E 4101/5101
Lecture 9: Deterministic trends vs integrated series; Typical spectral shape; Spurious regression; Dickey-Fuller distribution

Ragnar Nymoen

Department of Economics, University of Oslo

1 April 2014

Introduction I

Main references:

- H. Ch 15,16 and 17.
- D&M Ch 14.3 and 14.4
- Two Lecture Notes (RNy) about spectral analysis and complex numbers, in particular Section 9 in *Spectral analysis*.
- See the end of the slide set for additional references.

Deterministic trend—trend stationarity I

Let $\{y_t;\ t = 1, 2, 3, \ldots T\}$ define a time series (as before).
y_t follows a pure *deterministic trend* (DT) if

$$y_t = \phi_0 + \delta t + \varepsilon_t,\ \delta \neq 0 \tag{1}$$

where ε_t is white-noise and Gaussian.
y_t is non-stationary, since

$$E(y_t) = \phi_0 + \delta t \tag{2}$$

even though (in this case) the variance does not depend on time:

$$Var(y_t) = \sigma^2 \tag{3}$$

Deterministic trend—trend stationarity II

- In the pure DT model, the non-stationarity issue is resolved by **de-trending**. The de-trended variable:

$$y_t^s = y_t - \delta t$$

$Var(y_t^s) = \sigma^2$ and

$$E(y_t^s) = \phi_0$$

  - y_t^s is covariance stationary.
  - Since stationarity of y_t^s is obtained by subtracting the linear trend δt from y_t in (1), y_t is called a trend-stationary process.

Deterministic trend—trend stationarity III

- Assume that we are in period T and want a forecast for y_{T+h}. Assume that ϕ_0 and δ are known parameters in period T. The forecast is:

$$\hat{y}_{T+h|T} = \phi_0 + \delta(T + j)$$

  Assuming that the DGP is (1) also in the forecast period, the forecast error becomes:

$$y_{t+h} - \hat{y}_{T+h} = \varepsilon_{T+j}$$

  with

$$E\left[(y_{t+h} - \hat{y}_{T+h}) \mid T\right] = 0$$

Deterministic trend—trend stationarity IV

and variance:

$$Var(y_{t+h} - \hat{y}_{T+h}) \mid T] = \sigma^2$$

The conditional variance is the same as the unconditional variance (in the pure DT model).

Estimation and inference in the deterministic trend model I

Since the deterministic trend model can be placed within the ARMA class of models, it represents no new problems of estimation.

Still, the precise statistical analysis is not trivial, as Ch 16.1 in H shows.

Specifically, Hamilton (Ch. 16, p 460) shows that for (1)

$$y_t = \phi_0 + \delta\, t + \varepsilon_t,\; t = 1, 2, \ldots$$

and $\varepsilon_t \sim i.i.d.$ with $Var(\varepsilon_t) = \sigma^2$ and $E(\varepsilon_t^4) < \infty$,we have OLS estimators $\hat{\phi}_0$ and $\hat{\delta}$,

$$\begin{pmatrix} T^{1/2}(\hat{\phi}_0 - \phi_0) \\ T^{3/2}(\hat{\delta} - \delta) \end{pmatrix} \xrightarrow{D} N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \sigma^2 \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}^{-1} \right)$$

Estimation and inference in the deterministic trend model II

- The speed of convergence of $\hat{\delta}$ is $T^{3/2}$ (sometimes written as $O_p(T^{-3/2})$, for *order in probability*) while the usual speed of convergence for stationary variables is $T^{1/2}$
- $\hat{\delta}$ is so-called *super-consistent* and the OLS based $\widehat{Var(\hat{\delta})}$ has the same property.
- This means that the usual tests statistics have asymptotic N and χ^2 distributions as in the ARMA case., see H 16.2.

AR model with trend I

A more general DT model:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \delta t + \varepsilon_t, \ |\phi_1| < 1, \delta \neq 0 \tag{4}$$

Conditional on $y_0 = 0$, the solution is

$$\begin{aligned} y_t = \phi_0 \sum_{j=0}^{t} \phi_1^j - \delta \sum_{j=1}^{t-1} (\phi_1)^j j \\ + \delta \left(\sum_{j=1}^{t} \phi_1^{j-1} \right) \cdot t + \sum_{j=0}^{t} \phi_1^j \varepsilon_t \end{aligned} \tag{5}$$

If we define

$$y_t^s = y_t - \delta \left(\sum_{j=1}^{t} \phi_1^{j-1} \right) \cdot t$$

AR model with trend II

we get that also this de-trended variable is covariance stationary:

$$E(y_t^s) = \frac{\phi_0}{(1-\phi_1)} - \delta \frac{\phi_1}{(1-\phi_1)^2}$$
$$Var(y_t^s) = \frac{\sigma^2}{(1-\phi_1^2)}$$

where the result for $E(y_t^s)$ makes use of

$$\delta \sum_{j=1}^{t-1} \left(\phi_1^j\right) j \underset{t\to\infty}{\rightarrow} \delta \frac{\phi_1}{(1-\phi_1)^2}$$

OLS estimation of models with deterministic trend I

- Above, saw that $y_t \sim AR(1) + \textit{trend}$ can be transformed to $y_t^s \sim AR(1)$.
- The OLS estimators of **all** individual parameters, for example $(\hat{\phi}_0, \hat{\phi}_1, \hat{\delta})'$ are consistent at the usual rates of convergence ($\sqrt{T}$).

OLS estimation of models with deterministic trend II

- The reason why $\hat{\delta}$ is no longer super-consistent in the $AR(1) + trend$ model, is that $\hat{\delta}$ is a linear combination of moments that converge at different rates.
  - In such a situation, the slowest convergence rates dominates, it is $\sqrt{T}$.
  - See H Ch 16.3 p 463-467.
- The practical implication is that the stationary “asymptotic distribution theory” can be used also for dynamic models that include a DT
- For the $AR(p) + trend$ or $ARDL(p, p) + trend$ the conditional mean and variance of course depends on time, just as in the model without trend: Adds flexibility to pure DT model.

Other important forms of deterministic non-stationarity I

- The pure deterministic trend model (DT) can be considered a special case of

$$y_t = \phi_0 + \phi_1 y_{t-1} + \delta D(t) + \varepsilon_t$$

  where $D(t)$ is *any* deterministic (vector) function of time. It might be:
  - Seasonal dummies, or
  - Dummies for structural breaks (induce shifts in intercept and/or ϕ_1, gradually or as a deterministic shock))
- As long as the model with $D(t)$ can be re-expressed as a model with constant unconditional mean (with reference to the Frisch-Waugh theorem), this type of non-stationarity has no consequence for the statistical analysis of the model.

Stochastic (or local) trend I

$AR(p)$:

$$y_t = \phi_0 + \phi(L) y_{t-1} + \varepsilon_t \tag{6}$$

$$\phi(L) = \phi_1 L + \phi_2 L^2 + \ldots + \phi_p L^p.$$

Re-writing the model in the (now) well known way:

$$\Delta y_t = \phi_0 + \phi^{\ddagger}(L) \Delta y_{t-1} - \underbrace{(1 - \phi(1))}_{= p(1)} y_{t-1} + \varepsilon_t \tag{7}$$

The parameters $\phi_i^{\ddagger}$ in

$$\phi^{\ddagger}(L) = \phi_1^{\ddagger} L + \phi_2^{\ddagger} L^2 + \ldots + \phi_{p-1}^{\ddagger} L^{p-1} \tag{8}$$

are functions of the ϕ_i's.

Stochastic (or local) trend II

We know from before that y_t is stationary and causal if all roots of

$$p(\lambda) = \lambda^p - \phi_1 \lambda^{p-1} - \ldots - \phi_p \lambda \tag{9}$$

have modulus less than one. In the case of $\lambda = 1$ (one root is equal to 1),

$$p(1) = 1 - \phi(1) = 0. \tag{10}$$

and (7) becomes

$$\Delta y_t = \phi_0 + \sum_{i=1}^{p-1} \phi_i^{\ddagger} \Delta y_{t-i} \; + \varepsilon_t. \tag{11}$$

Stochastic (or local) trend III

Definition

y_t given by (6) is integrated of order 1, $y_t \sim I(1)$, if $p(\lambda) = 0$ has one characteristic root equal to 1.

- The stationary case is often referred to as $y_t \sim I(0)$, "integrated of order zero".
  - It follows that if $y_t \sim I(1)$, then $\Delta y_t \sim I(0)$.
  - An integrated series y_t is called *difference stationary*.
- With reference to our earlier discussion of stationarity and spectral analysis, we see that the definition above is not general:
  - The characteristic polynomial of an $AR(p)$ series can have other unit-roots than the real root 1.

Stochastic (or local) trend IV

- The unit-root defined by (10) corresponds to a root at the *long-run frequency*. It implies that $y_t \sim I(1)$ series has the so called "typical spectral shape".

- In the following, we will abstract from unit roots at the seasonal or business cycle frequencies.
  - We do not cover seasonal integration, of which Hylleberg et. al. (1990) is the classic reference.
- However, we note that the analysis of long-frequency unit-root can be extended to integration of order 2: $y_t \sim I(2)$ if $\Delta^2 y_t \sim I(0)$, where $\Delta^2 = (1 - L)^2$.
- In the $I(2)$ case, there must be a unit root in the characteristic polynomial associated with (11):

$$p(\lambda^{\ddagger}) = \lambda^{p-1} - \phi_1^{\ddagger}\lambda^{p-2} - \ldots - \phi_{p-1}^{\ddagger}.$$

Contrasting I(0) and I(1) I

:

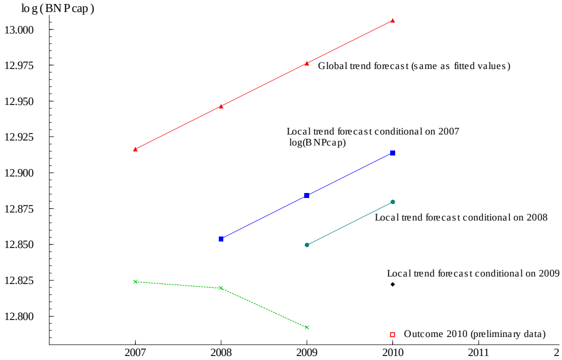
	I(1)	I(0)
1 $Var[y_t]$	$= \infty$	finite
2 $Corr[y_t, y_{t-p}]$	≈ 1	$\rightarrow 0$
3 Multipliers	Do not "die out"	$\rightarrow 0$
4a Forecasting y_{T+h}	$E(y_{T+h\|T})$ depends on y_T $\forall h$	$\underset{h\rightarrow\infty}{\rightarrow} E(y_t)$
4b Forecasting,y_{T+h}	*Var* of forecast errors $\rightarrow \infty$	$\rightarrow$ finite
5 PSD	Typical shape	Finite at all v
6 Inference	Non-standard theory	Standard

1-4 are easy to demonstrate for the Random Walk (RW) with drift:

$$y_t = \phi_0 + y_{t-1} + \varepsilon_t, \qquad (12)$$

Contrasting I(0) and I(1) II

in fact we have worked with this is in the exercises to Seminar 4. For example: The non-adaptive nature of the forecast of the DT-model of GDP per capita in Norway.



Forecasting GNP per capita

- by DT

and

- local trend model

Power Spectral Density of AR(1) I

With reference to the two lecture notes about complex numbers and spectral analysis we have

$$f_{y,ARMA(1,0)} = \frac{\sigma^2}{1 - 2\phi_1 \cos(2\pi v) + \phi_1^2}. \tag{13}$$

for the AR(1) process.

Power Spectral Density of AR(1) II

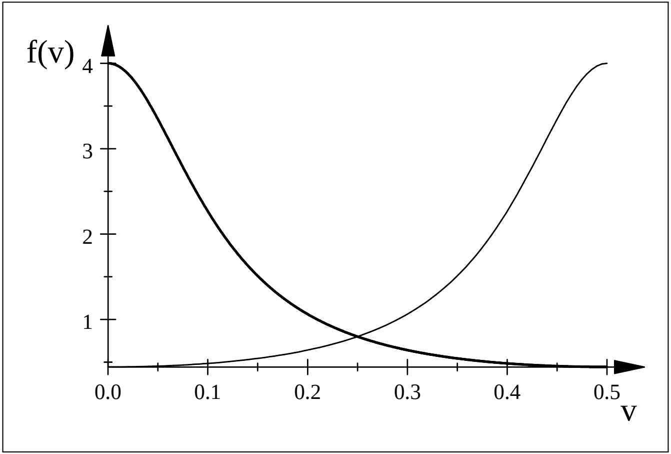


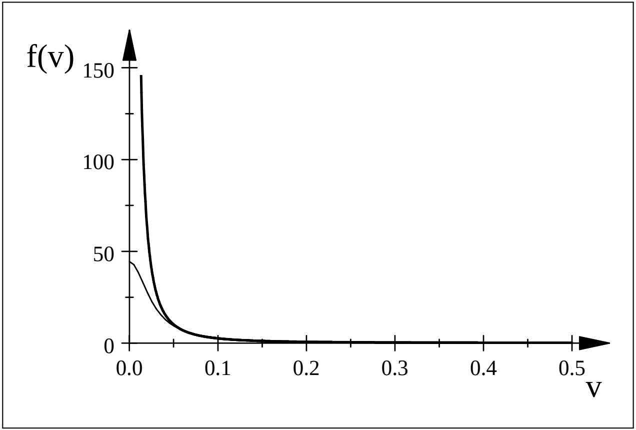
Figure: PSD of AR(1) with $\phi_1 = 0.5$, thick line and $\phi_1 = -0.5$, thin line

Power Spectral Density of Random Walk I

When $\phi_1 = 1$ in (13 we get the PSD of the Random Walk

$$f_{rw}(v) = \frac{\sigma^2}{2(1 - \cos(2\pi v))} \tag{14}$$

Power Spectral Density of Random Walk II

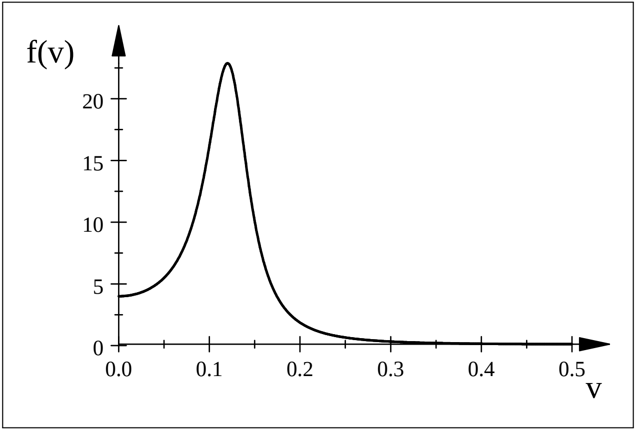


PSD of random-walk (thick line) and AR(1) with $\phi_1 = 0.85$, thin line.

Power Spectral Density of ARMA(2,1) I

$$f_{y,ARMA(2,1)}(v) = \sigma^2 \frac{1+\theta_1 2\cos(2\pi v)+\theta_2^2}{1+\phi_1^2+\phi_2^2-\phi_1(1-\phi_2)2\cos(2\pi v)-\phi_2 2\cos(4\pi v)}. \tag{15}$$

Power Spectral Density of ARMA(2,1) II



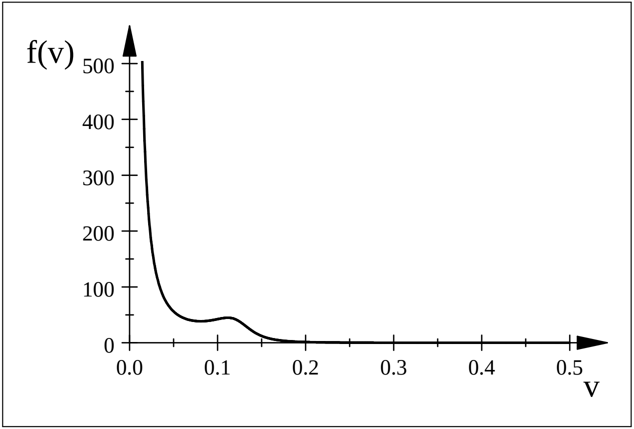
PSD of ARMA(2,0) with $\phi_1 = 1,2$ and $\phi_2 = -0,7$.

Power Spectral Density of ARIMA(p,1,q) I

- Like the ARMA(2,1), the PDF of the stationary ARMA(p,q) is finite at all frequencies.
- It can be shown (See the *Spectral analysis Lecture note*) that the PSD of a series that is ARMA(p,q) **after** differencing is

$$f_{ARIMA(p,1,q)}(v) = f_{rw}(v) f_{ARMA(p,q)}(v) \tag{16}$$

Power Spectral Density of ARIMA(p,1,q) II



PSD to ARIMA(2,1,0) with autoregressive parameters $\phi_1 = 1.2$ and $\phi_2 = -0.7$.

Spurious regression I

Granger and Newbold (1974) observed that

1. Economic time series were typically $I(1)$;
2. Econometricians used conventional inference theory to test hypotheses about relationships between $I(1)$ series

- G&N used Monte-Carlo analysis to show that 1. and 2. imply that to many "significant relationships are found" in economics
- Seemingly significant relationships between independent $I(1)-$variables were dubbed *spurious regressions*.

Spurious regression II

To replicate G&N results, we let YA_t and YB_t be generated by

$$YA_t = \phi_{A1} YA_{t-1} + \varepsilon_{A,t}$$
$$YB_t = \phi_{B1} YB_{t-1} + \varepsilon_{B,t}$$

where

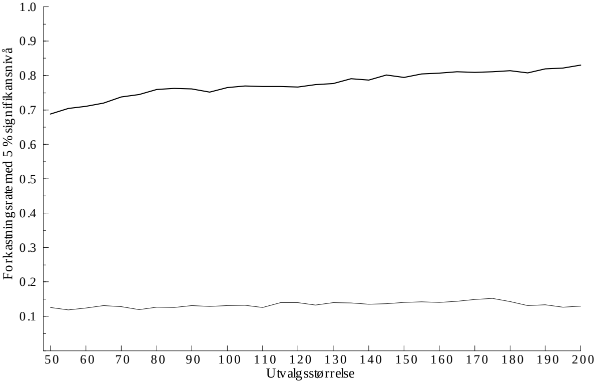
$$\begin{pmatrix} \varepsilon_{A,t} \\ \varepsilon_{B,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix} \right).$$

The DGP is a 1st order VAR. YA_t, YB_t are independent random walks if $\phi_{A1} = \phi_{B1} = 1$, and stationary if $|\phi_{A1}|$ and $|\phi_{B1}| < 1$.
The regression is

$$YA_t = \alpha + \beta YB_t + e_t$$

and the hypothesis is H_0: $\beta = 0$.

Spurious regression III



Rejection frequencies for H_0: $\beta = 0$ in the model $YA_t = \alpha + \beta YB_t + \varepsilon_t$ when ε_t is $I(0)$ (lowest line), and I(1) (highest). 5% nominal level.

Summary of Monte-Carlo of static regression

- With stationary variables:
  - wrong inference (too high rejection frequencies) because of positive residual autocorrelation
  - but $\hat{\beta}$ is consistent
- With *I(1)* variables:
  - rejection frequencies even higher and growing with T
  - Indication that $\hat{\beta}$ is inconsistent under the null of $\beta = 0$.
  - ... what *is* the distribution of $\hat{\beta}$?

Dynamic regression model I

In retrospect we can ask: Was the G&N analysis a bit of a strawman?

After all ,the regression model is obviously **mis-specified**.

And the true DGP is not nested in the model.

To check: use same DGP, but replace static regression by

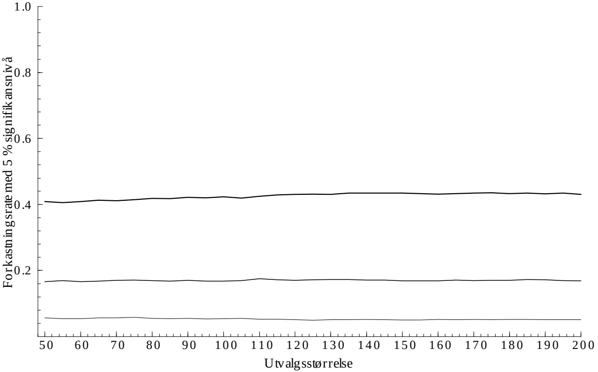
$$\Delta YA_t = \phi_0 + \rho YA_{t-1} + \beta_0 \Delta YB_t + \beta_1 YB_{t-1} + \varepsilon_{At} \qquad (17)$$

Under the null hypothesis:

$$\rho = 0$$
$$\beta_0 = \beta_1 = 0$$

and there is no residual autocorrelation, neither under H_0, nor under H_1.

Dynamic regression model II



Spurious regression in an ADL model Lines show rejection frequencised for H_0: $\rho = 0$ (highest), H_0: $(\beta_0 + \beta_1) = 0$ and H_0: $\beta_0 = 0$.

The Dickey Fuller distribution I

We now let the Data Generating Process (DGP) for $y_t \sim I(1)$ be the simple Gaussian Random Walk:

$$y_t = y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2) \tag{18}$$

We estimate the model

$$y_t = \rho y_{t-1} + u_t, \tag{19}$$

where our choice of OLS estimation is based on an assumption about white-noise disturbances u_t.
At the start, we know that, since the model can be written as

$$\Delta y_t = (\rho - 1) y_{t-1} + u_t$$

The Dickey Fuller distribution II

the OLS estimate $\widehat{(\rho - 1)}$ is consistent: The stationary (finite variance) series Δy_t cannot depend on the infinite variance variable y_{t-1}, see Hamilton p 488.

- However consistency alone doesn't guarantee that

$$\sqrt{T} \cdot (\hat{\rho} - 1)$$

  has a normal limiting distribution in this case ($\rho = 1$).
- In fact, $\sqrt{T} \cdot (\hat{\rho} - 1)$ has a degenerate asymptotic distribution since it can be shown that the speed of convergence is T when $\rho = 1$ in the DGP, another instance of super consistency.

The Dickey Fuller distribution III

We therefore seek the asymptotic distribution of the OLS based stochastic variable:

$$T \cdot (\hat{\rho} - 1) = \frac{\frac{1}{T} \sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2}. \tag{20}$$

when the DGP is (18).
Numerator in (20):

$$\frac{1}{T} \sum_{i=1}^{T} y_{t-1} \varepsilon_t \xrightarrow[T \to \infty.]{L} \frac{1}{2} \sigma^2 (X - 1), \tag{21}$$

where the stochastic variable X is distributed $\chi^2(1)$, see Hamilton's Ch 17.1

The Dickey Fuller distribution IV

The denominator in (20), see H p. 487

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 \underset{T\to\infty.}{\overset{L}{\longrightarrow}} \sigma^2 \int_0^1 [W(r)]^2\, dr, \qquad (22)$$

where $W(r)$, $r \in [0,1]$, is a "Standard Brownian motion", see Ch 17.2 in H for an explanation.

- $W(r)$ is a process that defines stochastic variables for any r. For example: $W(1) \sim N(0,1)$, but when $r < 1$, $W(r)$ is "something different" than the normal distribution.

The Dickey Fuller distribution V

- The result in (22) is an application of two theorems: "The functional limit theorem", which is a generalization of the central limit theorem to the random-walk case , and the "Continuous mapping theorem", which generalizes Cramer's theorem, see H Ch 17.3
- The asymptotic distribution of $T \cdot (\hat{\rho} - 1)$ turns out to be

$$T \cdot (\hat{\rho} - 1) \underset{T \to \infty.}{\overset{L}{\longrightarrow}} \frac{\frac{1}{2}(X - 1)}{\int_0^1 [W(r)]^2 \, dr} \tag{23}$$

- $\chi^2(1)$ is heavily skewed to the left (towards zero). Only 32% of the distribution lies to the right of 1. This means that values of X that makes the numerator in (23) negative have probability 0.68.

The Dickey Fuller distribution VI

- The denominator is always positive.
- As a result, we see that negative $(\hat{\rho} - 1)$ values will be over-represented when the true value of ρ is 1.

The distribution in (23) is called a **Dickey-Fuller (D-F)** distribution.
Under the H_0 of $\rho = 1$, also the "t-statistic" from OLS on (19) has a Dickey-Fuller distribution, which is of course relevant for practical testing of this H_0.

$$t_{DF} = \frac{\hat{\rho} - 1}{se(\hat{\rho})} \tag{24}$$

The Dickey Fuller distribution VII

where

$$se(\hat{\rho}) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^{T} y_{t-1}^2}}$$

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \hat{\rho} y_{t-1})^2$$

$\hat{\sigma}^2$ is consistent since $\hat{\rho}$ is consistent.
"Written out", t_{DF} is:

$$t_{DF} = \frac{T \cdot (\hat{\rho} - 1)\sqrt{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2}}{\sqrt{\hat{\sigma}^2}}$$

The Dickey Fuller distribution VIII

Substitute for $T \cdot (\hat{\rho} - 1)$ in the numerator:

$$t_{DF} = \frac{\frac{1}{T} \sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\sqrt{\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2} \sqrt{\hat{\sigma}^2}}$$

The numerator converges to (21), while we can use (22) in the denominator, so that

$$\sqrt{\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2} \sqrt{\hat{\sigma}^2} \xrightarrow[T \to \infty.]{L} \sqrt{\sigma^2 \int_0^1 [W(r)]^2 \, dr} \cdot \sigma = \sigma^2 \sqrt{\int_0^1 [W(r)]^2 \, dr},$$

The Dickey Fuller distribution IX

and

$$t_{DF} \xrightarrow[T\to\infty.]{L} \frac{\frac{1}{2}(X-1)}{\sqrt{\int_0^1 \left[W(r)\right]^2 dr}} \tag{25}$$

Which is not a normal distribution.

- Intuitively, because of the skewness of X, the left-tail 5 % fractile of this Dickey-Fuller distribution will be more negative than those of the normal.
- A very usefull, and pedagogical, reference is Ericsson and MacKinnon (2002), which also cover the extension to cointegration (as the title shows)

Dickey-Fuller tables and models I

The distributions (22) and (25) have been tabulated by Monte-Carlo simulation, see Hamilton's Table B.5 and B.6.

- The tables are based on the following, DGP, ("True process" in Hamilton)

$$y_t = y_{t-1} + \varepsilon_t, \quad , \quad \varepsilon_t \sim N(0, \sigma^2), \quad y_0 = 0. \tag{26}$$

Table B.6 gives the distribution of the "t-value" of $(\hat{\rho} - 1)$ from the following *models*

$$\Delta y_t = (\rho - 1) y_{t-1} + u_t \tag{a}$$

$$\Delta y_t = \mu + (\rho - 1) y_{t-1} + u_t \tag{b}$$

$$\Delta y_t = \mu + (\rho - 1) y_{t-1} + \gamma \cdot t + u_t \tag{c}$$

Dickey-Fuller tables and models II

- Model a. is the case of "model = DGP" under H_0: $\rho = 1$, the case above, and is "Case 1" in Table B.6.
- "Case 2" and "Case 4" in Table B.6 show that the two t_{DF} have different distributions when model (b) or (c) are estimated (the DGP is still (26).
- "Case 2". The model contains an intercept (which is not present in the DGP). The distribution is shifted to the left.
- "Case 4". Model contains μ and trend, γt, (drift may or may not be in the DGP).

Similar tests I

- The DF-test seems to depend on DGP = model (26).
- In practice we don't know if the DGP, in addition to potentially $\rho = 1$, also contains so called "nuisance" parameters, in the form of constant terms and trend.
- Tests that give correct inference about parameters of interest (e.g., 5 % prob of rejection of $\rho = 1$,when H_0 is true) even when DGP contains nuisance parameters are called *similar tests*.

Similar tests II

It turns out that the D-F tables can be used to construct "similar tests:

DGP	Models that give similar tests:*
i) $y_t = y_{t-1} + \varepsilon_t, \quad y_0 = 0.$	(a), (b), (c)
ii) $y_t = y_{t-1} + \varepsilon_t$, arbitrary y_0	(b), (c)
iii) $y_t = \mu + y_{t-1} + \varepsilon_t$, arbitrary y_0	(c)

* given that the correct tables are used!

(27)

The "rule of thumb" is that the model should be congruent with DGP, or include deterministic terms "of higher order" than those in the DGP.

A special case with standard inference I

If the DGP is

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

while the model is of type b:

$$\Delta y_t = \mu + (\rho - 1)y_{t-1} + u_t, \quad (28)$$

the t-value of $(\hat{\rho} - 1)$ becomes $N(0,1)$, as in Hamilton's "Case 3".
It is *as if* the regressor in (28) is a deterministic trend rather than y_{t-1}.
The reason is that, under H_0, there is a DT in y_t (remember the solution!) in addition to a stochastic trend (ST).
And, the DT dominates the properties of y_t when T becomes large.

Augmented Dickey-Fuller tests I

Let the DGP be the $AR(p)$

$$y_t - \sum_{i=1}^{p} \phi_i y_{t-i} = \varepsilon_t \tag{29}$$

with $\varepsilon_t \sim N(0, \sigma^2)$. We have the reparameterization:

$$\Delta y_t = \sum_{i=1}^{p-1} \phi_i^{\ddagger} \Delta y_{t-i} \; - (1 - \phi(1)) y_{t-1} + \varepsilon_t \tag{30}$$

$y_t \sim I(1)$ is implied by $(1 - \phi(1)) \equiv \rho = 0$
But a simple D-F regression will have autocorrelated u_t in the light of this DGP: one or more lag-coefficient $\phi_i^{\ddagger} \neq 0$ are omitted.

Augmented Dickey-Fuller tests II

The augmented Dickey-Fuller test (ADF), see Ch 17.7, is based on the model

$$\Delta y_t = \sum_{i=1}^{k-1} b_i \Delta y_{t-i} + (\rho - 1) y_{t-1} + u_t \tag{31}$$

Estimate by OLS and calculate the t_{DF} form this ADF regression.

- The asymptotic distribution is that same as in the first order case (with a simple random walk).
- The degree of augmentation can be determined by a specification search. Start with high k and stop when a *standard t-test* rejects null of $b_{k-1} = 0$

Augmented Dickey-Fuller tests III

- The determination of lag length" is an important step in practice since
  - Too low k destroys the level of the test (dynamic mis-specification),
  - Too high k lead to loss of power (over-parameterization).
- The ADF test can be regarded as one way of tackling "unit-root processes" with serial correlation
- Another much used test procedure, know as *Phillips-Perron* tests, is presented by Hamilton Ch 17.5.
- D&M also mentions alternatives to ADF, on page 623.
- The are several other tests for unit-roots as well—including tests where the null-hypotheses is stationarity and the alternative is non-stationary.

Augmented Dickey-Fuller tests IV

- As one example of the continuing interest in these topics: The book by Patterson (2011) contains a comprehensive review.

References

Ericsson N.R. and J.G. MacKinnon (2002): Distributions of error correction tests for cointegration,*Econometrics Journal*, 5,285—318
Granger C.W.J and P. Newbold (1974) Spurious Regressions in Econometrics, *Journal of Econometrics*, 2, 111-120.
Hylleberg, S, R. F. Engle, C. W. J. Granger and B. S. Yoo (1990) Seasonal Integration and Cointegration, *Journal of Econometrics*, 1990, 44, 215-238.
Patterson, K. (2011), Unit Root Tests in Time Series. Volume 1: Key Concepts and Problems, Palgrave MacMillan.
Sims, C., J. Stock, and M. Watson (1990) Inference in Linear Time Series Models with some Unit Roots *Econometrica*, 58, 113–44