STATISTICA, anno LXVI, n. 1, 2006

A NOTE ON UNIT ROOT TESTING IN THE PRESENCE OF LEVEL SHIFTS

G. Cavaliere, I. Georgiev

1. INTRODUCTION

For more than two decades economists have widely debated whether macroeconomic and financial time series can be modelled by nonstationary processes with an autoregressive unit root [UR] or, conversely, they are better described by processes which are stationary around a deterministic, generally linear, trend. In an important paper, Perron (1990) questions the empirical relevance of these two classes of processes. Specifically, he suggests that permanent shocks may occur only rarely in time, and hence, that economic time series may be better described by means of stationary processes around a deterministic trend subject to infrequent level shifts. Apart from its important macroeconomic implications, Perron's proposal is extremely important for its methodological implications: the presence of level shifts largely affects the performance of UR tests, with the evidence in favor of a UR being generally inflated. New tests which are robust to a *single* level shift have been developed, *inter alia*, by Perron (1990), Banerjee et al. (1992), Perron and Vogelsang (1992), Zivot and Andrews (1992), Amsler and Lee (1995), Saikkonen and Lütkepohl (2001, 2002), Lanne et al. (2002), Lütkepohl et al. (2004), among others.

When multiple level shifts are allowed for, the testing procedures proposed in the above mentioned papers no longer work, with the evidence in favour of a UR possibly being inflated. The few attempts to robustify these tests by allowing for more than one shift are generally feasible only when the number of shifts is very small; see Clemente et al. (1998) and the discussion in Lumsdaine and Papell (1997).

In this paper we attempt to fill the gap in the literature by proposing a simple procedure for UR testing under multiple additive level shifts. The basic idea is that level shifts should be treated as a random jump process where neither the number of shifts nor their location should be considered as known. Furthermore, opposite to the "ordinary" shocks affecting the process, level shifts should be large and should occur rarely. Given these premises, our procedure consists of two steps.

First, we suggest a preliminary detection of the level shift dates by appealing to the literature dealing with additive shift identification, see section 3. Once the level shift dates are identified, we propose to estimate the level shift component by a proper estimator. This allows us to eliminate the level shifts by simply subtracting from the original time series the estimated level shift component. The obtained time series – which will be called '*de-jumped*' in the following – can then be used to carry out standard UR tests. Notice that, by doing so, there is no need for new UR tests, as all the practitioner is required to do is to run a *standard* UR test on the de-jumped time series.

The structure of the paper is the following. In section 2 we present the reference autoregressive model with multiple level shifts and a possible UR. In section 3 we discuss the proposed UR testing procedure by focusing, in particular, on the de-jumping algorithm. In section 4 we briefly discuss some well-known level shift detection methods which can be used to implement the de-jumping technique. In section 5 we present a number of Monte Carlo simulations aiming at assessing the finite sample performance of our tests. Section 6 provides some extensions and Section 7 concludes.

2. THE MODEL

The model we are interested in is a standard unobserved components model made of (i) an autoregressive [AR] component, either stable or with a UR; (ii) a deterministic component; (iii) a (stochastic) level shift component. More specifically, the observable time series X_t satisfies:

$$X_t = \varphi' Z_t + Y_t + \mu_t ,\ t = 1-k,...,T \tag{1}$$

$$Y_t = \alpha Y_{t-1} + u_t \tag{2}$$

$$u_t = \sum\nolimits_{i=1}^{k} \gamma_i u_{t-i} + \varepsilon_t \tag{3}$$

where Y_t is the AR component, Z_t is a vector of known deterministic terms, and μ_t is the level shift component. Our interest is in testing the null hypothesis H_0: $\alpha = 1$ against the local alternative H_c: $\alpha = 1\text{-}c/T$, $c>0$, as well as against the fixed alternative H_f: $\alpha = \alpha^*$, $|\alpha^*| < 1$. The term u_t is assumed to be stable, i.e. the roots of the characteristic equation $\Gamma(z) := 1 - \sum\nolimits_{i=1}^{k} \gamma_i z^i = 0$ are all outside the unit circle. Morever, we assume that $\{\varepsilon_t\}$ is an IID$(0,\sigma^2)$ sequence. Notice that under these assumptions Y_t has at most one UR, all the other roots being stable roots. In the following, we regard the lag order k as known and we set $\varphi = 0$; the general case where k is unknown and $\varphi \neq 0$ is considered in section 6. Finally, the initial values are assumed to satisfy $\mu_{1-k},...,\mu_0 = 0$, $u_{1-2k},...,u_{-k} = O_p(1)$, with $X_{-k} = Y_{-k}$ being any real random variable whose distribution is fixed and independent of T.

The level shift component μ_t is assumed to be constant apart from a (small) number of time periods. Specifically, we assume that

$$\mu_t = \sum_{s=1}^{t} \delta_s \theta_s$$

where δ_t is a binary indicator variable which equals one if and only if a level shift occurs at time t (δ_t being zero otherwise). Notice that the single level shift model of Perron (1990), with the level shift occurring at time T^* and with shift magnitude equal to θ^*, obtains by setting $\delta_t = \mathbb{I}\{t = T^*\}$ ($\mathbb{I}\{.\}$ denoting the indicator function) and $\theta_{T^*} = \theta^*$. The number of level shifts occurring up to time t is given by $N_t := \sum_{s=1}^{t} \delta_s$. The procedure we propose in this paper is particularily designed for cases where the level shift component has the following features:

(a) N_T is bounded in probability;

(b) $\theta_t = T^{1/2}\eta_t$, where $\{\eta_t\}_{t=1}^{T}$ and $\{\eta_t^{-1}\}_{t=1}^{T}$ are sequences of $O_p(1)$ random variables as T diverges.

Condition (a) is in contrast with cases where the number of level shifts diverges together with the sample size. See, e.g., Balke and Fomby (1991a, 1991b) and Franses and Haldrup (1994). Basically, (a) allows us to preserve in the limit the distinction between ordinary shocks (frequent) and level shifts (rare).

Regarding (b), notice that this requirement fixes the stochastic magnitude order of level-shift sizes at $T^{1/2}$. It is not new, cf. Leybourne and Newbold (2000a, 2000b), Perron (1989, p. 1372), Müller and Elliott (2003). Specifically, under (b) the stochastic magnitude of the level shifts matches the stochastic magnitude order of Y_t under the null hypothesis and under local alternatives. That is, the magnitude of the shifts is not negligible if compared to the levels of the time series of interest.

Also, under (b) standard UR tests have non-pivotal asymptotic distributions in the presence of level shifts. Specifically, consider the so-called Augmented Dickey-Fuler [ADF] regression $X_t = \alpha X_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta X_{t-i}$ + error$_t$, and the ADF test statistics $ADF_{\hat{\alpha}} := T(\hat{\alpha}-1)/|\hat{\Gamma}(1)|$ and $ADF_t := (\hat{\alpha}-1)/s(\hat{\alpha})$ based on it, where $\hat{\Gamma}(1) := 1 - \sum_{i=1}^{k} \hat{\gamma}_i$ and $s(\hat{\alpha})$ is the (OLS) standard error of $\hat{\alpha}$, with $\hat{\alpha}$ and $\hat{\gamma}_i$ denoting respectively the OLS estimators of α and γ_i (i = 1,...,k). In the absense of level shifts, $ADF_{\hat{\alpha}}$ and ADF_t are pivotal; cf. e.g. Chang and Park (2002). On the other hand, if there is at least one level shift, the asymptotic distributions of these test statistics strongly depend on the level shift component, not only under the null but under local alternatives as well, leading to tests with low asymptotic power; cf. Cavaliere and Georgiev (2006, Theorem 1). To overcome this serious drawback, in the next section we will focus on how to test the UR null hypothesis H_0 in model (1)-(3) when the level shift component is present.

3. TESTS FOR UNIT ROOTS UNDER LEVEL SHIFTS

In the model (1)-(3), if Y_t was observable, then the asymptotic distributions of the ADF statistics from the regression

$$Y_t = \alpha Y_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta Y_{t-i} + \text{error}_t \tag{4}$$

would be of the standard Dickey-Fuller type, under the null hypothesis and under local alternatives. However, since Y_t is unobservable, this regression is not feasible empirically. Since $Y_t = X_t - \mu_t$ can be thought of as obtained from X_t once the unobservable level shift component has been removed, a simple UR test would obtain by conducting standard ADF tests on a time series obtained by subtracting from X_t an estimator of the level shift component μ_t. The idea is related to Saikkonen and Lütkepohl (2002), who suggest to adjust the original time series by removing an estimator of its deterministic component, including possible (deterministic) level shifts. Since in our case μ_t is a random jump process, in what follows this procedure is referred to as 'de-jumping'.

If δ_t were observable, but the shift sizes not, to optimize test power under alternatives close to the null we could estimate the shift sizes by pseudo-GLS under local alternatives, as suggested by Elliott et al. (1996). Furthermore, to keep under control the finite-sample size of the resulting tests, following the suggestion of Saikkonen and Lütkepohl (2002) and Lanne et al. (2002), we would set the localizing parameter to zero. This reduces to estimation of shift sizes under the UR null, i.e., to a regression of ΔX_t on impulse dummy variables, one per shift. The implied estimator of μ_t would be

$$\hat{\mu}_t := \sum_{s=1}^{t} \delta_s \Delta X_s = \sum_{s=1}^{t} \delta_s \theta_s + \sum_{s=1}^{t} \delta_s \Delta Y_s = \mu_t + \sum_{s=1}^{t} \delta_s \Delta Y_s .$$

It is not difficult to show that, under the assumptions discussed in section 2, if δ_t and ε_t are independent, the estimation error $\hat{\mu}_t - \mu_t = \sum_{s=1}^{t} \delta_s \Delta Y_s$ is uniformly bounded in probability, while μ_t (as well as Y_t under H_0 and H_c) has stochastic magnitude order $T^{1/2}$. This difference turns out to be sufficient for the ADF statistics based on the 'de-jumped' time series $\hat{X}_t^{\delta} := X_t - \hat{\mu}_t = Y_t + (\mu_t - \hat{\mu}_t)$, $t=1,...,T$, ($\hat{X}_t^{\delta} := X_t = Y_t$ for $t=-k,...,0$) to have the same null and local-to-null asymptotic distributions as in the standard case of no level shifts, and to diverge under fixed alternatives.

In the case of unobservable δ_t we can imitate the above de-jumping scheme through the following procedure:

1. the sequence of level shift indicators $\{\delta_t\}$ is estimated by an appropriate estimator $\{\tilde{\delta}_t\}$; the level shift component is estimated by $\tilde{\mu}_t^{\delta} := \sum_{s=1}^{t} \tilde{\delta}_s \Delta X_s$;

2. the level shifts are removed by constructing the de-jumped series $\tilde{X}_t^\delta := X_t - \tilde{\mu}_t^\delta$ $(t \geq 1)$ and $\tilde{X}_t^\delta := X_t$ $(t = -k,...,0)$;
3. standard ADF tests, say $ADF_{\hat{\alpha}}^\delta$ and ADF_t^δ, are computed using $\tilde{X}_t^\delta$ instead of the original time series; the UR null hypothesis is evaluated by comparing $ADF_{\hat{\alpha}}^\delta$ and ADF_t^δ to standard asymptotic critical values (see, e.g., Fuller, 1976).

In this framework, the key step is the estimation of level shift indicators. Technically, since $\tilde{X}_t^\delta = \hat{X}_t^\delta - \sum_{s=1}^{t} (\tilde{\delta}_s - \delta_s)\theta_s$, for the ADF test obtained from 1-3 above not to be influenced asymptotically by the fact that δ_t are estimated, $\sum_{t=1}^{T} | \tilde{\delta}_t - \delta_t |$ should converge to zero sufficiently fast.

In the next subsection we introduce a traditional estimator of the level shift dates which can be employed in conjunction with the de-jumping method. The estimator, which is due to Chen and Tiao (1990) and Chen and Liu (1993), is well-known in the outlier detection literature.

4. ESTIMATION OF THE LEVEL SHIFT INDICATORS

Let us start by considering the reduced form of equations (2)-(3):

$$\Pi(L)Y_t = \varepsilon_t,$$

where $\Pi(L) := (1 - \alpha L)\Gamma(L)$. The level shift detection procedure of Chen and Tiao (1990) and Chen and Liu (1993) is based on the observation that, under the additive level shift model, the counterpart of the above equation in terms of the observable variable X_t is

$$\Pi(L)X_t = \varepsilon_t + \Pi(L)\mu_t = \varepsilon_t + \sum_{s=1}^{T} \Pi(L)\mathbb{I}\{t \geq s\}\delta_s\theta_s,$$

where $\delta_{-s} := 0$ for $s = 0,...,k$. Thus, in the case of known $\Pi(L)$, to test the hypothesis that a level shift has occurred at a given time t_* against the alternative of no level shifts in the sample, one could regress $\Pi(L)X_t$ on $\Pi(L)\mathbb{I}\{t \geq t_*\}$ $(t = 1,...,T)$, and could base the test on the t-statistic

$$\hat{\lambda}_{t_*} := \frac{\hat{\omega}_{t_*}}{\hat{\sigma}_{t_*}} \left(\sum_{t=t_*}^{T} [\Pi(L)\mathbb{I}\{t \geq t_*\}]^2 \right)^{1/2},$$

where $\hat{\omega}_{t_*}$ is the OLS estimator of the regression coefficient, and $\hat{\sigma}_{t_*}$ is the residual standard deviation. On the other hand, a test of the hypothesis that there is at least one level shift in the sample, against the alternative of no shifts, could be based on the statistic

$$\Lambda := \max_{t=1,\dots,T} \left|\hat{\lambda}_t\right|.$$

In the practically relevant case where $\Pi(L)$ is unknown, Chen and Tiao (1990) and Chen and Liu (1993) suggest to replace it by the estimator $\tilde{\Pi}(L)$ obtained by regressing X_t on X_{t-1} and $(\Delta X_{t-1},...,\Delta X_{t-k})$; in this case we write $\tilde{\omega}$, $\tilde{\lambda}$ and $\tilde{\Lambda}$ for the associated statistics. An estimator of the shift indicators $\{\delta_t\}$ can then be defined as follows (cf. Chen and Tiao, 1990).

1. Let $\tilde{x}_t := \tilde{\Pi}(L)X_t$ and let $\tilde{D} := \varnothing$ denote the initial estimator of the set of dates where level shifts have occurred.
2. Compute $\tilde{\lambda}_t$ ($t = 1,...,T$) and $\tilde{\Lambda} := \max_{t=1,...,T} |\tilde{\lambda}_t|$. If $|\tilde{\Lambda}| < C$, where C is a critical value, then go to step 4.
3. Let t^* denote an observation such that $\tilde{\lambda}_{t_*} = \tilde{\Lambda}$. Replace $\tilde{D}$ by $\tilde{D} \cup \{t^*\}$ (updated set of dates with level shifts) and $\tilde{x}_t$ by $\tilde{x}_t - \tilde{\omega}_{t_*}\tilde{\Pi}(L)\mathbb{I}\{t \geq t^*\}$ (residuals corrected for the newly detected shift). Return to step 2 with the updated $\tilde{x}_t$.
4. Define $\tilde{\delta}_t := \mathbb{I}\{t \in \tilde{D}\}$ ($t = 1,...,T$).

The choice of the critical value C in the above procedure is obviously crucial. Since $\tilde{\Lambda}$ is an extremum statistic, it may converge to a distribution only upon appropriate normalization. Even so, due to difficulties related to the dependence structure of $\{\tilde{\lambda}_t\}$, no asymptotic critical values have been proposed in the literature, as far as we are aware. Instead, based on simulations with various DGPs, Chen and Tiao (1990) propose the value of $C = 2.8$ for what they call "high sensitivity level-shift detection", and we adopt their suggestion for the Monte Carlo simulations in the next sections.

Given $\{\tilde{\delta}_t\}$, we can obtain a de-jumped series $\tilde{X}_t := X_t - \sum_{s=1}^{t}\tilde{\delta}_s\Delta X_s$ as discussed in the previous section. In the simulations exercise we talk about Chen-Tiao de-jumping in this case, although Chen and Tiao do not actually propose de-jumping (they employ dummy variables) and do not address UR testing (by ruling out unit roots).

An alternative approach to de-jumping, discussed in Chen and Liu (1993), is to regress $\tilde{\Pi}(L)X_t$ on the vector $\tilde{\Pi}(L)(\mathbb{I}\{t \geq t_1\},...,\mathbb{I}\{t \geq t_{|\tilde{D}|}\})'$, $t = 1,...,T$, where $\{t_1,\dots,t_{|\tilde{D}|}\} = \tilde{D}$. If the estimated coefficient vector is $(\tilde{\omega}_1,\dots,\tilde{\omega}_{|\tilde{D}|})'$, then $\tilde{\omega}_i$ estimates the size of the level shift occurring at time t_i. A de-jumped series can be defined as $\tilde{X}_t := X_t - \sum_{s=1}^{t} \tilde{\delta}_s \tilde{\omega}_{\tilde{n}(s)}$, with $\tilde{n}(s) := \sum_{u=1}^{s} \tilde{\delta}_u$, and we talk about Chen-Liu de-jumping in this case.

It is seen that a main difference between the relatively simple Chen-Tiao de-jumping (the way we have defined it), and Chen-Liu de-jumping (the way it is implicitly defined in Chen and Liu, 1993, p. 287, stage II.3, and formalized above) is that there is a pseudo GLS (efficiency improving) idea behind the former, and an OLS idea behind the latter. Furthermore, for Chen-Tiao de-jumping the local GLS parameter, as explained in the previous section, is set to zero for size considerations. We evaluate next the finite sample implications of these differences.

5. FINITE SAMPLE SIMULATIONS

In this section the finite-sample size and power properties of standard ADF tests and of ADF tests based on prior level-shifts removal, as introduced in sections 3 and 4, are investigated by Monte Carlo simulation, for DGPs either with or without level shifts. We want to establish (i) whether allowing for multiple level shifts does not result in deteriorated size properties, and (ii) whether the power properties of the tests based on prior level-shifts removal are close to those of the usual ADF tests under standard conditions (i.e., without level shifts), at least as the sample size increases, under fixed and local alternatives.

The employed DGPs are as follows. Data are generated for sample sizes of $T =$ 100,200,400 observations according to model (1)-(3) with $k = 1$, $\gamma := \gamma_1 \in \{-0.5, 0, 0.5\}$, $\varphi = 0$ and zero-mean unit-variance IID innovations following a Gaussian distribution; the initial values are $Y_{-1} = 0$ and u_{-1} drawn from the stationary distribution implied by eq. (3), i.e., $u_{-1} \sim N(0,(1-\gamma^2)^{-1})$. We consider the UR case, which obtains by setting $\alpha = 1$ in (2), the sequence of local alternatives $\mathsf{H}_c : \alpha = 1-c/\mathrm{T}$ with $c := 7$, and the fixed alternative $\mathsf{H}_f : \alpha = 0.9$.

Three specifications of the level shift component are employed. First, the case of no level shifts ($\mu_t = 0$ for all t) is considered, with the resulting model denoted by S_0. Second, with S_4 we denote the case of four shifts occurring at fixed sample fractions t_i, i=1,...,4, with $t_1 := \lfloor 0.2T \rfloor$, $t_2 := \lfloor 0.35T \rfloor$, $t_3 := \lfloor 0.6T \rfloor$ and $t_4 := \lfloor 0.8T \rfloor$, and with size magnitudes $\eta_{t_1} = \eta_{t_4} := 0.4$ and $\eta_{t_2} = -\eta_{t_3} := 0.35$; consequently, the level shift component is

$$\mu_t := T^{1/2} [0.4\, \mathsf{I}(t \geq \lfloor 0.2T \rfloor) + 0.35\, \mathsf{I}(t \geq \lfloor 0.35T \rfloor) - 0.35\, \mathsf{I}(t \geq \lfloor 0.6T \rfloor) + 0.4\, \mathsf{I}(t \geq \lfloor 0.8T \rfloor)]$$

Third, we consider a case, S_r in the following, with a random number of level shifts $N_T \sim \mathrm{B}(T,2/T)$ (B denoting a Binomial distribution); i.e., at least two, and on average four, level shifts occur over the sample. The shift dates t_i, $i = 1,...,N_T$, are generated as $t_i := \lfloor \tau_i T \rfloor$, where the relative locations τ_i are independent and uniformly distributed on [0,1]; the (independent) shift sizes η_i are drawn from a uniform distribution on [−4,−3.5]∪[3.5,4]. Note that for our selection of T, this model generates level shifts of size between 4 and 8 standard deviations of the errors. These shift magnitudes, although large, are not unrealistic. For instance, Vogelsang and Perron (1998, p.1090) report, for the long GNP series considered in Perron (1992), level shifts "generally no larger (in absolute value) than 8 (...) relative to the standard deviation of the innovation errors". Furthermore, in some recent papers (see Papell *et al.*, 2000, *inter alia*), UR tests under level shifts are applied to unemployment series where level shifts are as large as 10 times the standard deviation of the innovations. Using, as in Perron (1990), quarterly US unemployment data from 1948 to 1988 with a break at 1974:1, we have estimated the size of the shift as 6.9 times the residual standard deviation from a 9 lags ADF regression on the de-jumped data (cf. Perron, 1990, Table 1). With T=163 this corresponds to η =0.54, which is actually larger than the values we have used in our Monte Carlo design. Similarly, using monthly unemployment over the same range (T=485) we find that the size of the shift is 10.4 times the residual standard deviation from the second-stage ADF regression; this corresponds to η =0.47.

TABLE 1

Comparison between standard Dickey-Fuller tests and the new unit root tests based on traditional level shift estimation. Model S_0 (no level shifts)

				Size ($\alpha = 1$)			
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	5.1	5.1	8.3	8.5	9.2	9.5
	0.0	5.2	5.1	8.2	8.3	8.6	8.7
	0.5	5.5	5.4	7.0	6.9	7.2	7.1
200	-0.5	5.3	5.4	10.0	10.2	10.9	11.2
	0.0	4.6	4.6	8.7	8.8	9.2	9.2
	0.5	5.0	4.8	6.4	6.5	6.7	6.7
400	-0.5	4.9	5.1	10.5	10.6	11.5	11.8
	0.0	4.8	4.8	9.5	9.5	9.9	10.1
	0.5	5.2	5.2	7.0	6.9	7.2	7.2
				Size-adjusted power ($\alpha = 1\text{-}c/T$)			
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	51.0	51.5	37.4	37.7	38.0	38.5
	0.0	48.0	49.3	38.9	39.5	38.8	39.9
	0.5	42.4	43.2	40.8	41.8	40.7	41.6
200	-0.5	47.6	48.4	36.3	35.3	36.5	36.1
	0.0	51.6	52.0	40.4	41.7	40.4	41.4
	0.5	46.2	47.5	43.9	44.7	43.5	45.1
400	-0.5	48.9	48.5	36.9	37.1	37.8	38.6
	0.0	50.4	50.6	39.6	39.7	40.1	40.3
	0.5	47.6	47.2	43.5	44.1	43.4	44.3
				Size-adjusted power ($\alpha = 0.9$)			
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	76.2	76.7	60.1	60.2	60.4	60.9
	0.0	72.7	73.4	59.4	60.9	59.9	60.9
	0.5	63.9	64.5	60.2	61.7	60.2	61.5
200	-0.5	99.8	99.5	94.4	94.2	97.1	97.2
	0.0	99.7	99.6	96.8	96.9	97.6	97.7
	0.5	98.7	98.6	97.1	97.1	97.5	97.5
400	-0.5	100.0	100.0	99.2	99.1	100.0	100.0
	0.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.5	100.0	100.0	100.0	100.0	100.0	100.0

We consider tests based both on Chen-Tiao and Chen-Liu de-jumping, see section 4. All tests are performed at the 5% (asymptotic) nominal level, with critical values taken from Fuller (1976, Tables 10.A.1 and 10.A.2). Computations are based on 10 000 Monte Carlo replications and are carried out in Ox v. 3.40, Doornik (2001).

In Table 1 results for the DGPs with no level shifts are reported. The three panels of the table refer to the UR case, to the case of local alternatives and to the case of a fixed alternative. Standard ADF tests are reported together with tests based on de-jumping. Three facts emerge from the table.

First, standard ADF tests have quite accurate finite-sample size, whereas tests based on de-jumping tend to be slightly oversized for all the sample sizes considered. There are no big differences between tests based on the Chen-Tiao procedure and tests based on the Chen-Liu procedure, although the latter are slightly less accurate.

Second, in terms of size-adjusted power against local alternatives, allowing for possible multiple level shifts when actually none is present implies a power loss with respect to standard ADF tests. This power loss does not seem to decrease when the sample size increases. Again, there are no strong differences between tests based on Chen-Tiao and Chen-Liu de-jumping.

Finally, all tests seem to be consistent against the fixed alternative $\alpha = 0.9$.

Results for model S_4 are reported in Table 2. The following results are worth to note. First, standard ADF tests tend to be slightly undersized, in particular for cases where γ is non zero. This feature does not characterize tests based on de-jumping: when the Chen-Tiao method is employed, the tests are slightly oversized for T=100 and, in particular, when $\gamma = 0$. The tests based on Chen and Liu's algorithm experience more severe oversizing problems. Again, these are more serious when $\gamma = 0$ and tend to disappear as T grows. However, even for T=400 the empirical size of Chen-Liu based tests can be up to 10%.

Second, under local alternatives the (size-adjusted) power of standard ADF tests is very low. For T=100, power is about 0 for γ=–0.5, about 0.2% for γ=0 and about 10% for γ=0.5. Conversely, the tests based on de-jumping have much better power properties. For T=100, when the Chen-Tiao procedure is employed power is about 11% for γ=–0.5, 15% for γ=0 and 30% for γ=0.5; for T=200 power grows to 24% for $\gamma = -0.5$, around 29% for γ=0 and 38% for $\gamma = 0.5$. For T=400 the rejection rate increases further, although it tends to stay below the 50% asymptotic power which characterizes standard ADF tests when no level shifts are present. When, instead, Chen-Liu's procedure is employed, local power slightly decreases.

Third, under non-local alternatives, standard ADF tests have (size-adjusted) power close to 0, even for large sample sizes. Conversely, the tests based on de-jumping have power which significantly increases with the sample size.

As in the case of local alternatives, (i) the Chen-Tiao procedure is preferable over the Chen-Liu method, and (ii) large sample sizes are needed for the tests to achieve a power level close to that of standard ADF tests under no level shifts (see Table 1).

TABLE 2

Comparison between standard Dickey-Fuller tests and the new unit root tests based on traditional level shift estimation. Model S_4 (four level shifts at fixed fractions of the sample)

				Size ($\alpha = 1$)			
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	3.0	3.2	7.4	7.5	10.9	11.1
	0.0	4.9	4.9	9.3	9.2	12.1	12.0
	0.5	3.6	3.7	7.1	6.7	9.3	9.3
200	-0.5	2.6	2.7	5.6	5.7	7.5	7.4
	0.0	4.8	4.9	6.4	6.4	10.2	10.1
	0.5	3.4	3.4	6.5	6.3	8.3	8.2
400	-0.5	2.5	2.7	5.2	5.1	6.1	6.1
	0.0	4.6	4.7	6.5	6.5	9.4	9.3
	0.5	3.5	3.5	6.8	6.8	8.2	8.4
				Size-adjusted power ($\alpha = 1\text{-}c/T$)			
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	0.0	0.0	11.4	11.2	10.1	10.0
	0.0	0.1	0.2	14.8	14.9	11.5	11.4
	0.5	11.0	10.9	29.9	30.6	27.3	28.5
200	-0.5	0.0	0.0	24.1	23.9	22.8	23.5
	0.0	0.2	0.2	28.8	28.4	19.4	20.4
	0.5	11.7	11.2	38.3	38.7	35.1	36.0
400	-0.5	0.0	0.0	27.3	27.6	27.7	28.7
	0.0	0.1	0.1	29.5	29.5	21.7	22.2
	0.5	10.6	10.4	38.8	38.8	36.7	37.1
				Size-adjusted power ($\alpha = 0.9$)			
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	0.0	0.0	14.2	14.1	14.1	14.1
	0.0	0.0	0.0	19.8	20.1	16.0	16.0
	0.5	7.5	7.8	40.2	41.0	36.9	36.9
200	-0.5	0.0	0.0	50.0	49.5	53.9	53.9
	0.0	0.0	0.0	60.9	60.4	52.4	52.4
	0.5	3.5	3.4	85.5	85.4	85.0	85.0
400	-0.5	0.0	0.0	63.7	63.3	70.4	70.4
	0.0	0.0	0.0	74.8	73.7	70.3	70.3
	0.5	0.7	0.6	97.0	96.8	98.3	98.3

Results for the random shift model S_r do not qualitatively differ from those for the S_4 model; in general, the power loss experienced by standard ADF tests is less severe than for S_4, and the local power of the tests based on de-jumping is closer to the asymptotic power envelope. Overall, the power results are encouraging, as they show that it is possible to distinguish between UR processes and processes which are stationary apart from several level shifts, even in samples of moderate dimension.

6. EXTENSIONS

In this section we briefly show how the tests discussed can be applied in the case of deterministic time trends and in the case of unknown autoregressive order.

TABLE 3

Comparison between standard Dickey-Fuller tests and the new unit root tests based on traditional level shift estimation. Model S_r (random level shifts, four on average, at least two)

Size ($\alpha = 1$)							
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	3.6	3.5	8.0	8.2	10.9	11.2
	0.0	4.9	4.8	9.4	9.4	11.9	12.0
	0.5	3.5	3.5	6.9	6.9	8.8	8.8
200	-0.5	3.9	3.9	6.1	6.3	9.2	9.2
	0.0	4.8	4.7	7.2	7.3	10.1	10.2
	0.5	3.7	3.6	6.3	6.3	7.9	7.8
400	-0.5	3.5	3.6	6.5	6.5	8.5	8.5
	0.0	4.8	4.7	7.5	7.6	9.7	9.8
	0.5	3.5	3.5	6.4	6.2	7.2	7.2

Size-adjusted power ($\alpha = 1$-c/T)							
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	0.0	10.2	15.9	15.7	18.9	18.6
	0.0	0.1	13.5	19.9	19.9	20.2	20.5
	0.5	11.0	23.4	31.9	32.1	32.0	32.7
200	-0.5	0.0	9.2	25.1	24.8	25.5	24.8
	0.0	0.2	14.1	31.2	30.9	27.4	27.7
	0.5	11.7	24.5	39.8	40.6	38.2	39.2
400	-0.5	0.0	9.5	26.3	26.6	27.4	27.5
	0.0	0.1	13.2	31.8	32.0	28.5	29.3
	0.5	10.6	24.7	40.9	41.5	40.0	41.1

Size-adjusted power ($\alpha = 0.9$)							
T	γ	ADF_α	ADF_t	ADF_α^{CT}	ADF_t^{CT}	ADF_α^{CL}	ADF_t^{CL}
100	-0.5	11.4	11.4	21.9	21.5	27.1	26.6
	0.0	15.7	15.7	25.9	26.1	27.3	27.7
	0.5	30.0	29.2	45.0	44.9	45.0	45.9
200	-0.5	12.0	12.0	53.4	52.5	64.5	63.6
	0.0	20.7	20.8	66.3	65.4	66.8	66.8
	0.5	43.8	42.6	87.9	87.8	89.8	89.8
400	-0.5	13.7	13.3	66.0	65.2	77.1	76.4
	0.0	23.8	23.7	78.6	78.0	81.6	81.2
	0.5	51.2	50.2	97.0	96.8	98.5	98.6

6.1 *Deterministic time trends*

So far we have considered the case of no deterministics in the DGP; i.e., φ_t of (1) is equal to 0, all t. Nevertheless, in the presence of deterministic time trends the approach discussed in sections 3-4 can be applied in conjunction with a proper detrending procedure.

In what follows only the cases of a constant ($Z_t := 1$) and of a linear time trend ($Z_t := (1,t)'$) are considered; the procedure may be extended to general deterministic trends as those considered, e.g., in Phillips and Xiao (1998).

A simple approach is to combine the de-jumping algorithm of section 3 with pseudo-GLS detrending (see Elliott *et al.*, 1996). For a time series X_t, the pseudo GLS detrended series at $\bar{\alpha} := 1 - \bar{c}/T$ is defined as $\tilde{X}_t^{\bar{\alpha}} := X_t^{\bar{\alpha}} - \hat{\varphi}^{\bar{\alpha}\prime} Z_t^{\bar{\alpha}}$, where $(X_0^{\bar{\alpha}}, X_t^{\bar{\alpha}}) := (X_0, (1-\bar{\alpha}L)X_t)$, $(Z_0^{\bar{\alpha}}, Z_t^{\bar{\alpha}}) := (Z_0, (1-\bar{\alpha}L)Z_t)$ and $\hat{\varphi}^{\bar{\alpha}}$ minimizes $S(\varphi) = \sum_t (X_t^{\bar{\alpha}} - \varphi' Z_t^{\bar{\alpha}})^2$.

In details, the de-jumping procedure can be combined with deterministic corrections in the following way:

1. level shifts are initially detected as suggested in section 4, but including the deterministic regressors Z_t in the regression for the estimation of $\Pi(L)$; δ_t are estimated using the residuals of this regression in place of $\tilde{\Pi}(L)X_t$;
2. level shifts are removed by de-jumping, i.e. by computing $\tilde{X}_t^{\delta}$ as explained in section 3;
3. pseudo-GLS de-trending with respect to Z_t is applied to $\tilde{X}_t^{\delta}$;
4. UR test statistics are obtained from an ADF regression for the de-trended version of $\tilde{X}_t^{\delta}$.

Obviously, the critical values which have to be used need to account for the fact that the data employed to compute the UR test statistics are de-trended. In the constant case $Z_t := 1$, critical values can still be taken from Tables 10.A.1 and 10.A.2 in Fuller (1976). In the linear trend case, $Z_t := (1,t)'$, when (as is standard) GLS detrending is conducted at $\bar{\alpha} := 1 - \bar{c}/T$ with $\bar{c} = 13.5$, critical values are reported in Ng and Perron (2001), Table I.

6.2 *Unknown autoregressive order*

So far, we have considered the case where the autoregressive order k of the errors u_t, see eq. (3), is known. In more realistic situations, the autoregressive order is unknown and must be estimated prior to the computation of the UR test statistics.

To deal with unknown lag order we may consider the following simple strategy, which combines the test procedure outlined earlier with standard econometric methods for determination of the lag order. More specifically, in the first round, level shifts can be detected as in section 4, with the lag order set to some maximum admissible number $k_{\max}$; for instance, as suggested in Ng and Perron (1995), one may use the following function of the sample size: $k_{\max} := \text{int}\{12(T/100)^{1/4}\}$. Once level shift detection has been performed, a de-jumped time series is obtained.

In the second round a standard criterion for the determination of the lag order is applied to the de-jumped time series; this delivers an estimate of k, say k^*. In the final round, the ADF statistics are computed on the de-jumped time series fixing the number of lags k at k^*.

7. CONCLUSIONS

In this note we have proposed a modification of the well-known augmented Dickey-Fuller (ADF) tests which allows to test for unit roots in the presence of multiple level shifts. Differently from existing works, we do not restrict the number of level shifts – which may occur at random dates and may have random sizes – apart from requiring it to be bounded in probability. The tests are based on a two-step procedure where possible level shifts are initially detected using the level shift indicator estimators suggested by Chen and Tiao (1990) and Chen and Liu (1993), and later removed by a novel procedure which is denoted as "de-jumping". A Monte Carlo simulation has shown that the new tests, used in con-juntion with standard critical values, behave well in finite samples.

Some remarks are worth noting. Althought the finite-sample properties seem largely acceptable when compared to those of standard ADF tests, there is still room for size improvements. The tests tend to be oversized for small values of T, and better size properties might be obtained by replacing Chen and Tiao's (1990) and Chen and Liu's (1993) level shift detection procedures with new procedures which explicitly account for the possible non-stationarity of the data. Moreover, the power of the modified tests seems not to be optimal even asymptotically: althought having much larger power than standard ADF tests, the local power of the modified tests does not seem to converge to the asymptotic power envelope (that is, approximately 50% power when the localizing parameter c equals -7) which obtains when no level shifts are present. This result is not related to the de-jumping method (which, in case the level shift dates are known, delivers 50% asymptotic local power). Rather, it can be attributed to the Chen-Tiao and Chen-Liu level shift detection procedure. Again, this result suggests that further research should be carried out in order to develop new methods for the estimation of the level shift indicators. The new procedures should probably be worked out under the null hypothesis, given that in our Monte Carlo experiment tests based on Chen-Tiao de-jumping (which implicitly imposes a UR) have better properties than tests based on Chen-Liu de-jumping (which are carried out without imposing the UR).

Dipartimento di Scienze Statistiche GIUSEPPE CAVALIERE
Alma Mater Studiorum - Università di Bologna

Faculdade de Economia ILIYAN GEORGIEV
Universidade Nova de Lisboa

REFERENCES

C. AMSLER, J. LEE (1995), *An LM test for a unit root in the presence of a structural change*, "Econometric Theory", 11, 359-368.

N.S. BALKE, T.B. FOMBY (1991a), *Shifting trends, segmented trends, and infrequent permanent shocks*, "Journal of Monetary Economics", 28, 61-85.

N.S. BALKE, T.B. FOMBY (1991b), *Infrequent permanent shocks and the finite-sample performance of unit root tests*, "Economics Letters", 36, 269-73.

A. BANERJEE, R.L. LUMSDAINE, J.H. STOCK (1992), *Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence*, "Journal of Business and Economic Statistics", 10, 271-287.

G. CAVALIERE, I. GEORGIEV (2006), *Testing for unit roots in autoregressions with multiple level shifts*, Quaderno del Dipartimento di Scienze Statistiche, Università di Bologna, Serie ricerche 2006 n. 2, Forthcoming in "Econometric Theory".

Y. CHANG, J.Y. PARK (2002), *On the asymptotics of ADF tests for unit roots*, "Econometric Reviews", 21, 431-447.

C. CHEN, G.C. TIAO (1990), *Random level shift time series models, ARIMA approximations and level shift detection*, "Journal of Business and Economic Statistics", 8, 170-186.

C. CHEN, L.-M. LIU (1993), *Joint estimation of model parameters and outlier effects in time series*, "Journal of the American Statistical Association", 88, 284-297.

J. CLEMENTE, A. MONTAÑÉS, M. REYES (1998), *Testing for a unit root in variables with a double change in the mean*. "Economics Letters", 59, 175-182.

J. DOORNIK (2001), *Ox: an object-oriented matrix programming language*, Timberlake Consultants, London.

G. ELLIOTT, T.J. ROTHEMBERG, J.H. STOCK (1996), *Efficient tests for an autoregressive unit root*, "Econometrica", 64, 813-836.

P.H. FRANSES, N. HALDRUP (1994), *The effects of additive outliers on tests for unit roots and cointegration*, "Journal of Business and Economic Statistics", 12, 471-78.

W. FULLER (1976), *Introduction to Statistical Time Series*, Wiley, New York.

M. LANNE, H. LÜTKEPOHL, P. SAIKKONEN (2002), *Comparison of unit root tests for time series with level shifts*, "Journal of Time Series Analysis", 23, 667-685.

S. LEYBOURNE, P. NEWBOLD (2000a), *Behaviour of the standard and symmetric Dickey-Fuller type tests when there is a break under the null hypothesis*, "Econometrics Journal", 3, 1-15.

S. LEYBOURNE, P. NEWBOLD (2000b), *Behavior of Dickey-Fuller t-tests when there is a break under the alternative hypothesis*, "Econometric Theory", 16, 779-789

R.L. LUMSDAINE, D.H. PAPELL (1997), *Multiple trend breaks and the unit root hypothesis*, "Review of Economics and Statistics", 79, 212-218.

H. LÜTKEPOHL, P. SAIKKONEN, C. TRENKLER (2004), *Testing for the cointegration rank of a VAR process with level shift at unknown time*, "Econometrica", 72, 647-662.

U.K. MÜLLER, G. ELLIOTT (2003), *Tests for unit roots and the initial condition*, "Econometrica", 71, 1269-1286.

S. NG, P. PERRON (1995),*Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag*, "Journal of the Americal Statistical Association", 90, 268-281.

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

D.H. PAPELL, C.J. MURRAY, H. GHIBLAWI (2000), *The structure of unemployment*, "Review of Economics and Statistics", 82, 309-315.

P. PERRON (1989), *The great crash, the oil price shock, and the unit root hypothesis*, "Econometrica", 57, 1361-401.

P. PERRON (1990), *Testing for a unit root in a time series with a changing mean*, "Journal of Business and Economic Statistics", 8, pp. 153-162.

P. PERRON (1992), *Trend, unit root and structural change: a multi-country study with historical data*, "Proceedings of the Business and Economic Statistics Section", Americal Statistical Association, 144-149.

P. PERRON, T.J. VOGELSANG (1992), *Nonstationarity and level shifts with an application to purchasing power parity*, "Journal of Business and Economic Statistics", 10, 301-320.

P.C.B. PHILLIPS, Z. XIAO (1998), *A primer on unit root testing*, "Journal of Economic Surveys", 12, pp. 423-470.

P. SAIKKONEN, H. LÜTKEPOHL (2001), *Testing for unit roots in time series with level shifts*, "Allgemeines Statistisches Archiv", 85, 1-25.

P. SAIKKONEN, H. LÜTKEPOHL (2002), *Testing for a unit root in a time series with a level shift at unknown time*, "Econometric Theory", 18, 313-348.

T.J. VOGELSANG, P. PERRON (1998), *Additional tests for a unit root allowing for a break in the trend function at an unknown time*, "International Economic Review", 39, 1073-1100.

E. ZIVOT, D. ANDREWS (1992), *Further evidence on the Great crash, the oil price shock and the unit root hypothesis*, "Journal of Business and Economic Statistics", 10, 251-270.

RIASSUNTO

Una nota sui test di radice unitaria in presenza di cambiamenti di livello

Nel lavoro vengono discusse le proprietà dei test Dickey-Fuller "aumentati" [ADF] per processi autoregressivi con una radice unitaria (o vicina all'unità) in presenza di cambiamenti di livello. Si mostra come la presenza dei cambiamenti di livello causa un significativo e sostanziale decremento nella potenza dei test. Viene discussa una nuova classe di test ADF "modificati" che possono essere applicati anche in assenza di informazioni sul numero (e sulla posizione) dei cambiamenti di livello. I test sono basati su una procedura a due stadi in cui, al primo passo, i cambiamenti di livello vengono individuati usando gli stimatori proposti da Chen e Tiao (1990, *Journal of Business and Economics Statistics*) e da Chen e Liu (1993, *Journal of the American Statistical Association*); successivamente, al secondo passo, i cambiamenti di livello vengono eliminati sulla base di una nuova procedura che, nel lavoro, viene denominata "de-jumping". Una simulazione Monte Carlo consente di notare come i test modificati, sebbene caratterizzati da un'ampiezza superiore al livello di significatività nominale in campioni di piccola dimensione, abbiano una potenza estremamente superiore a quella dei test ADF tradizionali.

SUMMARY

A note on unit root testing in the presence of level shifts

In this note we discuss the properties of Augmented-Dickey-Fuller [ADF] unit root tests for autoregressive processes with a unit or near-unit root in the presence of multiple level shifts of large size. Due to the presence of level shifts, the ADF tests experience severe power losses. We consider new modified ADF unit root tests which require no knowledge of either the location or the number of level shifts. The tests are based on a two-step procedure where possible level shifts are initially detected using the level shift indicator estimators suggested by Chen and Tiao (1990, *Journal of business and Economics Statistics*) and Chen and Liu (1993, *Journal of the American Statistical Association*), and later removed by a novel procedure which is denoted as "de-jumping". Using a Monte Carlo experiment we show that the new tests, althought partially oversized in samples of moderate size, have much higher power than standard ADF tests.