DO TECHNOLOGY SHOCKS DRIVE HOURS UP OR DOWN? A LITTLE EVIDENCE FROM AN AGNOSTIC PROCEDURE

ELENA PESAVENTO
Emory University

BARBARA ROSSI
Duke University

This paper analyzes the robustness of the estimate of a positive productivity shock on hours to the presence of a possible unit root in hours. Estimations in levels or in first differences provide opposite conclusions. We rely on an agnostic procedure in which the researcher does not have to choose between a specification in levels or in first differences. We find that a positive productivity shock has a negative impact effect on hours, but the effect is much shorter lived, and disappears after two quarters. The effect becomes positive at business-cycle frequencies, although it is not significant.

Keywords: Technology Shocks, Positive Productivity, Hours Worked

1. INTRODUCTION

According to real-business-cycle models, hours worked should rise after a positive permanent shock to technology. However, the empirical validity of this theoretical implication has been questioned in the recent literature. For example, Gali (1999) identifies technology shocks as the only shocks that have an effect on labor productivity in the long run, and estimates a persistent decline of hours in response to a positive technology shock. As Gali (1999) points out, this result is more consistent with the predictions of a New Keynesian model than those of standard real-business-cycle models. Other papers have reached similar conclusions [see, e.g., Shea (1999), and Francis and Ramey (2001)], which spurred a line of research aimed at developing general equilibrium models that can account for this empirical finding [see, e.g., Francis et al. (2003), Uhlig (2003), and Gali and Rabanal (2004)].

In a recent paper, Christiano et al. (2003) challenge these empirical results. Using the same identifying assumption as Gali (1999), Christiano et al. (2003)

478

find evidence that a positive technology shock drives hours worked up, not down. It seems that the estimated effects of technology shocks crucially depend on whether the empirical analysis is specified in levels or in differences. In fact, Gali (1999), Shea (1999), and Francis and Ramey (2001) specify hours in *first differences* and report that hours worked *fall* after a positive technology shock. On the other hand, Christiano et al. (2003) use hours in *levels* and report that hours worked *increase*. In the words of Christiano et al. (2003, p. 2) "the difference must be due to different maintained assumptions. As it turns out, a key culprit is how we treat hours worked."

Whether hours worked is a stationary or an exactly integrated process is then a key assumption in the current debate on the effects of technology shocks on business cycles. However, it is practically difficult to choose between specifications in levels or in first differences on the basis of unit root tests because of their low power. Pesavento and Rossi (2003) show that, in the presence of a root close to unity, impulse response function estimates and confidence bands that rely on unit root pretests have bad small sample properties (in terms of median unbiasedness and coverage rates). Impulse responses based on VAR's estimated in levels or first differences have bad coverage properties as well, unless the true data generating process is not persistent (in which case, levels are appropriate) or it has an exact unit root (in which case, first differences are appropriate).

We provide empirical evidence based on an agnostic empirical estimation procedure proposed by Pesavento and Rossi (2003). The estimation is agnostic in that it does not impose either a unit root or stationarity. These authors show that their method is robust to the presence of highly persistent processes and thus it is appropriate if the researcher aims at analyzing the long-run effect of technology shocks on hours worked without making assumptions on the order of integration of the series. We find that a positive productivity shock has a negative impact effect on hours worked, but this effect disappears more quickly than in Francis and Ramey (2001) (after only two quarters), and it becomes quickly positive.

2. METHODOLOGY

Let the data generating process (hereafter DGP) be

$$(I - \Phi L)w_t = u_t, \quad t = 1, 2, \ldots T, \tag{1}$$

where $w_t = [n_t \ f_t]'$ is a (2×1) vector of variables, where n_t is the log of per-capita hours worked in the business sector and f_t is average labor productivity; u_t is a (2×1) stationary and ergodic moving-average sequence,

$$u_t = \Theta(L)\epsilon_t, \tag{2}$$

ϵ_t is a martingale difference sequence with covariance matrix $\Sigma, \Theta(L) \equiv \sum_{i=0}^{\infty} \Theta_i L^i, \Theta_0 = I, I$ is the (2×2) identity matrix, and $\Omega^{1/2} \equiv \Theta(1)\Sigma^{1/2}$ is invertible.

Note that (1) and (2) are simply another way of writing a VAR, in terms of the roots rather than in the usual linear expression with lagged endogenous variables. This representation is convenient for our purposes because it distinguishes the long-run dynamics, captured by Φ, from the short-run dynamics, described by $\Theta(L)$. In fact, to allow a unit root in f_t and high persistence in n_t, we let

$$\Phi = \begin{bmatrix} \rho & 0 \\ 0 & 1 \end{bmatrix},$$

where ρ is close to 1 in a sense made precise below.

The objects of interest are the structural shocks, denoted by η_t, which are related to the VAR's residuals ϵ_t by the following relationship:

$$\varepsilon_t = A_0 \eta_t \tag{3}$$

We let $\eta_t = [\eta_t^m \ \eta_t^z]'$ where $\{\eta_t^z\}$ and $\{\eta_t^m\}$ denote, respectively, the sequence of technology and nontechnology shocks. Following Gali (1999), we identify the technology innovation as the only shock that can have a permanent effect on productivity. This long-run identification imposes a lower triangular structure to $\Theta(I) A_0$ that allows the identification of the technology shock.

Let us first provide some intuition about how our "agnostic" method works by discussing what our method would deliver at long horizons. As in Pesavento and Rossi (2003), we use a local-to-unity asymptotic theory to improve the asymptotic approximation to highly persistent processes in small samples. That is, we model the largest root associated with hours, ρ, as local-to-unity:

$$\rho = 1 + \frac{1}{T} c \tag{4}$$

To obtain better asymptotic approximations to IRF's in small samples, we also assume that the lead time of the impulse response function, h, is a fixed fraction of the sample size,

$$\frac{h}{T} \xrightarrow[T \to \infty]{} \delta. \tag{5}$$

Note that, because of assumption (5), the method works very well at horizons (h) that are large relative to the available sample size, which is what we refer to as "long horizons."

Considering the two assumptions (4) and (5) together, we have

$$\rho^h \xrightarrow[T \to \infty]{} e^{c\delta}.$$

Pesavento and Rossi (2003) show that the IRF of the effect of a technology shock, η_t^z, on n_t can be approximated by

$$\frac{\partial n_{t+h}}{\partial \eta_t^z} \simeq e^{c\delta} i_1' \Theta(1) A_0 i_2, \tag{6}$$

where i_s denotes the sth column of the $m \times m$ identity matrix. This provides a simple, closed-form formula for the IRF's at long horizons as a monotone increasing function of c. This formula can easily be used to construct confidence intervals for the IRF at *long horizons*.[1]

To construct IRF's that are valid at *short horizons* as well, which is what we do in this paper, the method is implemented in practice as follows:

(i) We construct a confidence interval for c [denoted by $(c_L; c_U)$] by inverting the acceptance region of a unit root test for hours. One could use any unit root test; in this paper, we use the Augmented Dickey-Fuller (ADF) test and Hansen's (1995) CADF test. In our case, the estimated ADF test statistic is -2.068. Thus, directly from Stock (1991, Table A1, pp. 455–456), inverting the ADF test delivers a confidence interval for c equal to $(-13.73, 2.411)$.

(ii) We run a VAR's in quasi differences, $(I - \hat{\Phi}L)w_t$,[2] to estimate $\Theta(L)$, and construct a 95% confidence interval for $\Theta(L)A_0$ by using a standard Monte Carlo simulation method [see Hamilton (1994) and Lütkepohl (1993) for details], where A_0 has been identified as above. More in detail, the confidence interval for $\Theta(L)A_0$ is obtained by simulating a confidence interval for $\Theta(L)$, and for every value belonging to the confidence interval, we estimate A_0 that satisfies the identification restriction, which we then use to obtain a confidence interval for $\Theta(L)A_0$.

(iii) For every horizon, we calculate a confidence interval for $[i_1'\Theta(L)A_0i_2']$ at the relevant horizon; call it (L_h, U_h). For example, at horizon $h = 1$ this confidence interval is $(-0.435, 0.096)$.

(iv) Finally, the Bonferroni confidence interval for the response of hours to a technology shock is $(e^{c_L\delta}L_h, e^{c_U\delta}U_h)$. In the example for $h = 1$, since $e^{c_L\delta} = 0.881$ and $e^{c_U\delta} = 1.023$, we have the confidence interval for the IRF that is $(-0.383, 0.098)$.[3] While confidence bands constructed in this way have good coverage properties at short horizons and are robust to the presence of a root close to unity, this comes at the cost of being pointwise and conservative [see Pesavento and Rossi (2003)].[4]

In the empirical section, we also report results by using Wright's (2000) method. The latter method is implemented by steps (i)–(iv) above, but replacing step (ii) with the following:

(ii′): $\Theta(L)A_0$ is reestimated conditional on every value of c within a grid over (c_L, c_U)—not only at the extremes, as we do. According to Pesavento and Rossi (2003), asymptotically the estimate of $\Theta(L)$ is consistent anyway, and we gain in computational simplicity and smaller confidence bands.

In our empirical section, we also report IRF's obtained from standard VAR's using n_t in levels as well as first differences. To estimate the confidence bands in both VAR's, we simulate the IRF distribution under a normality assumption with 1,000 Monte Carlo replications.

3. EMPIRICAL RESULTS

We use the same data as in Christiano et al. (2003), where per-capita hours are measured as the natural logarithm of hours worked in the business sector
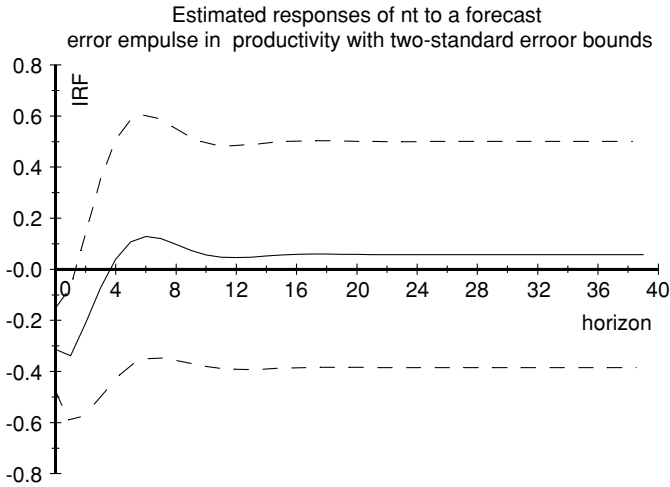
Estimated responses of nt to a forecast
error empulse in productivity with two-standard erroor bounds

FIGURE 1. Estimation of VAR in differences: estimated IRF (solid line) and IRF confidence bands (dotted line) of hours worked to a 1%-standard deviation increase in the productivity shock.

divided by a measure of the population. Productivity is measured as the natural logarithm of output per hour in the business sector. Data are quarterly observations from 1948:1 to 2001:4 and are ultimately taken from the DRI Economics Database.[5] As in the previous literature, we identify innovations to technology as the only shocks that have a permanent effect on the level of labor productivity. Figures 1 and 2 report the 90% confidence intervals and the estimated responses of per-capita hours to a 1%-standard-deviation positive shock to productivity by using either a VAR in differences or a VAR in levels.[6] Results from the VAR estimated *in differences* (Figure 1) are very similar to the results of Gali (1999) and Francis and Ramey (2001): Hours worked show a negative and persistent response to a technology shock in the short run. According to point estimates, the negative effect persists for one year (four quarters). Eventually, the effect becomes positive in the long run (although not significantly different from zero). When the VAR is estimated by using hours *in levels*, our results indicate that the initial response of hours is positive, although not significantly different from zero. The response is positive and statistically significant after 1 quarter, and for roughly 20 quarters.

Table 1 shows that, indeed, hours are a persistent process. The table provides both results on unit root tests on hours and empirical evidence on the magnitude of the persistence by using various methods to construct confidence intervals for the largest root. The methods are the Stock (1991) median unbiased method and those of Hansen (1995), Elliott et al. (1996), and Elliott and Jansson (2003). The Stock (1991) method is implemented as follows: First, we calculate the ADF test statistic for the time-series process of hours with four lags; then, by using the "inversion" Table A1 in Stock (1991), pp. 455–456, we recover the confidence

Estimated responses of nt to a forecast
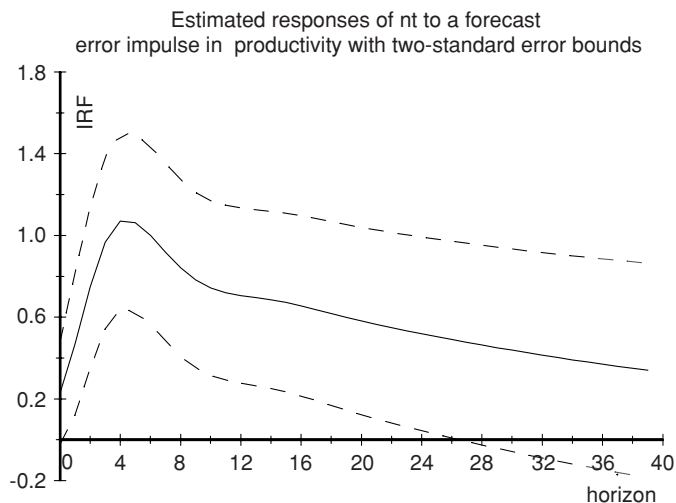error impulse in productivity with two-standard error bounds

FIGURE 2. Estimation of VAR in levels: estimated IRF (solid line) and IRF confidence bands (dotted line) of hours worked to a 1%-standard-deviation increase in the productivity shock.

interval for the largest root. Confidence intervals for the other methods are obtained in a similar fashion, although in the latter cases the inversion table may depend on nuisance parameters and thus needs to be calculated by the researcher for the specific database.

According to the Stock (1991) method, the largest root is between 0.93 and 1.01, with a median estimate equal to 0.98. With such a persistent process, it is not surprising that none of the tests is able to reject a unit root at 5% level (note that the CADF test rejects at 10%).

TABLE 1. Unit root tests on per-capita hours

			Largest Root	
Test[a]	Unit root test statistic	5% Critical value	Median estimate	95% confidence interval
ADF	−2.068	−2.88	0.977	(0.929; 1.015)
PT	23.192	3.17	1.003	(0.988; 1.022)
EJ	16.857	3.34	0.999	(0.982; 1.021)
CADF	−2.437	−2.65	0.971	(0.925; 1.007)
CADF*	−3.072	−2.54	0.951	(0.897; 0.997)

[a] Unit root tests are as follows: ADF is the Augmented Dickey Fuller *t*-test; PT is the Elliott et al. (1996) test. EJ and CADF are, respectively, Elliott and Jansson (2003) and Hansen (1995) tests, which use information on the stationary covariate, the first difference of productivity. The ADF, PT, and EJ tests are implemented with four lags, whereas the CADF test is implemented both with four lags and four leads (CADF) and with four lags (CADF*). Note that all tests reject the null hypothesis of a unit root when the test statistic is smaller than the critical value.

Given that unit root tests do not strongly support the presence of a unit root, it may not be desirable to take a stand on whether the process has a unit root or not. Kilian and Chang (2000) and Pesavento and Rossi (2003) show that in the presence of large roots the coverage rates of confidence intervals for impulse response functions constructed from VAR's in first differences or levels can be very bad in finite samples. The intuition is that a model that imposes a root equal to 1 when one of the variables is not $I(1)$ is misspecified. On the other hand, in small samples, a model in levels underestimates the largest root and the persistence of shocks. These apparently small mistakes and biases become extremely important at medium to long horizons, where the difference between stationary and nonstationary processes becomes more and more important. As a result, VAR's in levels and first differences have a very small probability of containing the true impulse response function, almost zero. Unit root pretests do not solve the problem because the actual coverage of impulse response bands obtained after a pretest can be quite different from the nominal one (due to the low power of unit root tests against persistent alternatives). Furthermore, even if the tests reject a unit root, asymptotic approximations that rely on highly persistent regressors are expected to provide better approximations in small samples. Thus, we use the Pesavento and Rossi (2003) *agnostic method* to estimate median unbiased impulse response functions and their confidence bands, which does not require the researcher to choose between the two specifications. By using the local-to-unity parameterization, we model the persistency of the process as a function of the location parameter c (see the preceding section for details), which measures how close to unity the largest root of the process is.

Figure 3 reports results for the agnostic method. It shows a negative and very short-lived impact effect, which is very much in accordance with the findings of Francis and Ramey (2001). The negative effect lasts only two quarters, less than in Francis and Ramey (2001), and it is significant on impact. At business-cycle frequencies, the median point estimate of the impulse responses is positive, although not significantly different from zero. The confidence bands show that the effect is very likely to be positive at long horizons and at business-cycle frequencies (between six quarters and eight years). Comparing our median unbiased estimate of the response with that of VAR's in differences, we find some evidence that the medium and long horizon effect is more positive and slightly larger in magnitude. On the other hand, the effect that we estimate is also more persistent than that obtained from VAR's in levels. Finally, for comparison, Figure 4 reports results obtained by using methods such as Wright (2000).[7] The results are similar, except that the confidence bands are larger. We also checked the robustness of the results to the use of Hansen's (1995) unit root test, which exploits information on stationary covariates, and we find very similar results (see Figure 5).[8]

Our results are also similar to those obtained by using the Anderson and Rubin (1949) robust confidence intervals, which also were reported by Vigfusson (2004, pp. 11–12). In fact, Vigfusson (2004) finds that, in a VAR's estimated in levels, the impact response of hours to a 1-standard-deviation shock can be negative
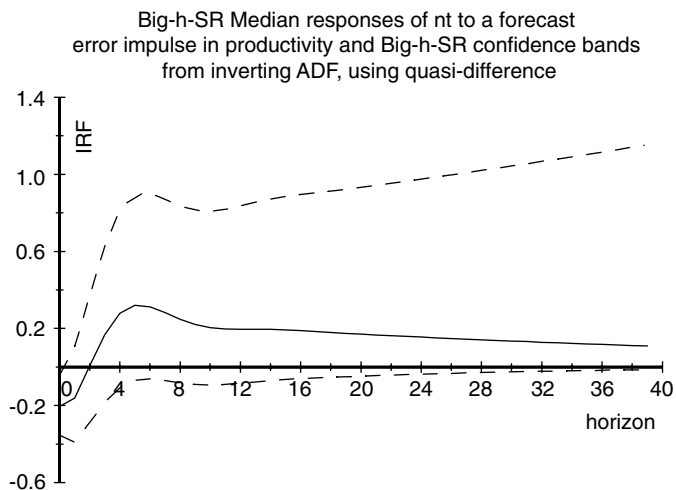
Big-h-SR Median responses of nt to a forecast
error impulse in productivity and Big-h-SR confidence bands
from inverting ADF, using quasi-difference

FIGURE 3. Agnostic estimation: estimated IRF confidence bands of hours worked to a 1%-standard-deviation increase in the productivity shock.

[the confidence interval is $(-0.05, 0.11)$ percent] and becomes more positive at business-cycle frequencies (the confidence interval is $(0.05, 0.27)$ percent after six quarters). In the present paper (see Figure 3), the agnostic estimation shows an impact effect that is negative, but the upper bound of the confidence interval is very close to zero; in addition, after five to six quarters, the confidence interval
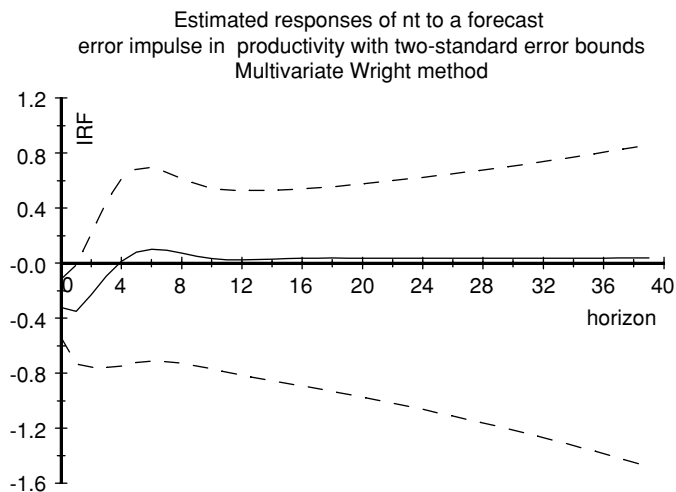
Estimated responses of nt to a forecast
error impulse in productivity with two-standard error bounds
Multivariate Wright method

FIGURE 4. Wright method: estimated IRF confidence bands of hours worked to a 1%-standard-deviation increase in the productivity shock.

Big-h-SR Median responses of nt to a forecast
error impulse in productivity and Big-h-SR confidence bands
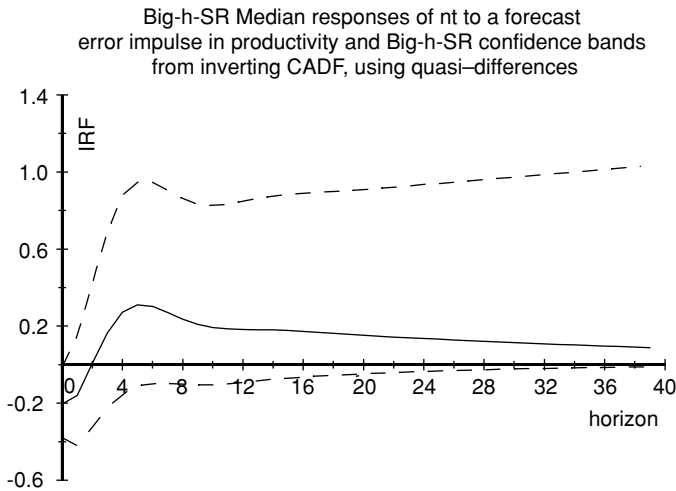from inverting CADF, using quasi–differences

FIGURE 5. Agnostic estimation CADF: estimated IRF confidence bands of hours worked to a 1%-standard-deviation increase in the productivity shock based on the Hansen (1995) test.

shifts more toward positive values, which is very much in line with what Vigfusson (2004) finds.

4. CONCLUSIONS

This paper analyzes the robustness of the estimate of the effect of a positive productivity shock on hours worked to the presence of a possible unit root in hours. Whereas the literature focuses on the cases in which hours are estimated either in levels or in first differences (a sort of "atheist" view), we rely on an agnostic procedure in which the researcher does not have to choose between the two specifications. We find that a positive productivity shock has a negative impact on hours, as in Francis and Ramey (2001), but the effect is much shorter lived than previously found, and disappears after only two quarters. The effect then becomes positive at business-cycle frequencies, as in Christiano et al. (2003), although it is not significantly different from zero.

Our empirical evidence extends the results of Christiano et al. (2003) in an important and crucial way. In their framework, the level specification implies that the first difference specification is misspecified whereas the difference specification implies that the level specification is correctly specified. The latter follows from the fact that the level VAR's allows for a unit root. Although this is true at very short horizons, this does not need to hold at horizons that are large relative to the sample size, where the possibly downward-biased estimate of the root becomes important. The importance of these biases depends on the economic problem at hand and on the particular parameters that the researcher faces. Our results show

that neglecting this effect may lead to very different economic results in measuring the effects of productivity shocks.

Possible alternative estimation methods include Bayesian methods, as described by Sims and Uhlig (1991). We do not attempt to pursue this approach in the present paper, but a thorough investigation of the performance of Bayesian methods in constructing confidence bands for impulse responses is provided by Kilian and Chang (2000).

NOTES

1. Simply use (6) to obtain the confidence interval as follows: $[e^{c_l, \delta} i_1' \hat{\Theta}(1) \hat{A}_0 \mathbf{i_2}; \ e^{c_l, \delta} i_1' \hat{\Theta}(1) \hat{A}_0 \mathbf{i_2}]$, where "hats" denote estimated values.

2. The quasi differences are obtained by taking the residuals of a VAR(1). In our empirical application,

$$\hat{\Phi} = \begin{pmatrix} 0.986 & 0.002 \\ -0.009 & 0.995 \end{pmatrix}.$$

As pointed out by a referee, since the estimated value of ρ (0.986) is very close to 1, quasi differencing gives very similar results to first differencing *at short horizons*.

3. The last two steps are equivalent (by monotonicity) to the following proceedure. For a given horizon $h = [\delta T]$, for each point on a grid within the confidence interval for $\Theta_i A_0$, construct two new sequences by multiplying each of the points in the confidence intervals by $e^{c_L \delta}$ and $e^{c_U \delta}$, respectively; call these sequences $e^{c_L \delta} \Theta_i A_0$ and $e^{c_U \delta} \Theta_i A_0$. The overall confidence interval for the IRF of hours to a productivity shock at horizon h is then obtained as the minimum over the first sequence and the maximum over the second sequence: $(\min e^{c_L \delta} i_1' \Theta_i A_0 i_2; \max i_1' e^{c_U \delta} \Theta_i A_0 i_2)$. By the Bonferroni inequality, the confidence interval should have a coverage of at least 90% at each horizon h. Because exponential functions are always positive, this procedure gives the same result as the procedure described in the main text. Intuitively, relative to simply using (6) with a consistent estimate of $\Theta(I)$ as described in a previous note, step (ii) adds information on the sampling variability of the short-run parameters, $\Theta(L)$, thus improving the performance of the method at short horizons.

4. Pesavento and Rossi (2003) investigate a variety of methods, all of which have good coverage. These methods build on the inversion of the following test statistics: ADF as in Stock (1991), Elliott et al. (1996), Elliott and Jansson (2003), Elliott and Stock (2001), and Elliott et al. (2005). Although we report results based on ADF only, our results are qualitatively robust to the use of the other methods mentioned above.

5. The mnemonics for business labor productivity, business hours, and the civilian population over the age of 16 are, respectively: LBOUT, LBMN, and P16. We thank Christiano et al. (2003) for the data.

6. The IRF's are multiplied by 100 so a value of 0.10 corresponds to a response of 0.10%. Following the cited literature, we include a constant, but not a time trend. We focus on a bivariate VAR with hours worked and the productivity measure. As in Francis and Ramey (2001) and in Christiano et al. (2003), we do not expect our results to change if we include additional variables. We use four lags (chosen by the BIC criterion) in order to compare our results directly to those of Francis and Ramey (2001) and Christiano et al. (2003). Results are robust to different lags (e.g., 1 to 6) if we use quasi differences to estimate the short-run dynamics.

7. The method originally proposed by Wright (2000) is univariate. We apply a method that is, in spirit, very similar to his, but it is extended to a multivariate VAR with one large root.

8. In unreported simulations, we found that the results for the Hansen's (1995) test are robust to whether the CADF test is estimated with both leads and lags (as in Figure 5) or with lags only. We also found that the results are robust to the use of other methods to construct confidence intervals for a unit root, such as the Elliott et al. (1996) PT test and the Elliott and Jansson (2003) test.

REFERENCES

Anderson, T.W. & Herman Rubin (1949) Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20, 46–63.

Christiano, Larry, Martin Eichenbaum & Robert Vigfusson (2003) What Happens After a Technology Shock? Mimeo, Northwestern University.

Elliott, Graham & Michael Jansson (2003) Testing for unit roots with stationary covariates. *Journal of Econometrics*, 115, 75–89.

Elliott, Graham & James H. Stock (2001) Confidence intervals for autoregressive coefficients near one. *Journal of Econometrics* 103, 155–181.

Elliott, Graham, Thomas J. Rothenberg & James H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Elliott, Graham, Michael Jansson & Elena Pesavento (2005) Optimal power for testing potential cointegrating vectors with known parameters for nonstationarity. *Journal of Business and Economic Statistics* 23, 34–38.

Francis, Neville & Valerie Ramey (2001) Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited. Mimeo, University of California at San Diego.

Francis, Neville, Michael Owyang & Athena T. Theodorou (2003) The use of long-run restrictions for the identification of technology shocks. *Federal Reserve Bank of St. Louis Review* 85(6), 53–66.

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

Gali, Jordi & Paul Rabanal (2004) Technology shocks and aggregate fluctuations: How well does the RBC model fit postwar U.S. data? In Mark Gertler & Kenneth Rogoff (eds.), *NBER Macroeconomic Annual*, pp. 225–288. Cambridge, MA: MIT Press.

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

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

Kilian, Lutz & Pao-Li Chang (2000) How accurate are confidence intervals for impulse responses in large VAR models? *Economics Letters* 69, 299–307.

Lütkepohl, Helmut (1993) *Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag.

Pesavento, Elena & Barbara Rossi (2003) Small sample confidence intervals for multivariate impulse response functions at long Horizons. *Working paper 03–19*, Duke University.

Shea, John (1999) What do technology shocks do? In Ben Bernanke & Julio Rotemberg (eds.), *NBER Macroeconomics Annual*, pp. 275–310. Cambridge, MA: MIT Press.

Sims, Chris & Harald Uhlig (1991) Understanding unit rooters: A helicopter tour. *Econometrica* 59, 1591–1599.

Stock, James H. (1991) Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, 435–459.

Uhlig, Harald (2003) Do Technology Shocks Lead to a Fall in Total Hours Worked? Mimeo, Umbolt University.

Vigfusson, Robert (2004) The Delayed Response to a Technology Shock: A Flexible Price Explanation. Discussion paper 810, Board of Governors of the Federal Reserve Board International Finance.

Wright, Jonathan (2000) Confidence intervals for univariate impulse responses with a near unit root. *Journal of Business and Economic Statistics* 18, 368–373.