L.J., Eichenbaum, M., Vigfusson, R. (2006). Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions, Journal of the European Economic Association, 4(2-3), 475-483.

Cogley, T., Nason, J. (1995). Output dynamics in Real-Business-Cycle models. American Economic Review, 84, 492-511.

Cooley, T. F., Dwyer, M. (1998). Business cycle analysis without much theory A look at structural VARs, Journal of Econometrics, Elsevier, 83(1-2), 57-88.

Dickey, D.A., Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with unit root. Journal of the American Statistical Association, 74, 427-481.

Dickey, D.A., Fuller, W.A. (1981). "Likelihood ratio statistics for autoregressive time series with unit root. Econometrica, 49, 1057-1072.

Dupaigne, M., Fève, P., Matheron, J. (2007). Some analytics on bias in DSVARs, Economics Letters, 97(1), 32-38.

Elliott, G., Rothenberg, T.J., Stock, J.H. (1996). Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.

Erceg, C.J., Guerrieri, L., Gust, C. (2005). Can Long-Run Restrictions Identify Technology Shocks?, Journal of the European Economic Association 3(6), 1237-1278.

Faust, J., Leeper, E. S. (1997). When Do Long-Run Identifying Restrictions Give Reliable Results?, Journal of Business & Economic Statistics, 15(3), 345-53.

Fève, P., Guay, A. (2009). The response of hours to a technology shock: A two-step structural VAR approach. Journal of Money, Credit and Banking, 41, 987-1013.

Francis, N., Ramey., V.A. (2005). Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited. Journal of Monetary Economics, 52, 1379-99.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? American Economic Review, 89, 249-71.

Galí, J., Rabanal, P. (2004). Technology shocks and aggregate fluctuations: How well does the real business cycle model fit postwar U.S. data? Working Paper No 10636, NBER.

Gust, C., Vigfusson, R. (2009). The power of long-run structural VARs, International Finance Discussion Papers 978, Board of Governors of the Federal Reserve System.

Haldrup, N., Jansson, M. (2006). Improving size and power in unit root testing. In Mills, T.C., Patterson, K. (eds.), Palgrave Handbook of Econometrics: Econometric Theory, Palgrave Macmillan.

Hansen, G.D. (1985). Indivisible labor and the business cycle. Journal of Monetary Economics, 16, 309-327.

Hansen, G.D. (1997). Technological progress and aggregate fluctuations. Journal of Political Economy, 16, 1005-1023.

Kascha, C., Mertens, K. (2009). Business cycle analysis and VARMA models, Journal of Economic Dynamics and Control 33(2), 267-282.

King, R., Plosser, C., Stock, J., Watson, M., (1991). Stochastic trends and economic fluctuations. American Economic Review, 81, 819-940.

Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationary against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics, 54, 159-178.

Kydland, F., Prescott, E. (1982). Time to build and aggregate fluctuations. Econometrica, 50, 1345-70.

Linde, J. (2005). Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach, Journal of Monetary Economics 52(6), 1135-1149.

Liu, P., Theodoridis, K. (2010). DSGE model restrictions for structural VAR identification, Bank of England working papers 402.

MacKinnon, J. (1991). Critical values for cointegration tests. In: Engle R. and Granger C. (eds), Long-run Economic Relationships, Readings in Cointegration, Oxford University Press, pp 267–276.

Nelson, C.R., Plosser, C.I. (1982). Trends and random walks in macroeconomic time series. Journal of Monetary Economics 10, 139-162.

Ng, S., Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. Journal of the American Statistical Association 90, 268-281.

Ng, S., Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-1554.

Perron, P., Ng, S. (1996). Useful modifications to unit root tests with dependent errors and their local asymptotic properties. Review of Economic Studies 63, 435-465.

Phillips, P.C.B., Perron, P. (1988). Testing for unit root in time series regression. Biometrika 75, 347-353.

Prescott, E.C. (1986). Theory ahead of business cycle measurement. Carnegie Rochester Conference Series on Public Policy, 25, 11-44.

Ravenna, F. (2007). Vector autoregressions and reduced form representations of DSGE models, Journal of Monetary Economics, Elsevier 54(7), 2048-2064.

Vougas, D.V. (2007). GLS detrending and unit root testing. Economics Letters, 97, 222–229.

Whelan, K.T. (2009). Technology shocks and hours worked: Checking for robust conclusions. Journal of Macroeconomics, 31, 231-239.

Tables and Figure

Table 1: Specifications of the DGPs.

N°	Specification	Parameter values
1.	Stationary hours worked without adjustment cost	$\alpha = 0.652; \beta = 0.995; \gamma = 0.004; \delta = 0.023; \nu = 0.527; \rho_B = 0.951$ $\alpha_A = 0.011; \sigma_B = 0.006; \ln A_0 = 5.708; \ln B_0 = 3.176; \varphi = 0$
2.	No-Stationary hours worked without adjustment cost	$\alpha = 0.654; \beta = 0.995; \gamma = 0.004; \delta = 0.024; \nu = 0.474; \rho_B = 1.000$ $\alpha_A = 0.011; \sigma_B = 0.006; \ln A_0 = 5.717; \ln B_0 = 3.166; \varphi = 0$
3.	Stationary hours worked with adjustment cost	$\alpha = 0.658; \beta = 0.995; \gamma = 0.004; \delta = 0.023; \nu = 0.433; \rho_B = 0.800$ $\alpha_A = 0.011; \sigma_B = 0.034; \ln A_0 = 5.748; \ln B_0 = 3.171; \varphi = 11.36$
4.	No-Stationary hours worked with adjustment cost	$\alpha = 0.661; \beta = 0.995; \gamma = 0.004; \delta = 0.024; \nu = 1.153; \rho_B = 1.000$ $\alpha_A = 0.011; \sigma_B = 0.012; \ln A_0 = 5.754; \ln B_0 = 3.194; \varphi = 8.054$

Source: Table 2 of Chang et al. (2007)

Table 2: Reject rates of unit-root test statistics – DGP: stationary hours worked.

Sample	MZ_α	MZ_t	DF-GLS	PT	MPT	ADF
Panel A: Without adjustment costs (1)						
$T = 1000$	0.9326	0.9281	0.9341	0.9242	0.9316	1.0000
$T = 500$	0.7984	0.8014	0.8320	0.7762	0.7953	0.9601
$T = 200$	0.4503	0.4580	0.4089	0.3952	0.4299	0.3688
$T = 100$	0.1859	0.1605	0.1460	0.1432	0.1643	0.1326
Panel B: With adjustment costs (3)						
$T = 1000$	0.9196	0.9154	0.9314	0.9125	0.9202	1.000
$T = 500$	0.8579	0.8620	0.8659	0.8422	0.8577	0.8642
$T = 200$	0.7286	0.7326	0.6821	0.6776	0.7148	0.3584
$T = 100$	0.5168	0.4874	0.3977	0.4369	0.4894	0.1186

Notes: (1) and (3) denote the specifications 1 and 3 in Table 1. MZ_α, MZ_t and MPT denote the Ng and Perron (2001) tests; DF-GLS and PT denote the Elliot et al (1996) tests; and ADF denotes the Augmented Dickey and Fuller (1981) test.

Table 3: Reject rates of unit-root test statistics – DGP: non-stationary hours worked.

Sample	MZ_α	MZ_t	DF-GLS	PT	MPT	ADF
Panel A: Without adjustment costs (2)						
$T = 1000$	0.0656	0.0615	0.0683	0.0639	0.0649	0.0620
$T = 500$	0.0642	0.0649	0.0658	0.0598	0.0624	0.0568
$T = 200$	0.0599	0.0603	0.0491	0.0486	0.0535	0.0590
$T = 100$	0.0550	0.0455	0.0403	0.0401	0.0466	0.0549
Panel B: With adjustment costs (4)						
$T = 1000$	0.0541	0.0501	0.0536	0.0530	0.0534	0.0390
$T = 500$	0.0547	0.0559	0.0535	0.0515	0.0538	0.0391
$T = 200$	0.0615	0.0616	0.0424	0.0500	0.0550	0.0389
$T = 100$	0.0722	0.0614	0.0390	0.0528	0.0609	0.0446

Notes: (2) and (4) denote the specifications 2 and 4 in Table 1. MZ_α, MZ_t and MPT denote the Ng and Perron (2001) tests; DF-GLS and PT denote the Elliot et al (1996) tests; and ADF denotes the Augmented Dickey and Fuller (1981) test.

Table 4: Reject rates of unit-root test statistics – DGP: AR(1) and ARMA(1,2) models.

Sample	MZ_α	MZ_t	DF-GLS	PT	MPT	ADF
Panel A: AR(1) with $\rho = 0.85$						
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$T = 500$	0.9998	0.9997	0.9999	0.9996	0.9996	1.0000
$T = 200$	0.9490	0.9507	0.9496	0.9430	0.9440	0.9958
$T = 100$	0.8033	0.7767	0.7528	0.7682	0.7756	0.6154
Panel B: ARMA(1,2) with $\rho = 0.80$,						
$\theta_1 = 0.65$ and $\theta_2 = 0.60$						
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$T = 500$	0.9999	0.9998	1.0000	0.9998	0.9998	0.9943
$T = 200$	0.9433	0.9442	0.9424	0.9335	0.9350	0.5069
$T = 100$	0.6627	0.6291	0.5512	0.6266	0.6330	0.1698
Panel C: ARMA(1,2) with $\rho = 0.80$,						
$\theta_1 = 0.16$ and $\theta_2 = 0.15$						
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$T = 500$	1.0000	1.0000	1.0000	0.9998	0.9998	0.9956
$T = 200$	0.9628	0.9633	0.9662	0.9585	0.9592	0.5885
$T = 100$	0.8460	0.8146	0.7936	0.8070	0.8119	0.2393

Notes: The AR(1) model is defined as $y_t = \rho y_{t-1} + \varepsilon_t$, and the ARMA(1,2) model as $y_t = \rho y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$, with ε_t is i.i.d.

MZ_α, MZ_t and MPT denote the Ng and Perron (2001) tests; DF-GLS and PT denote the Elliot et al (1996) tests; and ADF denotes the

Augmented Dickey and Fuller (1981) test.

Table 5: Results of unit root tests on hours worked series.

Dataset	MZ_α	MZ_t	DF-GLS	PT	MPT	k^a	ADF^b	k^b
Dataset 1	-12.40*	-2.47*	-2.48*	2.09*	2.05*	1	-2.80	4
Dataset 2	-3.65	-1.34	-1.42	7.93	6.71	1	-2.55	4
Dataset 3	-11.20*	-2.34*	-2.43*	2.29*	2.30*	1	-2.44	4
Critical value	-8.10	-1.98	-1.98	3.17	3.17		-2.86	
at the 5% level								

Notes: * indicates rejection of the unit-root null hypothesis at the 5% level of significance. [a] the lag order k in the regression is selected by using the Modified Information Criteria (MIC) proposed by Ng and Perron (2001). [b] the values of the ADF tests and lag order k are taken in Chang et al. (footnote 7, p. 1363, 2007). The three datasets have been collected by Chang et al. (2007). Dataset 1 is constructed by the Bureau of Labor Statistics and corresponds to the average weekly hours of all people in the non-farm business sector. Dataset 2 has been constructed by Christiano et al. (2004) (LBMN, DRI-Global Insight). Dataset 3 has been constructed by Gali and Rabanal (2004) and corresponds to non-farm business sector hours (LXNFH, Haver Analytics' USECON). MZ_α, MZ_t and MPT denote the Ng and Perron (2001) tests; DF-GLS and PT denote the Elliot et al (1996) tests; and ADF denotes the Augmented Dickey and Fuller (1981) test.
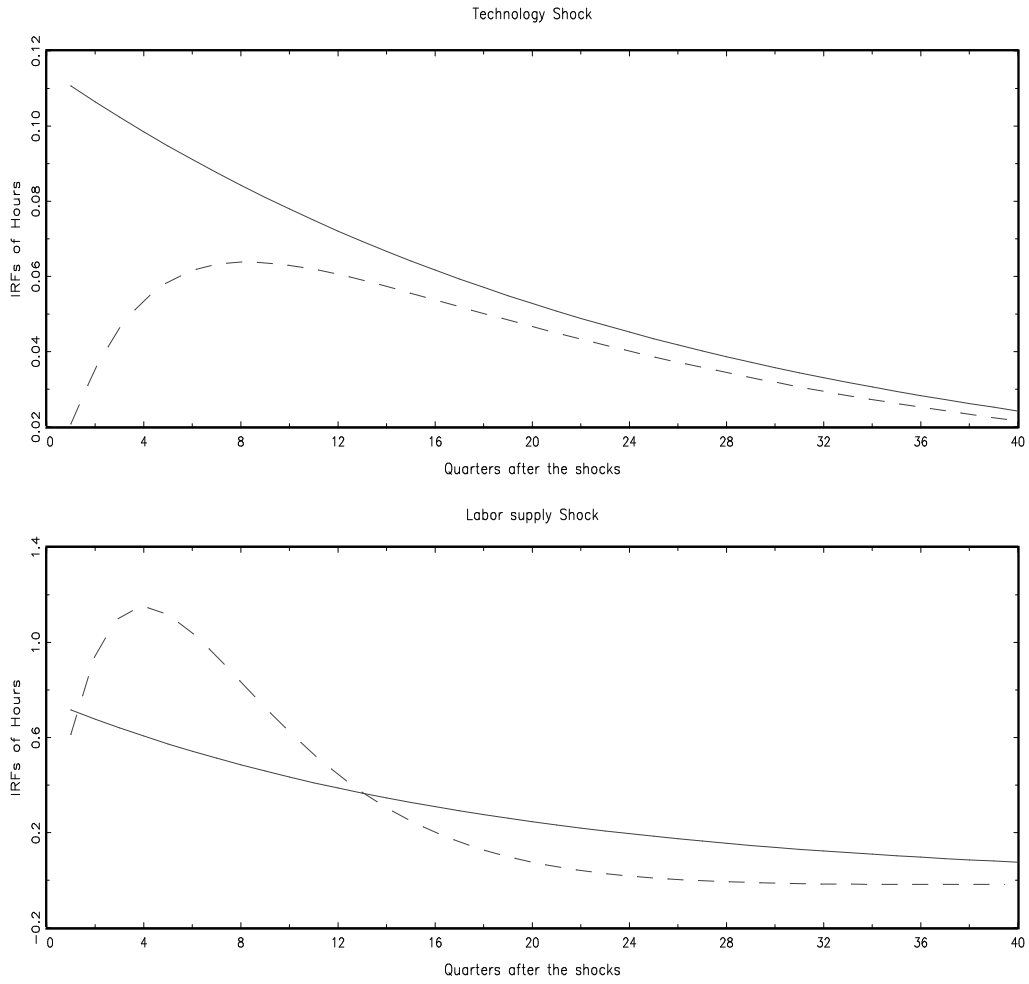
Figure 1: IRFs of Hours Worked to shocks with (dotted lines) and without labor adjustment costs (solid lines) for the model.
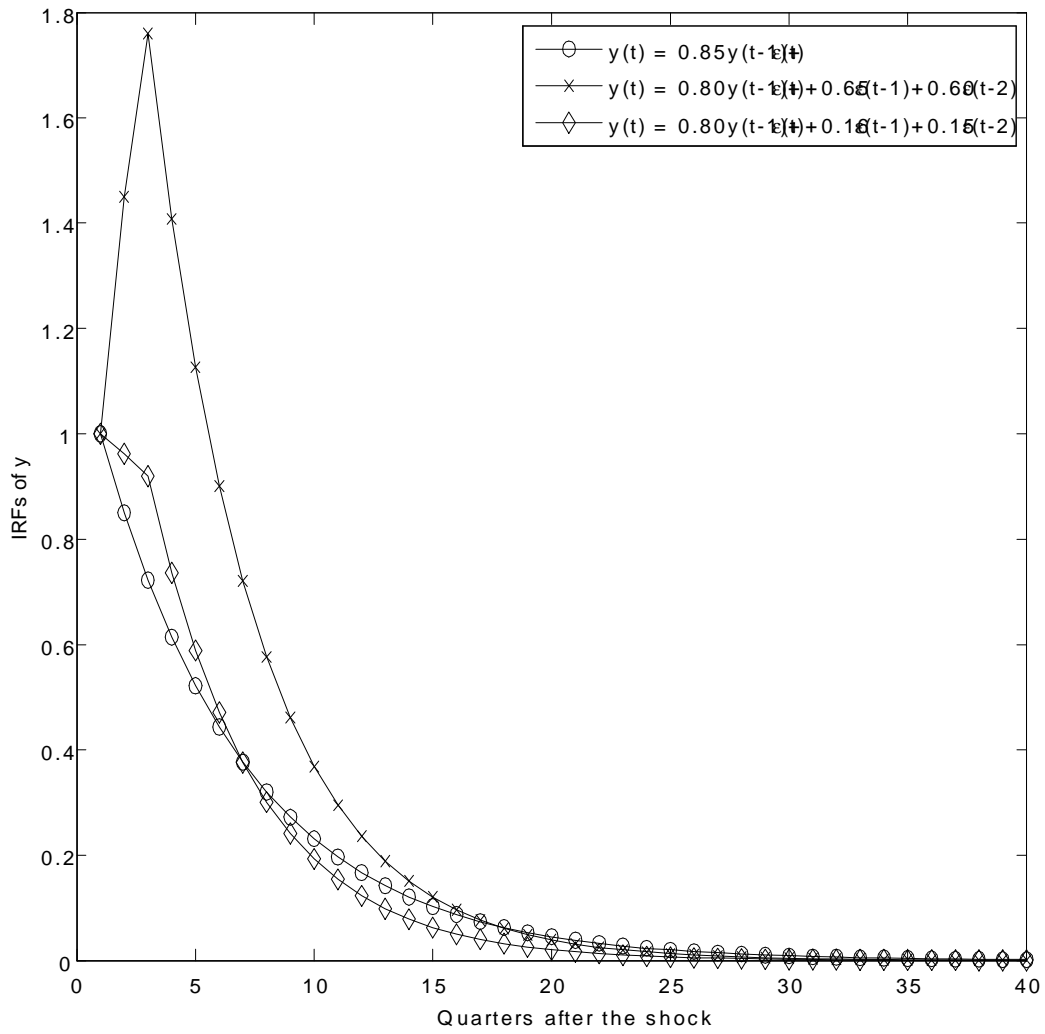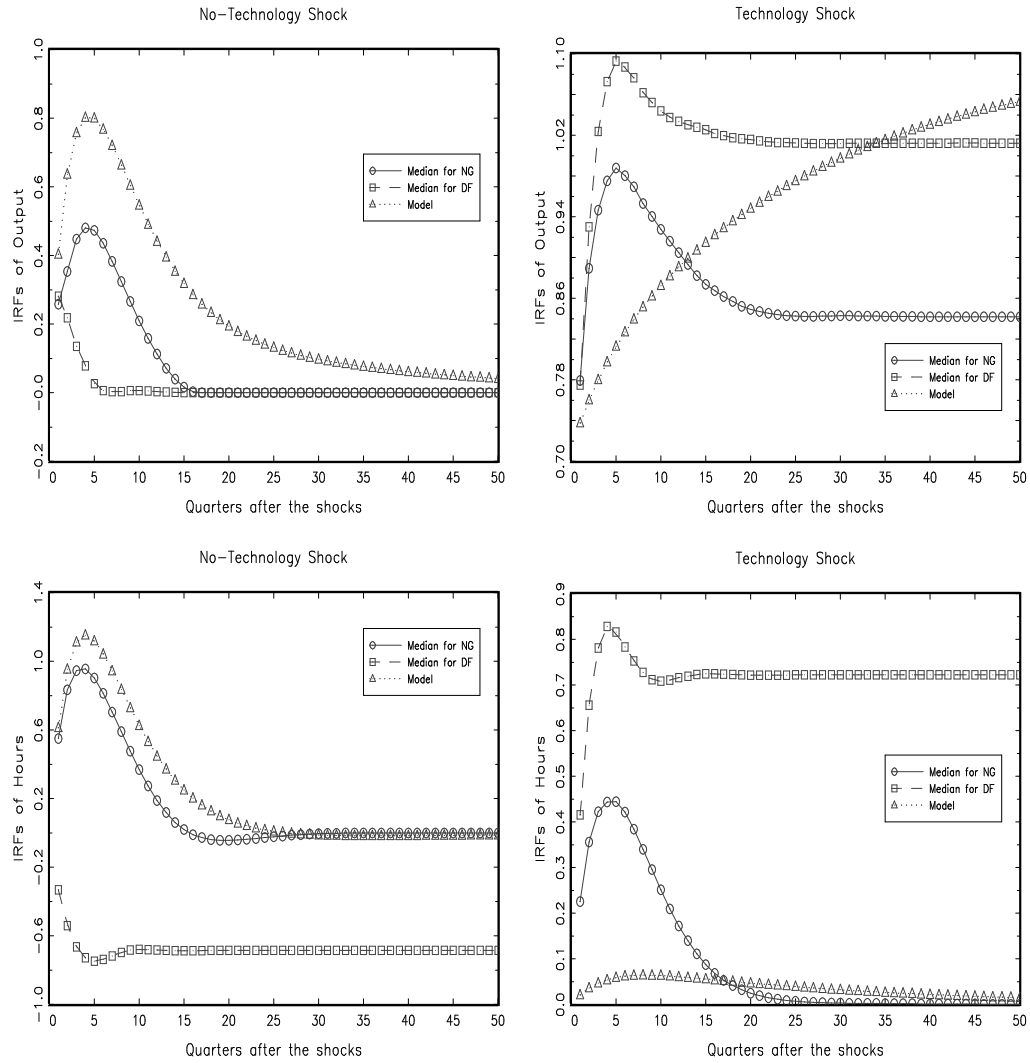
Figure 2: IRFs for three ARMA processes.

Figure 3: Model and SVAR population responses of output and hours to no-technological and technological shocks. For SVAR lines with circles are the mean of IRFs using the NG's test and lines with suqares using the DF's test.
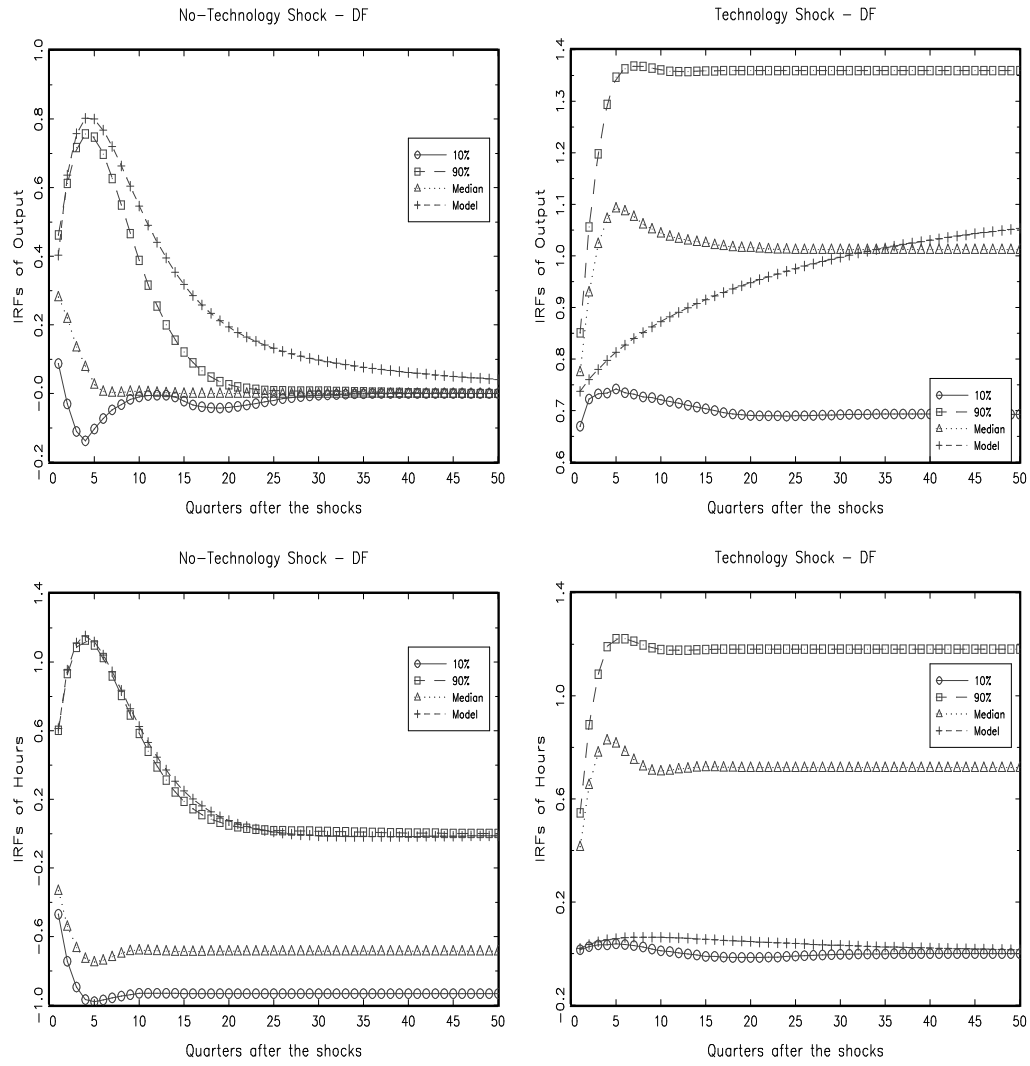
Figure 4: Model and SVAR population responses of output and hours to no-technological and technological shocks (mean and deciles) using the DF's test.
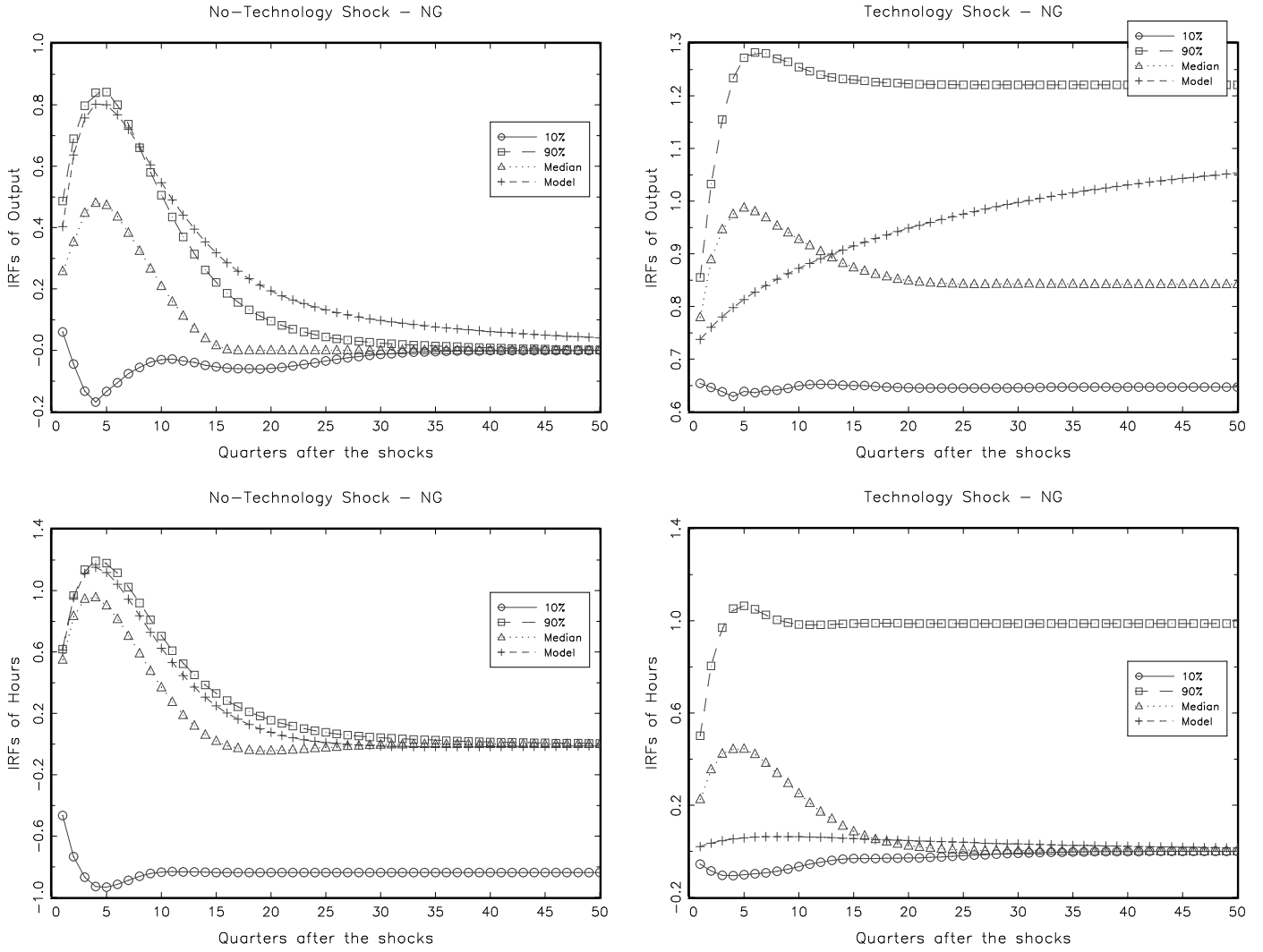
Figure 5: Model and SVAR population responses of output and hours to no-technological and technological shocks (mean and deciles) using the NG's test.