BE990-8-SP - Research Methods in Financial Econometrics

Module Lecturer (Topics 1-3)

Professor Robert Taylor
Room EBS.3.12, Essex Business School
University of Essex

e-mail: robert.taylor@essex.ac.uk

Spring Term 2023

Topic 1: Recent Developments in Unit Root Testing

- Models with an autoregressive [AR] unit root have very different properties than stationary AR models. It is therefore important to test whether a unit root is present in a given time series.
- We will review the most popular class of unit root tests, the so-called *Dickey-Fuller unit root tests*.
- We will then look at extensions of the these tests, including the semi-parametric unit root tests of Phillips-Perron and the M unit root tests.
- We then consider *efficient* tests of the unit root hypothesis.
- We will also discuss important practical issues relating to trend function determination and the initial condition of the process.
- Finally, we will discuss how to perform bootstrap unit root tests (which will be followed up on in the lab session with Dr Sam Astill later today).

1.1 Dickey-Fuller Unit Root Tests

- Assume that we have $T+1$ observations on a time series y_t generated by the following DGP

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad t = 1, ..., T \tag{1}$$

where the driving shocks $\varepsilon_t \sim IID(0, \sigma^2)$, and where we assume for now that the (observed) *initial value* y_0 is a random variable with finite variance.

- We can re-write the series as

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t, \qquad t = 1, ..., T \tag{2}$$

where $\Delta y_t := y_t - y_{t-1}$ and $\gamma := \phi - 1$.

- Dickey and Fuller (1979, JASA; 1981, Econometrica) [DF] consider tests of the following hypothesis structure

$$
\begin{aligned}
& H_0 : \phi = 1 \text{ (or equally } \gamma = 0) \Leftrightarrow y_t \sim I(1) \\
vs \quad & H_1 : |\phi| < 1 \text{ (or equally } -2 < \gamma < 0) \Leftrightarrow y_t \sim I(0).
\end{aligned}
$$

- So the null hypothesis is that y_t is an integrated (unit root) process; $y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$. Under the alternative the process is stationary [$I(0)$].

DF recommend using OLS to estimate (2), regressing Δy_t on y_{t-1} for $t = 1, ..., T$. They then propose testing H_0 versus H_1 using a one-sided regression t-statistic for the significance of the regressor y_{t-1} in (2)

$$t_{DF} := \hat{\gamma}/s.e.(\hat{\gamma})$$

where

$$\hat{\gamma} := \sum_{t=1}^{T} \Delta y_t . y_{t-1} / \sum_{1}^{T} y_{t-1}^2, \qquad s.e.(\hat{\gamma}) := \left(\hat{\sigma}^2 / \sum_{1}^{T} y_{t-1}^2 \right)^{1/2}$$

and where

$$\hat{\sigma}^2 := \sum_{1}^{T} \left(\Delta y_t - \hat{\gamma} y_{t-1} \right)^2 / (T-1).$$

Notice that the formulae given on the last page can be obtained using the standard matrix formulae

$$\hat{\gamma} = (X'X)^{-1}X'Y$$

and

$$\text{s.e.}(\hat{\gamma}) = (\hat{\sigma}^2(X'X)^{-1})^{1/2}$$

with

$$\hat{\sigma}^2 = (Y - X\hat{\gamma})'(Y - X\hat{\gamma})/(T-1)$$

where $X = (y_0, y_1, ..., y_{T-1})\prime$ and $Y = (\Delta y_1, \Delta y_2, ..., \Delta y_T)'$.

In order to establish the limiting null distribution of t_{DF} we need to use the functional central limit theorem [FCLT] (see, for example, Hamilton's, 1994 textbook) which states that under H_0 (noting that $T^{-1/2}y_0 \xrightarrow{p} 0$)

$$(T\sigma^2)^{-1/2}y_{\lfloor Tr \rfloor} \equiv (T\sigma^2)^{-1/2}\sum_{j=1}^{\lfloor Tr \rfloor}\varepsilon_j \xrightarrow{w} B(r),\ r \in [0,1]$$

where $\lfloor \cdot \rfloor$ denotes integer part, '$\xrightarrow{w}$' denotes weak convergence as $T \to \infty$, and $B(r)$ is a *standard Brownian motion* on $[0,1]$, $B(0) = 0$.

From this basic building block it can be established, noting that $\hat{\sigma}^2$ is a consistent estimator of σ^2, that (again, see Hamilton, 1994)

$$t_{DF} \xrightarrow{w} \frac{0.5(B(1)^2 - 1)}{(\int_0^1 B(r)^2 dr)^{1/2}} \equiv \frac{\int_0^1 B(r) dB(r)}{(\int_0^1 B(r)^2 dr)^{1/2}}.$$

This limiting distribution is not the standard Normal [$N(0,1)$]. So, a crucial property of t_{DF} is that it *does not* have a $N(0,1)$ limiting distribution under the unit root null hypothesis. Consequently, comparing the outcome of this statistic with critical values from the usual regression t_{T-1} tables for the Student t-distribution will *not* deliver tests with the anticipated *size* (probability of rejection under the null hypothesis).

- The large sample critical values for t_{DF} are in fact given by

	1%	2.5%	5%	10%
t_{DF}	-2.58	-2.23	-1.95	-1.62
$N(0,1)$	-2.33	-1.96	-1.65	-1.28

- So, we can reject the unit root null hypothesis, H_0, in favour of stationarity at, for example the 5% level if $t_{DF} < -1.95$.
- Note that using the standard normal critical value of -1.65 would yield a test with size closer to 10%.

1.2 Allowing for a Non-Zero Mean

- Suppose we consider a more sophisticated model which allow for a non-zero, possibly trending, mean for the series

$$\begin{aligned} y_t &= d_t + u_t, \quad t = 0, ..., T \\ u_t &= \phi u_{t-1} + \varepsilon_t \end{aligned} \tag{3}$$

  where d_t is purely deterministic and $E(u_0^2) < \infty$.
- For example, $d_t = \alpha$ yields a series with constant mean α, while $d_t = \alpha + \beta t$ yields a series whose mean follows a linear trend.
- The mean of the series is d_t regardless of whether there is a unit root ($\phi = 1$) or not ($|\phi| < 1$).

- We may still test $H_0 : \phi = 1$ versus $H_1 : |\phi| < 1$, but we must include appropriate deterministic regression variables in the regression we estimate.
- To that end, observe that for the special case of $d_t = \alpha + \beta t$, (3) can be re-written as

$$\Delta y_t = \gamma y_{t-1} + \alpha^* + \beta^* t + \varepsilon_t, \quad t = 1, ..., T, \tag{4}$$

where

$$\begin{aligned} \gamma &:= \phi - 1 \\ \alpha^* &:= \phi(\beta - \alpha) + \alpha \\ \beta^* &:= \beta(1 - \phi). \end{aligned}$$

- To allow for a linear trend in the data, DF estimate (4) by OLS [regress Δy_t on y_{t-1}, a constant and a time trend - so the regressors at time t are $(y_{t-1},\ 1,\ t)'$] and calculate the regression t-statistic for the significance of γ in (4). Denote this statistic by $t^{\tau}_{DF} = \hat{\gamma}/\text{s.e.}(\hat{\gamma})$.
- We can obtain precise formulae for $\hat{\gamma}$ and $\text{s.e.}(\hat{\gamma})$ using the matrix formulae given previously with Y as given before, but with

$$X = \begin{bmatrix} y_0 & 1 & 1 \\ y_1 & 1 & 2 \\ \vdots & \vdots & \vdots \\ y_{T-1} & 1 & T \end{bmatrix}$$

By the Frisch-Waugh theorem, this statistic is the same as that from a regression of Δy_t on $\hat{y}_{t-1}$ where $\hat{y}_{t-1}$ is the residual from the OLS regression of y_{t-1} onto a constant and linear trend over $t = 1, ..., T$. This is also approximately equivalent to calculating the DF statistic from a regression of $\Delta \bar{y}_t$ onto $\bar{y}_{t-1}$ where $\bar{y}_t$ is the residual from a regression of y_t onto a constant and trend for $t = 0, ..., T$.

- Where only a constant is needed omit the linear trend from the above regression and now call the t-statistic t_{DF}^{μ}. This is the same as the t-statistic from the regression of Δy_t on $\tilde{y}_{t-1}$ where $\tilde{y}_{t-1}$ is the residual from the OLS regression of y_{t-1} on a constant over $t = 1, ..., T$.
- Matrix formulae can again be used, where X now contains only the first two columns of the form for X given in the linear trend case above.
- Again, this is approximately the same as calculating the DF statistic from a regression of $\Delta \breve{y}_t$ onto $\breve{y}_{t-1}$ where $\breve{y}_t$ is the residual from a regression of y_t onto a constant for $t = 0, ..., T$.

- The limiting null distribution of t_{DF}^{μ} is given by

$$t_{DF}^{\mu} \xrightarrow{w} \frac{0.5(\tilde{B}(1)^2 - \tilde{B}(0)^2 - 1)}{(\int_0^1 \tilde{B}(r)^2 dr)^{1/2}} \equiv \frac{\int_0^1 \tilde{B}(r) dB(r)}{(\int_0^1 \tilde{B}(r)^2 dr)^{1/2}}$$

  where $\tilde{B}(r) := B(r) - \int_0^1 B(s) ds$ is a *de-meaned standard Brownian motion*.
- The limiting null distribution of t_{DF}^{τ} is given by

$$t_{DF}^{\tau} \xrightarrow{w} \frac{0.5(\hat{B}(1)^2 - \hat{B}(0)^2 - 1)}{(\int_0^1 \hat{B}(r)^2 dr)^{1/2}} \equiv \frac{\int_0^1 \hat{B}(r) dB(r)}{(\int_0^1 \hat{B}(r)^2 dr)^{1/2}}$$

  where $\hat{B}(r) := \tilde{B}(r) - 12\left(r - \frac{1}{2}\right) \int_0^1 \left(s - \frac{1}{2}\right) \tilde{B}(s) ds$ is a *de-trended standard Brownian motion*.
- Again neither of these distributions are standard normal and they differ from each other and from the distribution given previously for t_{DF}.

- The critical values from the limiting null distributions of t_{DF}^{μ} and t_{DF}^{τ} are therefore different from each other and from those of t_{DF} given above:

	1%	2.5%	5%	10%
t_{DF}^{μ}	-3.43	-3.12	-2.86	-2.57
t_{DF}^{τ}	-3.96	-3.66	-3.41	-3.12

- Notice that the critical values of t_{DF} are shifted to the left (become more negative for a given significance level) relative to the standard normal, and that there is a further left shift in t_{DF}^{μ} and again in t_{DF}^{τ}. E.g. using the 1% critical value for $N(0,1)$ [-2.33] for t_{DF}^{μ} would actually give a test with size above 10%.

- Why is it important to allow for non-zero means where they occur?
- Suppose again that we have the time-series

$$y_t = \alpha + \beta t + u_t$$
$$u_t = \phi u_{t-1} + \varepsilon_t, \ \ \varepsilon_t \sim IID(0, \sigma^2)$$

  with initial condition u_0.
- Suppose then that we calculate t_{DF}. In doing so we are implicitly assuming that $\alpha = \beta = 0$. But what if this is not true?
- If $\beta \neq 0$, it can be shown that $t_{DF} \xrightarrow{w} N(0,1)$ when $\phi = 1$, so it does now have a standard normal limiting null distribution.
- However, if $|\phi| < 1$ (the process is $I(0)$) then $t_{DF} \xrightarrow{p} 0$ if $\beta \neq 0$.

- So what? Ideally, of course, we wish a test to be *consistent*, so that if the alternative is true, the probability that the test rejects the null hypothesis tends to one as the sample size diverges.
- Recall that we reject the null if $t_{DF} < 1.95$, e.g. for a 5% test. However, if $t_{DF} \xrightarrow{p} 0$, then $Pr(t_{DF} < -1.95)$ in fact converges to *zero*. Consequently, the test *never* rejects the unit root null when $\beta \neq 0$, even though the process is $I(0)$.
- So, while it is neat that when $\beta \neq 0$ we get a standard normal limiting null distribution for t_{DF} [this is also true of t_{DF}^{μ}] the result is of little use because under H_1 t_{DF} will never be able to reject the (false) unit root null and we would wrongly be led to taking first differences when not needed [same holds for t_{DF}^{μ}].

- If $\beta = 0$, then t^{μ}_{DF} has the asymptotic critical values we gave before and, when $|\phi| < 1$ it is such that $t^{\mu}_{DF} \rightarrow -\infty$ as $T \rightarrow \infty$, and, hence, $Pr(t^{\mu}_{DF} < -2.86) \rightarrow 1$ as $T \rightarrow \infty$.
- Consequently, the behaviour of t^{μ}_{DF} depends on whether $\beta = 0$ or $\beta \neq 0$. This related to statistical concepts of *similarity* and *invariance*.
- *A (asymptotically) similar test is one whose (asymptotic) null distribution does not depend on nuisance parameters.*
- *An (asymptotically) invariant test is one whose null and alternative (asymptotic) distributions do not depend on nuisance parameters.*

- Here is a summary of similarity/invariance properties for the ADF tests (note the results in this table hold for arbitrary u_0):

	u_0	α	β
t_{DF}	AS,NI	NS, NI	NS, NI
t_{DF}^{μ}	S,NI	S,I	NS,NI
t_{DF}^{τ}	S,NI	S,I	S,I

  where: S=similar; I=invariant; NS = not similar; NI = not invariant, and AS = asymptotically similar.
- So, should we always use t_{DF}^{τ}? Yes, provided all we care about is getting a test which is invariant to β.
- However, for a given value of ϕ under H_1 the *power* (the probability of rejecting the null when the alternative is true, for a given significance level) of t_{DF}^{τ} is significantly lower than for t_{DF}^{μ}, which in turn is much lower than that for t_{DF}. So if $\beta = 0$, by using t_{DF}^{τ} we would be needlessly 'throwing power away'.

- This is a serious problem! If the data contain a trend ($\beta \neq 0$) then not using t^{τ}_{DF} is disastrous, because we know that power will approach zero. However, if there is no trend in the data ($\beta = 0$) then we lose a large amount of power in using t^{τ}_{DF} rather than t^{μ}_{DF}.
- A *conservative* strategy is to always use t^{τ}_{DF}.
- Other strategies are available ...

  Strategy (a): we might pre-test the data for the presence or otherwise of a linear trend by running the regression

$$y_t = \alpha + \beta t + \text{error}_t$$

  and testing $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ using a standard regression t-test.
- This approach is, however, problematic and the behaviour of the t-test depends on whether y_t is $I(1)$ or $I(0)$. This is an example of the so-called *spurious regression problem*, which we now explore.

Example: Spurious De-trending

- Suppose the true model is a random walk

$$y_t = y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim I(0)$$

  and we decide to de-trend the data by regressing y_t on an intercept and trend and taking the residuals from that regression.
- What happens?
- This is an example of the *spurious regression* phenomenon: there is no trend in the data - it's mean zero, but in the regression

$$y_t = \alpha + \beta t + \mathrm{error}_t$$

(i) The OLS estimate of β converges to zero (the true value of β), but the estimate of α diverges.

(ii) The t-statistic for $\beta = 0$ does not have a t-distribution. In fact it diverges as the sample size increases - *bad*!

- So, under strategy (a), we would use t^{τ}_{DF} if the trend test is significant, and use t^{μ}_{DF} if it is not. Notice then that if y_t is $I(1)$ then we will basically always be using t^{τ}_{DF}. If y_t is $I(0)$ then the efficacy of the procedure will come down to the power of trend test. If it doesn't have good power in small samples then there would be a higher chance that we used t^{μ}_{DF} even though β was non-zero - this would be disastrous!
- Could use trend tests that are robust to whether y_t is $I(0)$ or $I(1)$. Examples include Perron and Yabu (2009) and Harvey, Leybourne and Taylor (2007). Rather complicated though!

- *Strategy (b)*: *Union of Rejections* approach. Calculate t^{μ}_{DF} *and* t^{τ}_{DF} and reject the unit root null hypothesis if either t^{μ}_{DF} *or* t^{τ}_{DF} is significant at a chosen significance level.
- Strategy (b), although very simple, works very well in practice and seems to be the best compromise. See

  Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2009). Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. Econometric Theory 14, 587-636.

  for further details.

1.3 Allowing for Serially Correlated Shocks

- Thus far we have conducted unit root tests in the context of the first-order autoregressive [$AR(1)$] process in (1) or (3). This is very restrictive as simple $AR(1)$ processes do not seem in practice to model all of the serial correlation in economic series.
- To generalise, consider first the case where the shocks follow an $AR(p)$, $0 \leq p < \infty$, process (so that y_t is an $AR(p+1)$ process):

$$\begin{aligned} y_t &= \phi y_{t-1} + u_t, \quad t = 1, ..., T \\ u_t &= \sum_{i=1}^{p} \phi_i u_{t-i} + \varepsilon_t \end{aligned} \tag{5}$$

with $\varepsilon_t \sim IID(0, \sigma^2)$ and where u_t is a stationary $AR(p)$. For the present assume that p is *known*.

- Equation (5) can be re-written as

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \varepsilon_t, \quad t = p+1, ..., T. \tag{6}$$

- DF show that $\gamma = 0$ yields a unit root [or $I(1)$] process, while $-2 < \gamma < 0$ yields an $I(0)$ series.
- So to test $H_0 : \phi = 1$ against $H_1 : |\phi| < 1$, now estimate (6) by OLS and calculate the t-statistic for the significance of γ in (6). This is the *augmented* Dickey-Fuller, or ADF, statistic.
- The critical values for this test, at least in large samples, are the same as for the case of $p = 0$ which we tabulated before.

- Deterministic variables, such as a constant and linear trend, are dealt with as before - e.g. we either include a constant or a constant and linear trend when estimating (6).
- Alternatively, again as before, we could equally run regressions using the projection residuals from regressing each of the regressors from (6) individually on a constant or a constant and linear trend, or we could run the regressions using the de-trended (or de-meaned data).

- So far we have assumed that the data follow a finite-order autoregression whose order is known. How do we proceed if these assumptions are not justified?
- *Case 1: The series follows an* $AR(p+1)$, $0 \leq p < \infty$, *but* p *is unknown*:

  STEP 1: Run the (OLS) regression

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p^*} \gamma_i \Delta y_{t-i} + \varepsilon_t, \qquad (7)$$

  over $t = p^* + 1, ..., T$, where $p^* \geq p$. Choose p^* "large".

  STEP 2: Estimate (7) by OLS and test the null hypothesis $H_{0,p^*} : \gamma_{p^*} = 0$ against $H_{1,p^*} : \gamma_{p^*} \neq 0$. If we reject H_{0,p^*} then perform the ADF unit root test on γ in (7). If we accept H_{0,p^*}, go to step three.

STEP 3: Estimate the regression

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p^*-1} \gamma_i \Delta y_{t-i} + \varepsilon_t,$$

over $t = p^* + 1, ..., T$. Repeat step 2 above for γ_{p^*-1}.

- Continue this procedure until you find a maximum lag length, k^* say, that is significant. Then using that regression perform the ADF test. Proposed in Hall (1994, JBES).
- The large sample critical values for the resulting ADF test are as for the $p = 0$ case.
- Deterministic variables can be handled as before, including these regressors in the test regression used in each step.

- *Case 2: The shocks, u_t, follow a general linear process:*

  Here the time-series model for y_t is

$$\Delta y_t = \gamma y_{t-1} + u_t \tag{8}$$
$$u_t = \psi(L)\varepsilon_t \tag{9}$$

  where $\varepsilon_t \sim IID(0, \sigma^2)$ and where $\psi(L) := \sum_{j=0}^{\infty} \psi_j L^j$, $\psi_0 \equiv 1$, with L the *lag operator* such that $L^k y_t := y_{t-k}$, $k = 0, 1, 2, ...$, with the following conditions holding on $\psi(z)$:

$$\sum_{j=0}^{\infty} j|\psi_j| < \infty \tag{10}$$
$$\psi(z) \neq 0 \quad \forall |z| \leq 1. \tag{11}$$

- Equation (9) specifies an $MA(\infty)$ for u_t. This is often called a *linear process*.

- The condition in (10) implies that u_t is stationary, and (11) that it is invertible. All standard stationary and invertible $ARMA$ processes are therefore covered under the linear process structure.
- Under these conditions we can re-write (8)-(9) as

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{\infty} d_i u_{t-i} + \varepsilon_t. \tag{12}$$

- We then make use of the truncated, or *sieve*, autoregression,

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{k_T} d_i^* \Delta y_{t-i} + \text{error}_t \tag{13}$$

where k_T is a function of the sample size, T. We need to choose k_T such that $k_T \to \infty$ as $T \to \infty$ but $(k_T)^3/T \to 0$ as $T \to \infty$.

- This sieve-based approach was originally proposed in the context of the ADF tests by Saïd and Dickey (1984, Biometrika). They assumed further that $\epsilon_t \sim NIID(0, \sigma^2)$.
- Saïd and Dickey show that under the *rate conditions* on k_T stated above, the ADF regression t-statistic from (13) again has the usual $p = 0$ DF limiting null distribution (Section 1.1).
- But we can go further ...

- Using k_T as the maximum lag length, we can apply the Hall (1994) procedure we used in Case 1 above (setting $p^* = k_T$).
- This procedure is due to Ng and Perron (1995, JASA). This again retains the usual $p = 0$ DF asymptotic critical values.
- Deterministic variables handled as before.
- Full technical details can be found in Chang and Park (2002, Econometric Reviews) who generalise the theory in Saïd and Dickey (1984) to allow u_t to follow a more general (not necessarily Gaussian) linear process driven by conditionally heteroskedastic innovations (interestingly, the limiting null distribution of the ADF statistic does not depend on any nuisance parameters arising from the conditional heteroskedasticity).

1.4 Other Ways to Select the Lag Truncation, k^*: Information Criteria

1. Again select a maximum order, k_T, satisfying the rate conditions stated above.
2. Estimate the set of regressions

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta y_{t-i} + \text{error}_t \qquad (14)$$

   for each of $k = 0, 1, ..., k_T$, in each case over the sample data $t = k_T + 1, ..., T$, including deterministic regressors if needed.
3. Calculate the maximum likelihood estimate of the error variance in each case: $\hat{\sigma}_k^2 := T^{-1} \sum \hat{u}_{t,k}^2$, where $\hat{u}_{t,k}$ are the residuals from estimating (14) for k, $k = 0, ..., k_T$.

- Two famous information criteria are the Akaike and Schwart-Bayes information criteria, AIC and BIC respectively:

$$
\begin{aligned}
AIC &:= \min_{k=0,\dots,k_T} \left\{ \log \hat{\sigma}_k^2 + 2k/T \right\} \\
BIC &:= \min_{k=0,\dots,k_T} \left\{ \log \hat{\sigma}_k^2 + k(\log T)/T \right\}
\end{aligned}
$$

- Neither criterion seems to work especially well in practice in controlling the nominal finite sample size of the ADF test when the process contains MA behaviour. See Ng and Perron (2001, Econometrica) [NP] for extensive simulations and also the discussion in Haldrup and Jansson (2006).
- AIC also does not consistently estimate the true lag length in a pure AR case.

- Both AIC and BIC were originally designed for *stationary* time series models and so, as NP point out, this limits their use in the ADF regression context where the regressor y_{t-1} is non-stationary under the unit root null.
- Consequently, NP propose a new information criterion which accounts for the fact that y_{t-1} is a non-stationary regressor under the null. Their MAIC criterion is given by

$$MAIC \quad := \quad \min_{k=0,\ldots,k_T} \left\{ \log \tilde{\sigma}_k^2 + 2(\tau(k) + k)/(T - k_T) \right\}$$

where:

$$\tilde{\sigma}_k^2 := (T - k_T)^{-1} \sum_{t=k_T+1}^{T} \hat{u}_{t,k}^2, \quad \tau(k) := (\tilde{\sigma}_k^2)^{-1} \hat{\gamma}(k)^2 \sum_{k_T+1}^{T} y_{t-1}^2$$

and $\hat{\gamma}(k)$ is the OLS estimator of γ from (14).

1.5 Semi-Parametric Unit Root Tests

- An alternative to the parametric correction outlined above (the so-called ADF tests, where we include lags of Δy_t in the test regression), is considered in Phillips (1987, Econometrica) and Phillips and Perron (1988, Biometrika) [PP] using non-parametric methods to account for any serial correlation present in the shocks.
- Condition (11) may be dropped, except at $z = 1$.
- Phillips (1987) showed that if the data are generated as in (8)-(9) and we use the standard unaugmented DF test, t_{DF} (with no lagged dependent variables included) of $\gamma = 0$ from the regression

$$\Delta y_t = \gamma y_{t-1} + \text{error}_t$$

  that the limiting null distribution of t_{DF} depends on two nuisance parameters:

(i) the “short-run” variance of u_t,

$$\sigma_u^2 := E(u_t^2) = \sigma^2 \left(\sum_{j=0}^{\infty} \psi_j^2 \right)$$

(ii) the “long-run variance of u_t”,

$$\omega^2 := \lim_{T \to \infty} T^{-1} E \left[\left(\sum_{t=1}^{T} u_t \right)^2 \right] = \sigma^2 \left(\sum_{j=0}^{\infty} \psi_j \right)^2 = \sigma^2 \psi(1)^2.$$

- For any stationary process, the long run variance can also be written as

$$\omega^2 := \sigma_u^2 + 2\sum_{j=1}^{\infty} E(u_t\, u_{t-j}).$$

Example: Consider the $MA(1)$ process

$$u_t = \varepsilon_t + \theta\varepsilon_{t-1}, \;\; \varepsilon_t \sim IID(0, \sigma^2).$$

- In this case the short-run variance of u_t is given by

$$\sigma_u^2 = \sigma^2(1+\theta^2).$$

- The process is stationary and so its long-run variance is given by

$$\begin{aligned} \omega^2 &= \sigma_u^2 + 2E(u_t\, u_{t-1}) \\ &= \sigma^2 + \theta^2\sigma^2 + 2\theta\sigma^2 \\ &= \sigma^2(1+\theta)^2, \; (= \sigma^2\psi(1)^2). \end{aligned}$$

- PP use non-parametric estimators of the nuisance parameters σ_u^2 and ω^2 to modify the unaugmented DF statistic, t_{DF}, such that it has the usual (pivotal) limiting null distribution given before for the $AR(1)$ case. They propose the statistic

$$Z_t = \left(\frac{\hat{\sigma}_u^2}{\hat{\omega}^2}\right)^{1/2} t_{DF} - \frac{\hat{\omega}^2 - \hat{\sigma}_u^2}{2\left(\hat{\omega}^2 T^{-2} \sum_{t=1}^{T} y_{t-1}^2\right)^{1/2}}$$

where $\hat{\omega}^2$ and $\hat{\sigma}_u^2$ are consistent estimators of ω^2 and σ_u^2 respectively.

- These estimators are given by

$$\hat{\sigma}_u^2 := T^{-1} \sum_{t=1}^{T} \hat{u}_t^2 = T^{-1} \sum_{t=1}^{T} \left(y_t - \hat{\phi} y_{t-1} \right)^2$$

where $\hat{\phi}$ is the OLS estimator from regressing y_t on y_{t-1}, and

$$\hat{\omega}^2 := \hat{\sigma}_u^2 + 2 \sum_{j=1}^{m_T} w(j) \left[T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j} \right]$$

where $w(\cdot)$ is a *kernel function* and $m_T + 1$ is the *bandwidth*.

- The choice of m_T is rather like the choice of lag length in the ADF regression. The bandwidth needs to satisfy the condition that $m_T \to \infty$ as $T \to \infty$ such that $(m_T)^4/T \to 0$ as $T \to \infty$.
- What of the choice of kernel function? See the discussion in Haldrup and Jansson (2006,p.258), but a simple and popular choice is the Bartlett kernel: $w(j) := 1 - (j/(m_T + 1))$.
- Deterministic variables can again be handled by substituting t_{DF} by either t^{μ}_{DF} or t^{τ}_{DF} in the expression for Z_t and by calculating $\hat{\phi}$ and the residuals $\{\hat{u}_t\}$ from the regression of either $\bar{y}_t$ on $\bar{y}_{t-1}$ or $\breve{y}_t$ on $\breve{y}_{t-1}$. These statistics will then have the the limiting null distributions given previously for the pure $AR(1)$ case.

- An alternative estimator of $\hat{\omega}^2$ which was originally proposed by Berk (1974, Ann. Stats) and recently popularised by NP is calculated as follows.
- First estimate the ADF regression (possibly with deterministic regressors)

$$\Delta y_t = \hat{\gamma} y_{t-1} + \sum_{j=1}^{p} \hat{\gamma}_j \Delta y_{t-j} + \mathrm{error}_t \tag{15}$$

  where the lag truncation p could be chosen as outlined previously.
- Then calculate the *autoregressive spectral density estimator* of ω^2:

$$\hat{\omega}^2_{AR}(p) := \frac{\hat{\sigma}^2_p}{\left(1 - \sum_{j=1}^{p} \hat{\gamma}_j\right)^2}$$

  where $\hat{\sigma}^2_p$ and $\{\hat{\gamma}_j\}_{j=1}^{p}$ are respectively the OLS variance and slope estimators from (15). NB Assumes u_t in (9) satisfies (10) and (11).

1.6 The M Class of Unit Root Tests

- Originally proposed in Stock (1999, A Festschrift in Honour of Clive W.J. Granger, OUP) but popularised by NP the trinity of M tests take the form

$$
\begin{aligned}
MZ_\alpha &:= \frac{T^{-1}y_T^2 - \hat{\omega}_{AR}^2(p)}{2T^{-2}\sum_{t=1}^{T} y_{t-1}^2} \\
MSB &:= \left(T^{-2}\sum_{t=1}^{T} y_{t-1}^2/\hat{\omega}_{AR}^2(p)\right)^{1/2} \\
MZ_t &:= MZ_\alpha \times MSB
\end{aligned}
$$

  where $\hat{\omega}_{AR}^2(p)$ is the autoregressive spectral density estimator we just considered.

- The associated M tests reject H_0 for large negative values of MZ_α and MZ_t, and for small (close to zero) values of MSB.

- NP suggest using the MAIC criterion to select the lag truncation, p.
- The limiting null distributions of these statistics are given in NP. For example, MZ_t has the usual Dickey-Fuller distribution (and hence critical values), as given for t_{DF} before.
- To allow for deterministic regressors, simply construct the M tests as above but replacing y_t throughout (including in the ADF regression for constructing $\hat{\omega}^2_{AR}(p)$) by $\bar{y}^d_t$, the residuals from the OLS regression of y_t on d_t, $t = 0, ..., T$. Also, in the numerator of the expression for MZ_α add the term $-T^{-1}\bar{y}^d_0$. This ensures that, for example, the statistic MZ_t has the same limiting null distribution as t^μ_{DF} when $d_t = \alpha$ and as t^τ_{DF} when $d_t = \alpha + \beta t$.

1.7 Efficient (and Near-Efficient) Unit Root Tests

- The ADF test has no optimality properties, even in large samples - it is an example of what is called an *ad-hoc* test. We now turn to developing optimal unit root tests.
- Following Elliott, Rothenberg and Stock (1996, Econometrica) [ERS], consider the model

$$\begin{aligned} y_t &= d_t + u_t, \quad t = 1, ..., T \\ u_t &= \phi u_{t-1} + v_t \end{aligned}$$

where $d_t = \boldsymbol{\beta}' \mathbf{z}_t$ is a purely deterministic process, and v_t is an $I(0)$ process. Again we focus on testing the unit root null $H_0 : \phi = 1$ against the $I(0)$ alternative $H_1 : |\phi| < 1$.

- *Case 1:* Let $\{v_t\}$ be generated according to the linear process

$$v_t = \sum_{j=0}^{\infty} \delta_j \eta_{t-j}, \quad \eta_t \sim NIID(0,1)$$

  with $\sum_{j=0}^{\infty} j|\delta_j| < \infty$ (one summability), $u_0 = 0$ and the MA lag parameters $\{\delta_j\}_{j=0}^{\infty}$ all *known*.
- These conditions are, obviously, totally unrealistic but they allow us to obtain a theoretically optimal test, using the ...
- *Neyman-Pearson Lemma*, which states that a most powerful [MP] test of H_0 versus H_1 is obtained by rejecting for *large* values of the likelihood ratio, L_1^*/L_0^*, where L_j^* is the likelihood under H_j, $j = 0, 1$. Equivalently we can base the test on log-likelihoods using $\ell(1) - \ell(0)$, where $\ell(j) := \log(L_j^*)$, $j = 0, 1$.

- Now in this case we assume first that $d_t = 0$ [or that d_t is known]. Then we obtain that -2 times the log likelihood (ignoring the additive constant) is given by

$$L(\phi) = [\Delta \mathbf{u} - (\phi - 1)\mathbf{u}_{-1}]' \, \boldsymbol{\Sigma}^{-1} \, [\Delta \mathbf{u} - (\phi - 1)\mathbf{u}_{-1}]$$

  where $\Delta \mathbf{u} := (u_1, u_2 - u_1, ..., u_T - u_{T-1})'$ and $\mathbf{u}_{-1} := (u_0, u_1, ..., u_{T-1})' = (0, u_1, ..., u_{T-1})'$, and $\boldsymbol{\Sigma}$ is the (known) covariance matrix of $\mathbf{v} := (v_1, v_2, ..., v_T)'$.
- So, using the N-P Lemma, a MP test of $H_0 : \phi = 1$ versus $H_1 : \phi = \bar{\phi}$ rejects for *small* values (recall the -2 factor taken out of the log likelihoods above) of the statistic $L(\bar{\phi}) - L(1)$.
- To distinguish between the power properties of tests, consistency is too blunt an instrument since any sensible test should have power = 1 under the alternative in large samples.

- Instead we usually compare tests according to their power against *local alternatives* (known also as *local drift* or *Pitman drift*); i.e. power against $H_{1,c} : \phi = 1 + \bar{c}/T = \bar{\phi}$, or equally $\bar{c} = T(\bar{\phi} - 1)$, where $\bar{c} \leq 0$.
- So, by the N-P lemma, for such a local alternative the optimal (MP) test will reject for small values of $S_{\bar{c}} := L(1 + \bar{c}/T) - L(1)$.
- ERS demonstrate that under $H_{1,c} : \phi = 1 + c/T$,

$$S_{\bar{c}} \xrightarrow{w} \bar{c}^2 \int_0^1 W_c(r)^2 - \bar{c}[W_c(1)^2 - 1]$$

where $W_c(r) := \int_0^s \exp\left(-c\left(s - r\right)\right) dB\left(r\right)$, $B(s)$ a standard Brownian motion. This is known as a *standard Ornstein–Uhlenbeck* [OU] process with mean reversion parameter c. Notice that $W_0(r) = B(r)$. Observe that $\bar{c}$ is the value of the drift that we test against and c is the true drift.

- Setting $c = 0$ gives the limiting null distribution of the LR test, while setting $c = \bar{c}$ gives the optimal test against $\bar{c}$. From this we can obtain the *power envelope* for a given nominal significance level by 'joining up' all such tests: viz, $\pi(c) := \pi(c, c)$, where $\pi(c, \bar{c}) := \Pr\left[\bar{c}^2 \int_0^1 W_c(r)^2 - \bar{c}[W_c(1)^2 - 1] < cv_\xi\right]$, where cv_ξ denotes the critical value of the test for a significance level $\xi\%$.
- Unlike in classical testing problems, the N-P lemma does not yield a uniformly most powerful [UMP] test in the unit root testing problem. Each test, $S_{\bar{c}}$ is only MP against $\bar{c}$.
- Now we move onto the case where it is not assumed that $d_t = 0$.

- *Case 2:* Let $\{v_t\}$ be generated as in Case 1, but now we no longer assume that $d_t = 0$.
- This situation is more complicated than Case 1, and we need to use what are called *maximal invariants*. Let's make some notation first. Let $\mathbf{y}_a := (y_1, y_2 - ay_1, ..., y_T - ay_{T-1})'$ and let $\mathbf{Z}_a := (\mathbf{z}_1, \mathbf{z}_2 - a\mathbf{z}_1, ..., \mathbf{z}_T - a\mathbf{z}_{T-1})'$.
- Now (-2 times) the log-likelihood (again ignoring the additive constant) is given by

$$L(\phi, \boldsymbol{\beta}) = (\mathbf{y}_\phi - \mathbf{z}_\phi \boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_\phi - \mathbf{z}_\phi \boldsymbol{\beta})$$

and ERS show that a most powerful invariant [MPI] (where the invariance is w.r.t. $\boldsymbol{\beta}$) rejects for small values of the statistic

$$L_T^* := \min_{\boldsymbol{\beta}} L(\bar{\phi}, \boldsymbol{\beta}) - \min_{\boldsymbol{\beta}} L(1, \boldsymbol{\beta}).$$

- For the case where $d_t = \alpha$ (an unknown constant), ERS demonstrate that $L_T^* \xrightarrow{w} \bar{c}^2 \int_0^1 W_c(r)^2 - \bar{c}[W_c(1)^2 - 1]$, just as $S_{\bar{c}}$ did under Case 1.
- For the case where $d_t = \alpha + \beta t$, $L_T^* \xrightarrow{w} \bar{c} + \bar{c}^2 \int_0^1 V_c(s, \bar{c})^2 ds + (1 - \bar{c}) V_c(1, \bar{c})^2$, where $V_c(s, \bar{c}) := W_c(s) - s \left[\lambda W_c(1) + 3(1 - \lambda) \int_0^1 r W_c(r) dr \right]$ and $\lambda := (1 - \bar{c})(1 - \bar{c} + \bar{c}^2/3)^{-1}$.
- The representations for the power function of L_T^* in both cases then follow immediately using these limits, as before.
- Regardless of the form of d_t, no uniformly most powerful invariant test exists. Power functions depend on the deterministic elements in $\mathbf{z}_t$ and differ, in general, from Case 1 (the constant case being an exception).

- Thus far we have made some heroic assumptions: (i) $u_0 = 0$ and $\boldsymbol{\Sigma}$ is known; (ii) the process is Gaussian. ERS subsequently weaken these conditions to: (i) u_0 has a finite variance, which implies that $T^{-1/2}u_0 \xrightarrow{p} 0$, and v_t is as given before but it is no longer assumed that the $\{\delta_j\}$ parameters are known. We now denote the long run variance of v_t as ω_v^2. Assumption (ii) can be dropped in favour of $\eta_t \sim IID(0,1)$.
- Now consider the test which rejects for small values of the statistic

$$P_T := \left[S(\bar{\phi}) - \bar{\phi}S(1)\right]/\hat{\omega}_v^2$$

where $S(a) := (\mathbf{y}_a - \mathbf{z}_a\boldsymbol{\beta})'(\mathbf{y}_a - \mathbf{z}_a\boldsymbol{\beta})$ and $\hat{\omega}_v^2$ is any consistent estimator for ω_v^2.
- ERS show that this test has the same asymptotic (but not exact) power functions as the MP(I) tests detailed previously. This is a powerful (!) result.

- *Case 3:* Now suppose we want practical tests, recognising the fact that we do not know the true value of ϕ (i.e. of c).
- ERS suggest a modified version of P_T, denoted $P_T(\pi)$, which is run for a value of $\bar{c}$ (and, hence, $\bar{\phi}$) such that for a test run at the ξ % significance level the test has 50% power against $c = \bar{c}$. They show that such tests have asymptotic local power which lies very close to the asymptotic Gaussian local power envelope.
- ERS show that the appropriate choices for 50% power on 5% significance level tests are: $\bar{c} = -7$ for the constant case; and $\bar{c} = -13.5$ for the linear trend case.

- The most widely used tests proposed in ERS are their local-GLS de-trended ADF tests. These are constructed as follows. Notice that this procedure only differs from the standard ADF tests where de-trending is employed (deterministic variables are included).
- *STEP 1:* Regress (OLS) $\mathbf{y}_{\bar{\phi}}$ (where $\bar{\phi}$ is the value defined just above, corresponding to $\bar{c}$) on $\mathbf{Z}_{\bar{\phi}}$ to obtain the local GLS de-trended data $y_t^d := y_t - \hat{\boldsymbol{\beta}}' \mathbf{Z}_t$, where $\hat{\boldsymbol{\beta}}$ denotes the OLS estimator from this regression.
- *STEP 2:* Run the ADF-type regression

$$\Delta y_t^d = a_0 y_{t-1}^d + \sum_{j=1}^{p} a_j \Delta y_{t-j}^d + \text{error}_t \tag{16}$$

and calculate an ADF-type test for $a_0 = 0$.

- ERS show that for $d_t = \alpha$ the resulting test, $t_{DF}^{GLS,\mu}$ say, has the same local power function as t_{DF}; they show that this is given by,

$$t_{DF},\ t_{DF}^{GLS,\mu} \xrightarrow{w} \frac{0.5\left(W_c(1)^2 - 1\right)}{\left(\int_0^1 W_c(s)^2 ds\right)^{1/2}}$$

  while for the linear trend case, $d_t = \alpha + \beta t$, the GLS-type test, $t_{DF}^{GLS,\mu}$ say, satisfies

$$t_{DF}^{GLS,\tau} \xrightarrow{w} \frac{0.5\left(V_c(1,\bar{c})^2 - 1\right)}{\left(\int_0^1 V_c(s,\bar{c})^2 ds\right)^{1/2}}.$$

- Critical values for $t_{DF}^{GLS,\mu}$ and $t_{DF}^{GLS,\tau}$ are given in ERS on p.825. For the constant case, these are obviously the same as for t_{DF}.

- The choice of the lag truncation p in (16) can be made as outlined before.
- ERS show that these local GLS de-trended ADF-type tests lie arbitrarily close to the asymptotic Gaussian local power envelope.
- The PP tests and M tests that we discussed before can also be constructed using local GLS de-trended data. Let's use the M tests to illustrate.
- The M tests are formed as at the start of Section 1.6 except that the statistics are constructed using $\{y_t^d\}$ rather than $\{y_t\}$.

- The autoregressive spectral density estimator can also be formed using the analogue of (15) based on $\{y_t^d\}$, although Perron and Qu (2007, Economics Letters) show that superior finite sample power is obtained against fixed alternatives when constructing the autoregressive spectral density estimator using (15) augmented with the relevant deterministic regressors (there is no change to the asymptotic local power properties).
- The limiting distributions of the M tests based on local GLS de-trending are provided in NP, along with critical values. The MZ_t type test, for example, has the same limit distribution (and, hence, asymptotic local power function) as that for $t_{DF}^{GLS,\mu}$ in the constant case, and $t_{DF}^{GLS,\tau}$ in the linear trend case.

- NP show that the local GLS de-trended M tests all lie arbitrarily close to the asymptotic Gaussian local power envelope in both the constant and linear trend environments.
- NP show that the M tests using the autoregressive spectral density estimator of the long-run variance, coupled with local GLS de-trending show the best size and power of available unit root tests. See also, in particular, the detailed discussion in Haldrup and Jansson (2006), who also discuss the form of the limiting local power envelope for non-Gaussian cases.

1.8 The Impact of the Initial Condition on Unit Root Testing

- Thus far we have made quite strong assumptions on the *initial condition*, defined as the deviation of the first observation from its deterministic component, of the process. Throughout we have made assumptions which imply that $T^{-1/2}y_0$ (or $T^{-1/2}u_0$, depending on the precise model) converges in probability to zero. The scale factor $T^{-1/2}$ is crucial because, as is clear from the FCLT in Section 1.1, this implies that the initial condition will be asymptotically irrelevant under such an assumption. That is, the initial condition is assumed to be of smaller order in probability than the rest of the data.
- While this is convenient, since it implies that the limiting null distribution and asymptotic local power functions of the unit root tests do not depend on the initialisation, it is nonetheless unlikely to be true in reality.

- As discussed in Elliott and Müller (2006, Jnl. Econometrics), while there may be situations in which one would not necessarily expect the initial condition to be unusually large or, indeed, unusually small, relative to the other data points, equally the initial condition might be relatively large in other situations. The former case occurs, for example, where the first observation in the sample is dated quite some time after the inception of a mean-reverting process, while the latter could perhaps happen if the sample data happen to be chosen to start after a break (perceived or otherwise) in the series, or where the beginning of the sample coincides with the start of the process. This latter example can also allow for the case where an unusually small (even zero) initial condition occurs.

- In practice it is therefore hard to rule out small or large initial conditions, *a priori*. This is problematic because the magnitude of the initial condition can have a substantial impact on the power properties of unit root tests in practice (see, *inter alia*, Müller and Elliott, 2003, Econometrica) and, as discussed in and Elliott and Müller (2006), we observe only the initial observation rather than the initial condition.
- Müller and Elliott (2003) find that although the local-GLS de-trended ADF tests of ERS (Section 1.7) are considerably more powerful than the conventional OLS-detrended ADF tests (Section 1.2) when the initial condition is zero, the reverse is true when the initial condition is large.
- More discussion on this issue can again be found in Harvey, Leybourne and Taylor (2009, Econometric Theory) who propose union of rejections and weighed tests, based on the local-GLS de-trended ADF test and the OLS de-trended ADF test, for dealing with this issues.

1.9 Bootstrap Unit Root Tests

- As we have saw in Section 1.3, so called *sieve* methods have been proposed in the context of the ADF tests.
- Recall that we obtain the ADF t-ratio from the ADF regression (d_t is the deterministic term):

$$\Delta y_t = d_t + \gamma y_{t-1} + \sum_{i=1}^{k_T} d_i^* \Delta y_{t-i} + \text{error}_t \tag{17}$$

  where $k_T \to \infty$ as $T \to \infty$ with $(k_T)^3/T \to 0$ as $T \to \infty$.
- The ADF limiting null distribution can however be a very poor approximation to the finite sample null distribution of the sieve-based ADF statistic. This is therefore a case where a bootstrap implementation might be useful to obtain unit root tests with better finite sample properties.

- Park (2003, Econometrica) for the $AR(p)$ case and Chang and Park (2003, JTSA) for the $AR(\infty)$ (sieve) case show how to develop valid bootstrap implementations of the ADF tests. Both use what is called an *i.i.d. residual bootstrap* approach.
- **Aside:** Given a set of sample data, say $\{x_1, ..., x_T\}$, the i.i.d. bootstrap samples T points from $\{x_1, ..., x_T\}$ *with replacement*. The selected data points are chosen as random and independent draws from a given distribution, usually (though not necessarily) assigning equal probability to each data point; ie draws from a uniform distribution over $\{1, ..., T\}$.
- A residual i.i.d. bootstrap applies the same resampling principle to a set of regression residuals.
- The statistic of interest is then calculated from the bootstrap sample. If this is done B times we can obtain the empirical distribution function [EDF] of the bootstrap statistic.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17) satisfying S&D's rate condition on k_T.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17) satisfying S&D's rate condition on k_T.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k_T} \phi_j e_{t-j} + u_{k_T,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k_T$, and the residuals, $\tilde{u}_t$.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17) satisfying S&D's rate condition on k_T.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k_T} \phi_j e_{t-j} + u_{k_T,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k_T$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \overline{\tilde{u}}$, to get bootstrap residuals, u_t^*.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17) satisfying S&D's rate condition on k_T.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k_T} \phi_j e_{t-j} + u_{k_T,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k_T$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \overline{\tilde{u}}$, to get bootstrap residuals, u_t^*.
- Step 4: Recursively generate $e_t^* = \sum_{j=1}^{k_T} \tilde{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17) satisfying S&D's rate condition on k_T.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k_T} \phi_j e_{t-j} + u_{k_T,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k_T$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \bar{\tilde{u}}$, to get bootstrap residuals, u_t^*.
- Step 4: Recursively generate $e_t^* = \sum_{j=1}^{k_T} \tilde{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the e_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} e_j^*$, with y_0^* set to eg zero.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17) satisfying S&D's rate condition on k_T.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k_T} \phi_j e_{t-j} + u_{k_T,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k_T$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \bar{\tilde{u}}$, to get bootstrap residuals, u_t^*.
- Step 4: Recursively generate $e_t^* = \sum_{j=1}^{k_T} \tilde{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the e_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} e_j^*$, with y_0^* set to eg zero.
- Step 6: Calculate the bootstrap analogue of t_{ADF} in (17) applied to y_t^*.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17) satisfying S&D's rate condition on k_T.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k_T} \phi_j e_{t-j} + u_{k_T,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k_T$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \bar{\tilde{u}}$, to get bootstrap residuals, u_t^*.
- Step 4: Recursively generate $e_t^* = \sum_{j=1}^{k_T} \tilde{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the e_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} e_j^*$, with y_0^* set to eg zero.
- Step 6: Calculate the bootstrap analogue of t_{ADF} in (17) applied to y_t^*.
- Step 7: Perform Steps 2-6 B times to form the estimated bootstrap EDF. Obtain bootstrap p-value.

- C&P Demonstrate the asymptotic validity of their sieve bootstrap unit root test. However, they impose that the shocks, u_t are i.i.d. Their bootstrap is still valid with conditionally heteroskedastic errors, but won't replicate such effects in the bootstrap data.
- Their bootstrap is not, in general, valid if there is unconditional heteroskedasticity present in u_t. This is often called *non-stationary volatility*. A simple example occurs where the variance of ε_t in (9) displays a one-time break at some point in the sample.
- Although C&P argue that their sieve bootstrap loses no power relative to the test based on asymptotic critical values, their own simulations show large power losses relative to the standard ADF test under H_1. This occurs because Step 2 imposes H_0 on the sieve stage. Under H_1, e_t is non-invertible, violating the conditions for sieve validity.

- Cavaliere and Taylor (2008, Econometric Theory) address these problems proposing wild bootstrap ADF tests.
- **Aside:** With an original set of sample data, say $\{x_1, ..., x_T\}$, the wild bootstrap data is given by $x_t^* = x_t \times w_t$, where the w_t's are a sequence of independent random variables with mean zero and variance 1. Examples used for w_t include $NIID(0,1)$, and independent draws from the Rademacher distribution, which takes either the value 1 or -1, each with probability 0.5.
- Again the wild bootstrap resampling scheme can also be applied to regression residuals.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17), for some lag length k.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17), for some lag length k.
- Step 2: (Optional): Estimate (17) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17), for some lag length k.
- Step 2: (Optional): Estimate (17) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17), for some lag length k.
- Step 2: (Optional): Estimate (17) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.
- Step 4: (Optional) Recursively generate $e_t^* = \sum_{j=1}^{k} \hat{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17), for some lag length k.
- Step 2: (Optional): Estimate (17) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.
- Step 4: (Optional) Recursively generate $e_t^* = \sum_{j=1}^{k} \hat{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the u_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} u_j^*$, with eg y_0^* set to zero [replace u_j^* by e_j^* if Step 4 is performed].

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17), for some lag length k.
- Step 2: (Optional): Estimate (17) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.
- Step 4: (Optional) Recursively generate $e_t^* = \sum_{j=1}^{k} \hat{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the u_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} u_j^*$, with eg y_0^* set to zero [replace u_j^* by e_j^* if Step 4 is performed].
- Step 6: Calculate the wild bootstrap analogue of t_{ADF} in (17) applied to y_t^*.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{ADF}, from (17), for some lag length k.
- Step 2: (Optional): Estimate (17) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.
- Step 4: (Optional) Recursively generate $e_t^* = \sum_{j=1}^{k} \hat{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the u_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} u_j^*$, with eg y_0^* set to zero [replace u_j^* by e_j^* if Step 4 is performed].
- Step 6: Calculate the wild bootstrap analogue of t_{ADF} in (17) applied to y_t^*.
- Step 7: Perform Steps 2-6 B times to form the estimated bootstrap EDF. Obtain bootstrap p-value.

- Because the wild bootstrap kills weak correlations, there's no need to perform the sieve element for asymptotic validity, unlike with C&P's bootstrap. But including a sieve stage can improve finite sample size. Indeed k can be set to zero in the bootstrap version of (17) in Step 6 if the sieve stage is omitted.
- Notice that C&T do not impose H_0 when performing the (optional) sieve and, as a result, C&T's wild bootstrap ADF tests avoid the power losses seen with C&T's tests.
- C&T show that the wild bootstrap statistic, t^*_{ADF} say, has the same first order limiting distribution as the limiting null distribution of t_{ADF} under null, local and fixed alternatives. Hence, behaves like an infeasible size-corrected ADF test.
- C&T in various papers (eg Econometric Theory, 2009, and Econometric Reviews, 2009) show that the bootstrap ADF tests perform very well in the presence of both conditional heteroskedasticity and unconditional heteroskedasticity of many forms (eg volatility breaks, trending volatility, IGARCH, AR-SV, various GARCH -type models).

References

- Berk, K.N. (1974). Consistent autoregressive spectral estimates. Annals of Statistics 2, 489–502.
- Cavaliere G. and Taylor, A.M.R. (2008). Bootstrap unit root tests for time series with non-stationary volatility. Econometric Theory 24, 43-71.
- Cavaliere, G. and Taylor, A.M.R. (2009), Heteroskedastic time series with a unit root. Econometric Theory, 25, 1228–1276.
- Cavaliere G. and Taylor, A.M.R. (2009). Bootstrap M Unit Root Tests. Econometric Reviews 28, 393-421.
- Chang, Y. and Park, J. (2003). A sieve bootstrap for the test of a unit root. Journal of Time Series Analysis, 24, 379–400.
- Chang Y. and J.Y. Park (2002), On the asymptotics of ADF tests for unit roots, Econometric Reviews 21, 431–447.
- Dickey, D.A. and W.A. Fuller (1979), Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427-431.

- Dickey, D.A. and W.A. Fuller (1981), Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica 49, 1057-1072.
- Elliott, G. and Müller, U.K. (2006). Minimizing the impact of the initial condition on testing for unit roots. Journal of Econometrics 135. 285–310.
- Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.
- Haldrup, N. and Jansson, M. (2006) Improving Size and Power in Unit Root Testing. In: Miller, T.C. and Patterson, K., Eds., Palgrave Handbook of Econometrics, Vol.1, Palgrave Macmillan, London. https://dx.doi.org/10.2139/ssrn.1147594 Also available as a working paper at SSRN: https://ssrn.com/abstract=1147594

- Hall, A. (1994), Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection, Journal of Business and Economic Statistics, 12, 461–70.
- Hamilton, J.D. (1994). Time Series Analysis. Princeton University Press.
- Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2007). A simple, robust and powerful test of the trend hypothesis. Journal of Econometrics 141, 1302–1330.
- Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2009). Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. Econometric Theory 14, 587-636.
- Müller, U.K. and Elliott, G. (2003). Tests for unit roots and the initial condition. Econometrica 71, 1269-1286.

- Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. Journal of the American Statistical Association 90, 268-281.
- Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-1554.
- Park, J.Y. (2003), Bootstrap unit root tests, Econometrica 71, 1845–1895.
- Perron P. and Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. Review of Economic Studies 63, 435–463.
- Perron, P. and Qu, Z. (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. Economics Letters 94, 12-19.

- Perron, P. and Yabu, T. (2009). Estimating deterministic trends with an integrated or stationary noise component, Journal of Econometrics 151, 56–69.
- Phillips P.C.B. (1987), Time series regression with a unit root, Econometrica 55, 277–301.
- Phillips, P.C.B. and Perron, P. (1988). Testing for a unit root in time series regression. Biometrika 75, 335-346.
- Said S.E. and D.A. Dickey (1984), Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika 89, 1420–1437.
- Stock J.H. (1999), A class of tests for integration and cointegration, In Engle, R.F., and White, H. (Eds.), Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger, Oxford: Oxford University Press, 137–167.