Nonstationarity and Nonlinearity in the US Unemployment Rate: A Re-examination

Dipak Ghosh and Swarna (Bashu) Dutt[1]

ABSTRACT

Conventional econometric tests cannot distinguish nonstationarity from nonlinearity because of the joint modeling of unit roots with threshold effects. Caner–Hansen (CH, 2001) provides a new test which for the first time can simultaneously test for both (without any prior assumption of stationarity). Their threshold unit root tests are more powerful than conventional Augmented Dickey-Fuller tests, especially when the true process is nonlinear. They look at unemployment among adult males, and find contrary to many previous studies, that it is a "stationary nonlinear threshold process". This paper attempts to re-examine and reconfirm the CH methodology by using unemployment in the civilian labor force. We extend the data up to December 2004, to see if the results hold up to the recent turbulent times, when unemployment changed dramatically from 3.9 % (1999) to 6.2 % (2003). Our results support the premise that US unemployment is a stationary threshold autoregressive process.

Introduction

> "Two key features of US unemployment, which are well documented in the literature, are that shocks to the series seem rather persistent and that it seems to rise faster during recessions than it falls during expansions. The first feature is commonly called long memory, ……..The second feature is commonly called nonlinearity……"[2]

The concern over the slow recovery of the U.S. unemployment rate even when the U.S. economy is growing out of a recession ties in directly to this statement by van Dijk et. al. (2002). A study of nonlinearity in the unemployment data is therefore particularly appropriate. Since US unemployment has always exhibited an asymmetric behavior (for example steep increases ending in sharp peaks, alternating with a gradual and longer decline), theory suggests the presence of nonlinearities, and hence the application of nonlinear statistical methods seems appropriate, which is what has been attempted here.

Presence of a unit root (nonstationarity, absence of mean reversion) would imply that the data series in question moves in a random manner (a random walk) over time, whereas absence of a unit root (stationarity, mean reversion) implies that the data reverts to a mean value over time. The traditional tests cannot, however, distinguish between non-stationarity and non-

[1] Dipak Ghosh, Associate Professor of Economics, Department of Accounting & IS, Emporia State University, Emporia, KS; Swarna (Bashu) Dutt, Professor of Economics, Richards College of Business, University of West Georgia, Carrollton, GA. The authors would like to thank Bruce Hansen for making the GAUSS programs for estimating the Caner-Hansen procedure available. Dutt would like to thank the University of West Georgia for their faculty research grant (#: 10000-1014408-12100-11000: 2007-08).

[2] Dijk, et. al. (2002)

linearity. Linearity refers to the property that the econometric model describing the data remains stable over time. When the model changes during the sample period, the data are non-linear. For example, if the Fisher equation for the United States is estimated, a change in the model in the late 1970s and early 1980 is expected due to the oil price shocks and subsequent Federal Reserve policy. Traditional unit root tests, such as the Augmented Dickey-Fuller (ADF, 1979, 1981), the Phillips-Perron (1988), and the KPSS (1992), interpret this change in the model parameters as non-stationarity. Nevertheless, the model has undergone a shift in the parameters before and after the event (oil price shocks) and could very well be stationary if we run the tests in the pre and post event data separately. Since we do not know *a priori* whether there is a shift in the model parameters, we cannot test the two sub-samples separately. Therefore, we need an econometric test that can distinguish between non-stationarity and non-linearity. The Caner-Hansen procedure is one such test.

The threshold autoregressive (henceforth TAR) models introduced by Tong (1978) lie in the forefront of nonlinear techniques, but these models cannot simultaneously distinguish between nonstationarity and nonlinearity. Recent examples include studies by Chan (1991, 1993), Chan and Tsay (1998) and Hansen (1996, 1997, and 2000). In all of these, the maintained hypothesis is that the data are stationary (no unit roots), then nonlinearity and regime shifts were tested for. To date there is no statistical distribution theory to distinguish non-stationarity from nonlinearity, without assuming stationarity *a priori*.

Caner-Hansen (henceforth CH, 2001) is the first attempt at developing a rigorous asymptotic theory which simultaneously tests for both effects. A Wald test is developed to detect thresholds, with both Wald and "t" tests for unit roots. A Wald test is a test of restrictions on a model (similar to the Lagrange Ratio and Lagrange Multiplier tests). It is more effective than the other available tests when the model under the null hypothesis is easier to estimate than the model under the alternate hypothesis. This is the case for our model, since the model under the null hypothesis is the linear model and the model under the alternate hypothesis is the non-linear model. CH test for both stationarity and linearity in US unemployment (among adult males). They find that it is a stationary nonlinear process. We confirm these results by using a broader unemployment series, unemployment in the civilian labor force, and extend it to 2003. This takes into consideration the most recent volatile period when unemployment ranged from as low as 3.9 % (1999) to as high as 6.2% (2003). Our results support the findings of CH, signifying that this other measure of unemployment is also a stationary, but nonlinear process.

Literature Review

This study is prompted by the lack of unanimity in the literature on US unemployment. Here, a few recent, but important contributions in unemployment dynamics are discussed. Our starting point is Hansen (1997), who constructed a confidence interval estimate under TAR models. He tests for the presence of nonlinearities in US unemployment among males age 20 and over, using standard ADF tests of nonstationarity. Nonlinearity is rigorously tested using two different threshold choices. He reports clear regime shifts, one for decreasing and one for increasing unemployment. Our concern with Hansen's tests is that conventional ADF unit root tests have very low power in TAR models, as demonstrated by Pippenger and Goering (1993).

Montgomery, *et al* (1998) study the forecasting performance of multiple econometric time series models (ARIMA, VARMA, TAR and MSA etc.) in regard to the US unemployment rate. Both linear and nonlinear techniques, as well as a combination of the two, are applied to determine their relative strengths and weaknesses. Since US unemployment had always exhibited an asymmetric pattern (for example, steep increases ending in sharp peaks, alternating with gradual and longer declines), theory suggests the presence of nonlinearities, and hence the application of nonlinear statistical methods. Using minimum mean square error as the testing criteria for model credibility, they find that nonlinear models significantly improve forecasting performance. Even better results were evident when these models were combined with univariate TAR methods. Nonlinearities could not be fully exploited given the state of the literature at that point. This would change only after the CH 2001 test. Finally, Chen and Tsay (henceforth CT, 1998) start with a two regime TAR model, a substantial improvement over standard TAR models.[3] This continuous autoregressive model is applied to the US civilian unemployment data, which exhibits clear nonlinear characteristics, evident from its asymmetric cyclical behavior.

Caner-Hansen Model:

In all the studies mentioned above and in the literature examining regime shifts (shifts in the parameters of the model describing the data), the maintained assumption is that the data under consideration are ergodic[4] and stationary. The tests then conducted are for data series nonlinearity and its type. CH (2001) is the first rigorous treatment of the simultaneous existence of both nonstationarity and nonlinearity. There are Wald and "t" tests for unit roots, and a sequential Wald test for threshold effects. The Wald test of nonlinearity has a nonstandard asymptotic null due to an unidentified parameter under the null hypothesis (Hansen, 1996). This null hypothesis has two components, one reflecting the unit roots, but free from nuisance parameters, and the other similar to the stationary case, but dependent on nuisance parameters. The resulting distributions are non-standardized and have to be derived in every case. The unit root Wald test has an asymptotic null distribution, depending on whether there is a threshold effect or not. These tests are more powerful than the conventional Augmented Dickey-Fuller unit root tests when the true process is indeed nonlinear.[5] Moreover, conventional unit root tests consistently fail to reject the hypothesis that post war unemployment is non-stationary, mainly because of their inability to jointly model unit roots and regime shifts. A technical description of the Caner-Hansen procedure is given in the appendix.

Data

We use monthly data for unemployment in the civilian labor force for the period January 1948 to December 2004. The data were obtained from the website of the Bureau of Labor Statistics. A graph of the unemployment data is provided in Figure 1. The nonlinearity seems to show up in the rapid rise in the unemployment rate during recessions, followed invariably by a more gradual decline during an expansion.

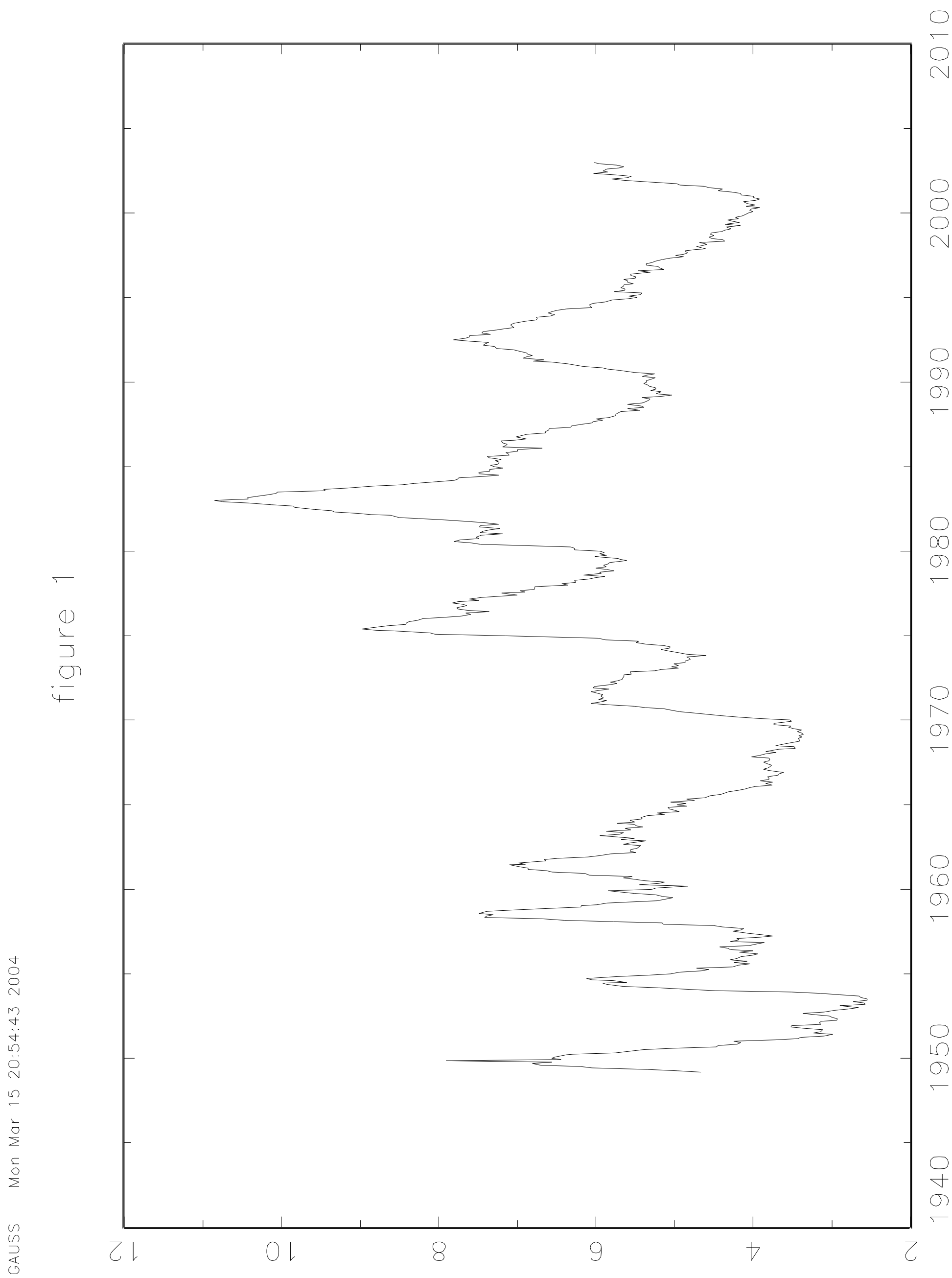
---

[3] See Tiao and Tsay (1994) who built a two-regime TAR model to study the dynamics of the US real GNP. It shows clear evidence that the true autoregressive function is continuous everywhere. Hence a new model is applied to unemployment in Chan and Tsay (1998).

[4] Ergodicity implies that in a time series, every observation will contain at least some unique information. Ergodicity and stationarity are necessary for estimation of parameters.

[5] Tsay (1997) introduces unit root tests in the presence of threshold effects, but the autoregressive lags are constant across regimes (which is not true here) making it a special case of the CH methodology. Also, Gonzalez and Gonzalo (1998) examine a TAR(1) model with nonstationarity, but of a particular geometrically ergodic type.

Unemployment in the U.S. Civilian labor force



Non-stationarity and Nonlinearity Test Results [6]

Preliminary tests conducted using the standard Augmented Dickey-Fuller (ADF) procedure indicates the unemployment series is nonstationary in line with the literature (estimate of ρ is –0.015 and its t-statistic, which is the ADF statistic, is –2.73, which is insignificant (less than the critical value), and indicates a unit root in the linear model. Even if the model is non-linear (model parameters change from one sub-sample to another beyond a certain threshold), however, a unit root could result. This may occur even if the data are indeed stationary in the two sub-samples separately. The jump in the model from one set of parameters to another (caused by a precipitating economic event) could, in itself, indicate the apparent presence of a unit root in the data. In order to determine whether we have a linear model (no change in model parameters) with a unit root, or a non-linear model with no unit root, or a non-linear model with a unit root, we apply the Caner-Hansen procedure to our data.

Table 1
Threshold and Unit Root Tests: Unconstrained Model

Bootstrap Threshold Test				Unit Root Tests, p-Value					
				R_{1T}		t_1		t_2	
m	W_T	1%C.V.	p-Value	Asym	Boot	Asym	Boot	Asym	Boot
1	62.3	41.3	0.0002	0.174	0.0974	0.0896	0.0358	0.955	0.749
2	46.2	41.8	0.0049	0.226	0.134	0.403	0.174	0.486	0.210
3	51.2	42.0	0.0011	0.240	0.145	0.136	0.0590	0.944	0.690
4	69.8	41.2	0.0000	0.148	0.0908	0.256	0.110	0.511	0.229
5	60.0	41.4	0.0002	0.0526	0.0350	0.411	0.178	0.127	0.0513
6	68.7	40.7	0.000	0.0855	0.0571	0.258	0.114	0.328	0.140
7	83.6	41.1	0.000	0.117	0.0747	0.344	0.156	0.323	0.132
8	87.9	40.5	0.000	0.0544	0.0370	0.341	0.153	0.164	0.0654
9	79.9	41.4	0.000	0.0877	0.0562	0.111	0.0480	0.643	0.304
10	77.9	40.3	0.000	0.145	0.0947	0.216	0.0941	0.575	0.262
11	80.3	40.7	0.000	0.159	0.0977	0.163	0.0732	0.728	0.372
12	71.5	40.4	0.000	0.100	0.0651	0.532	0.248	0.167	0.0654

Notes: Bootstrap p-values are calculated from 10,000 replications.

The Wald statistic in Table 1 tests for the existence of a threshold, a point in the data where the model parameters change from one level to another, or, in other words, the existence of non-linearity in the data. From Table 1, the Wald statistic, W_t, for threshold variables of the form $Z_t = y_t - y_{t-m}$ with delay parameters m =1,…,12, is highly significant across all lags (the p-value is less than 0.01). This implies rejection of the null hypothesis of a linear model in favor of a threshold model at the 1 percent level. The results are sensitive to the choice of "m", making it necessary to select "m" endogenously. That is, one must first estimate m, instead of assuming a certain value of m, and use the estimated value of m in the rest of the procedure. The

6 The econometric tests were done using GAUSS. The software has been made available by Bruce Hansen on his website.

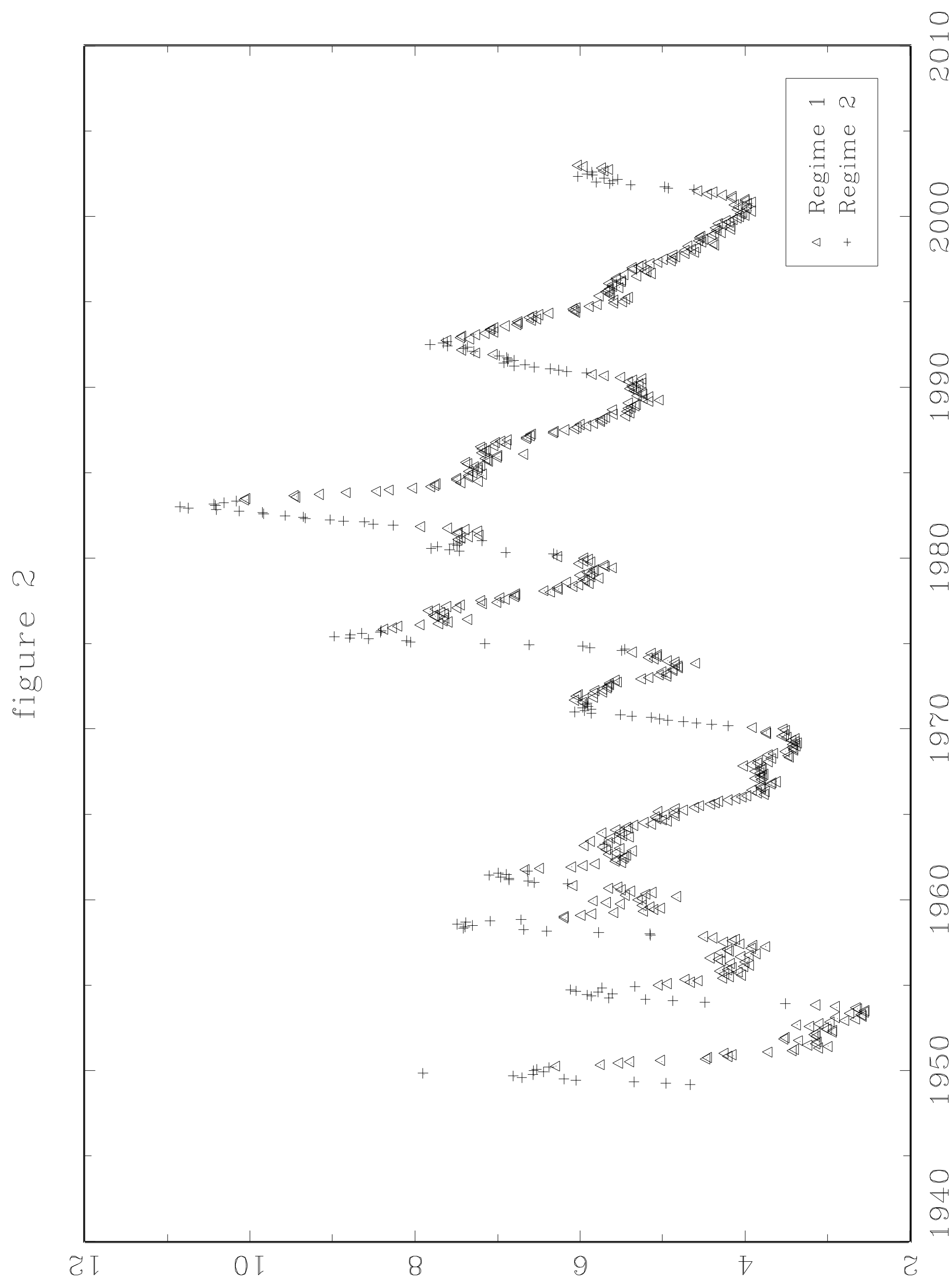
least squares estimate of m is equivalent to determining "m" such that W_T is maximized. According to Table 1, this corresponds to a value of m=8. We then calculate the threshold unit root test statistic $R^1{}_t$, t_1 and t_2, for all lags. Out of the 12 bootstrap p-values of $R^1{}_t$, 10 are significant at the 10 percent level and 2 are significant at the 5 percent level. At m=8, the bootstrap p-value of t_1=0.153 (insignificant) and of t_2=0.0654 (significant at the 10 percent level), which is evidence of partial unit roots. In addition, examination of the actual estimates of ρ_1 and ρ_2 (omitted, but available on request), indicates stationarity in the data. These indicate a non-linear but stationary data set (there is a shift in the model, but each sub-sample, with different parameters, is individually stationary).

Conclusion

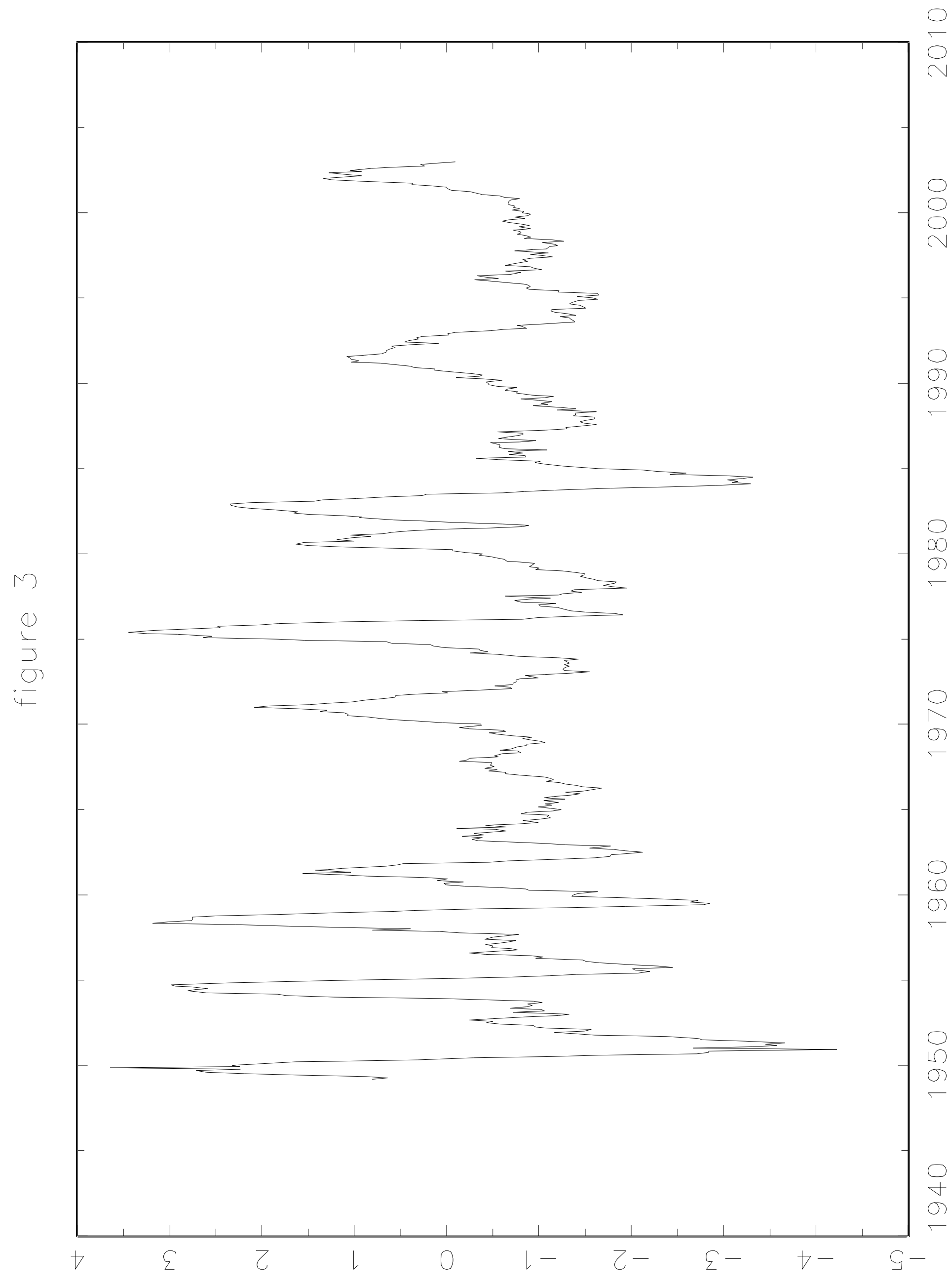
We have shown evidence in favor of the presence of stationarity in the U.S. unemployment rate after the Second World War. The pattern is visible in Figure 2, but is even more evident in Figure 3. Figure 3 shows the deviations of the change in the unemployment rate from the threshold estimate. We can see that there are significant changes around major economic events: the oil price shocks in the 1970s; the late 1970s and the early 1980s right around the time President Reagan came into office; the late 1980s and the early 1990s during the previous recession and the slow recovery; and again around 2001-2002 during that recession and subsequent recovery.

Arestis, *et al* (2002) reach a similar conclusion concerning the presence of nonlinearities in U. S. budget deficits, concluding that this is due to "asymmetries in the adjustment process." D. van Dijk, *et al* (2002) use a FI-STAR model to analyze U.S. unemployment, as suggested by Caner and Hansen (2001), which also supports both the CH results and ours. These results are particularly important in light of the recent concern over the slow recovery of the unemployment rate in the 1990s, probably due to "asymmetries" in the labor market (eg. outsourcing). Our results, as well as a glance at Figure 2, show that these asymmetries are not a new occurrence. They have always existed in the U.S. labor market. This suggests an inevitable slow recovery of unemployment during an expansion, since this has occurred frequently in the past. There may not be any new government policy that could quicken the adjustment process.

U.S. civilian unemployment rate, classified by regime



Deviations of the change in unemployment from the threshold estimate



References

Arestis, Philip, Andrea Cipollini, and Bassam Fattouh. 2002. “Threshold effects in the U.S. budget deficit,” *Working Paper, Levy Institute of Bard College.*

Caner, Mehmet and Bruce. E. Hansen. 2001. “Threshold autoregression with a unit root,” *Econometrica*, 69: 1555-96.

Chan, Kung Sik. 1991. “Percentage points of likelihood ratio tests for threshold autoregression.” *Journal of the Royal Statistical Society*, *Series B,* 53: 691-96.

Chan, Kung Sik. 1993. “Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model.” *The Annals of Statistics*, 21: 520-33.

Chan, Kung Sik and Ruey S. Tsay. 1998. “Limiting properties of the least squares estimator of a continuous threshold autoregressive model.” *Biometrika,* 45: 413-26.

Dijk, Dick van, Philip Hans Franses, and Richard Paap. 2002. “A nonlinear long memory model, with an application to US unemployment. *Journal of Econometrics*,110: 135-165.

Dickey, David A. and Wayne A. Fuller. 1979. “Distribution of the Estimators for Autoregressive Time Series with a Unit Root.” *Journal of the American Statistical Association,* 74: 421- 31.

Dickey, David A. and Wayne A. Fuller. 1981. “Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root.” *Econometrica,* 49: 1057-72.

Gonzalez, Martin and Jesus Gonzalo. 1998. “Threshold unit root models.” *U. Carlos II de Madrid.*

Hansen, Bruce E. 1996. “Inference when a nuisance parameter is not identified under the null hypothesis.” *Econometrica,* 64: 413-30.

Hansen, Bruce E. 1997. “ Inference in TAR models.” *Studies in Nonlinear Dynamics and Econometrics*, 1: 119-31.

Hansen, Bruce E. 2000. “ Sample splitting and threshold estimation.” *Econometrica*, 68: 575-603.

Kwiatkowski, Dennis, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root.” *Journal of Econometrics* 54: 159-78.

Montgomery, Alan L., Victor Zarnowitz, Ruey S. Tsay, and George C. Tiao. 1998. “ Forecasting the US unemployment rate.” *Journal of the American Statistical Association,* 93: 1035-70.

Phillips, Peter C.B. and Pierre Perron, (1988), "Testing for a unit root in time series regression", *Biometrika*, 75: 335-46.

Pippenger, Michael K. and Gregory E. Georing. 1993. “A note on the empirical power of unit root tests under threshold processes.” *Oxford Bulletin of Economics and Statistics*, 55: 473-81.

Tiao, George C. and Ruey S. Tsay. 1994. “Some advances in non-linear and adaptive modeling in time series.” *Journal of Forecasting*, 13: 109-31.

Tong, Howell. 1978. “On a threshold model.” in *Pattern Recognition and Signal Processing,* C. H. Chen, ed., Amsterdam: Sijhoff and Noordhoff.

Tsay, Ruey S. 1997. “Unit root tests with threshold innovations.” *University of Chicago.*

Appendix

A standard TAR model is

$$\Delta y_t = \theta_1' x_{t-1} 1_{(z_{t-1} < \lambda)} + \theta_2' x_{t-1} 1_{(z_{t-1} \geq \lambda)} + \varepsilon_t \tag{1}$$

where e_t is an i.i.d. error process and λ is an unknown threshold, within the interval $\lambda \in \Lambda = [\lambda_1, \lambda_2]$ where each segment has a significant presence to be dubbed a regime. The i.i.d errors ensure that the first difference of the series Δy_t is stationary and ergodic, so that y_t is itself integrated of order one. The regimes (i.e, the TAR models) are estimated by least squares.

$$\Delta y_t = \hat{\theta}_1(\lambda)' x_{t-1} 1_{(Z_{t-1<\lambda})} + \hat{\theta}(\lambda)' x_{t-1} 1_{(z_{t-1 \geq \lambda})} + \hat{\varepsilon}_t(\lambda) \tag{2}$$

The threshold λ is estimated by minimizing $\sigma^2(\lambda)$:

$$\hat{\lambda} = \arg\min \hat{\sigma}^2(\lambda), \lambda \in \Lambda \tag{3}$$

The first difference model :

$$\Delta y_t = \theta_1' x_{t-1} 1_{(z_{t-1} < \lambda)} + \theta_2' x_{t-1} 1_{(z_{t-1} \geq \lambda)} + \hat{\varepsilon}_t \tag{4}$$

is estimated using the standard Wald and "t" statistic. Here the statistics are standard, but the sampling distribution is non standard. The test in equation1 is for the presence of threshold effects, under the joint hypothesis $H_0 : \theta_1 = \theta_2$, implying no regimes. This is the Wald statistic where

$$w_t = T\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}} - 1\right) \tag{5}$$

Under the null of no threshold effects, λ is not identified, hence the testing procedure is nonstandard. The null hypothesis of $H_0 : \theta_1 = \theta_2 = \theta$, simplifies the model to

$$\Delta y_t = \rho y_{t-1} + \mu + \alpha' . \tilde{\Delta} y_{t-1} + \varepsilon_t \tag{6}$$

where $\tilde{\Delta} y_{t-1} = (\Delta y_{t-1} \ldots\ldots \Delta y_{t-k})'$ are two bootstrap methods, one for stationarity and other for the nonstationary case. Since the order of integration is unknown (true for most situations) the authors recommend calculating "ρ" both ways, and drawing our inference on the larger value. In testing for unit roots and nonstationarity CH discuss three possibilities. In equation 1, ρ_1 and ρ_2 are the determinants of stationarity of y_t. Under the null hypothesis,

1) $H_0 : \rho_1 = \rho_2 = 0$, Δy_t is stationary, indicating y_t is an I(1) process

2) If $\rho_1 < 0$, $\rho_2 < 0$ and $(1+\rho_1)(1+\rho_2) < 1$, then the series is stationary and ergodic

3) $H_1 : \rho_1 < 0$ and $\rho_2 < 0$

What if it is the intermediary case of a partial unit root ? Then,

$$H_2 : \begin{array}{c} \rho_1 < 0 \text{ and } \rho_2 = 0 \\ \text{or} \\ \rho_1 = 0 \text{ and } \rho_2 < 0 \end{array}$$

Here y_t is stationary in one regime and nonstationary in another. Their test can distinguish amongst the three. The difficulty is that the null of a unit root ($\rho_1 = \rho_2 = 0$) is compatible with both the existence of a threshold ($\theta_1 \neq \theta_2$) and the nonexistence of a threshold ($\theta_1 = \theta_2$). But CH determines that the assumptions of these two situations are different and hence we can simultaneously distinguish between nonstationarity and nonlinearity. Using theorems 5 and 6 of CH, the distinction between linearity and nonlinearity lies in the identification of the threshold parameter λ. With no threshold effects, λ is not identified, and so its estimate, $\hat{\lambda}$, is random and so is R_t. With threshold effects, λ is identified, and with no randomness in R_t, it is equivalent to the case where λ_0 is known. CH recommend (with caution) the implementation of bootstraps since both the identified and unidentified effects can be imposed. The unidentified threshold bootstrap imposes the restriction $\theta = \theta_1 = \theta_2$ (no thresholds) and $\rho = 0$ (unit root). In this case the bootstrap p-value is the percentage of simulated test statistic R^b_t that exceeds R_t. The identified threshold bootstrap requires simulation of the TAR process, and calculating R^b_t. Again the bootstrap p-value is the percentage of simulated R^b_t that exceeds R_t.[(5)] [7] Thus we conclude that in the presence of nonlinearity, the CH threshold unit root tests have more power than the standard ADF tests.

[7] CH run Monte Carlo simulations to show their relative strength vis-à-vis the conventional Dickey-Fuller (ADF) tests in the presence of thresholds. ***Case 1:*** This is where the condition $\rho_1 = \rho_2$ is imposed and $\Delta\mu = 0$ (no regimes), the ADF is more powerful than the CH threshold unit root test. But as $\Delta\mu$ increases, the R_{1t} and R_{2t} tests gather more power than the ADF test. ***Case 2***: This is where $\rho_1 = 0$, ρ_2 varies and $\Delta\mu = 0$, a partial unit root model. Here R_{1t} and R_{2t} have substantially greater power than the ADF test. The ADF test is particularly weak when $\Delta\mu$ is large. Here the t-ratio test is itself enough to distinguish between the pure unit root, the partial unit root and the stationary cases. ***Case 3***: This is where ρ_1 is fixed, ρ_2 varies and $\Delta\mu = 0$, the stationary case. Here also R_{1t} is the most powerful test with R_{2t} a close second.