E 4160 Autumn term 2015. Lecture 9: Deterministic trends vs integrated series; Spurious regression; Dickey-Fuller distribution

Ragnar Nymoen

Department of Economics, University of Oslo

14 October 2015

Introduction I

Main references:

- HN Ch 16.
- DM Ch 14.3 and 14.4
- BN201422 Kap 10
- See the end of the slide set for additional references (one of them posted on the web side as an important resource when testing for unit-root and later cointegration)

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:

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

$Var(Y_t) = \sigma^2$ 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*.
- Assume that we are in period T and want a forecast for Y_{T+h}. Assume that ϕ_0 and δ are known to us from history (mainly to simplify notation).

Deterministic trend—trend stationarity III

- The forecast is then:

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

- Assume (and this is critical) that the parameters ϕ_0 and δ remain constant over the whole forecast period, The forecast error becomes:

$$Y_{t+h} - \hat{Y}_{T+h} = \varepsilon_{T+j}$$

with

$$E[(Y_{t+h} - \hat{Y}_{T+h}) \mid T] = 0$$

and variance:

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

The conditional variance of the forecast error 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 stationary time series framework, it represents no new problems of estimation.
- Nevertheless, the precise statistical analysis is non-trivial. For example, for (1)

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

and $\varepsilon_t \sim IID$. with $Var(\varepsilon_t) = \sigma^2$ and $E(\varepsilon_t^4) < \infty$, it has been shown for the 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,**
- $\widehat{Var(\hat{\delta})}$ has the same property in this model, meaning that the usual tests statistics have teh usual asymptotic N and χ^2 distributions.

AR model with trend I

A more general DT model:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \delta t + \varepsilon_t, \; |\phi_1| < 1, \, \delta \neq 0 \qquad (4)$$

The solution is (take as an exercise!):

$$Y_t = \phi_0 \sum_{j=0}^{t} \phi_1^j - \delta \sum_{j=1}^{t-1} (\phi_1)^j j \qquad (5)$$
$$+ \delta \left(\sum_{j=1}^{t} \phi_1^{j-1} \right) \cdot t + \phi_1^t Y_0 + \sum_{j=0}^{t} \phi_1^j \varepsilon_t$$

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 find 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

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

OLS estimation of models with deterministic trend I

- We have seen 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 $(\sqrt{T})$.
- The reason why $\hat{\delta}$ is no longer super-consistent in the $AR(1) + \textit{trend}$ 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 $\sqrt{T}$.

OLS estimation of models with deterministic trend II

- The practical implication is that the standard asymptotic distribution theory can be used also for dynamic models that include a DT, as long as the homogenous part of the AR part of the model satisfies the conditions of weak stationarity.
- 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

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

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

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

(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 also called *difference stationary*.
- With reference to our earlier discussion of stationarity, we see that this definition (although common) is not general:
  - The characteristic polynomial of an $AR(p)$ series can have other unit-roots than the real root 1.
  - In fact, the unit-root defined by (10) corresponds to a unit-root at the "zero frequency" or *"long-run frequency"*. In order to make this concept precise, spectral analysis is needed.

Stochastic (or local) trend IV

  - In practice, the preclusion of unit-roots at "non-zero frequencies" means that we abstract from seasonal integration ("summer may become winter") and unit-roots at the business-cycle frequencies (boom may become bust).

- 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(1)	I(0)
1. $E(Y_t)$	depends on Y_0	constant
2 $Var[Y_t]$	$= \infty$	constant
3 $Corr[Y_t, Y_{t-p}]$	≈ 1	$\underset{p \to \infty}{\longrightarrow} 0$
4 Multipliers	Do not "die out"	$\to 0$
5a Forecasting Y_{T+h}	$E(Y_{T+h\|T})$ depends on Y_T $\forall h$	$\underset{h \to \infty}{\longrightarrow} E(Y_t)$
5b Forecasting, Y_{T+h}	Var of forecast errors $\to \infty$	$\to$ finite
6 Inference	Non-standard theory	Standard

Try to show 1-5 for the Random Walk (RW) with drift:

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

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 the data generating process (DGP):

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

where

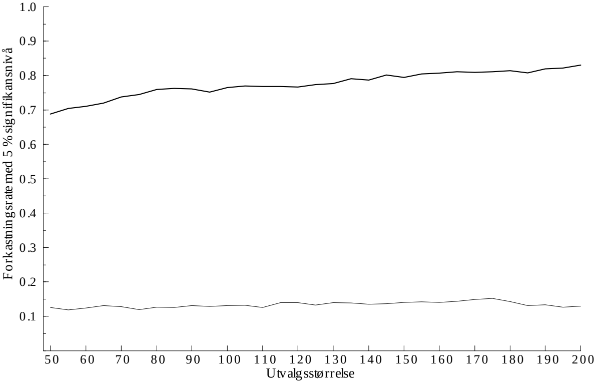
$$\begin{pmatrix} \varepsilon_{A,t} \\ \varepsilon_{B,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix} \right).$$

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 tested is $H_0\colon\ \beta = 0$.

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

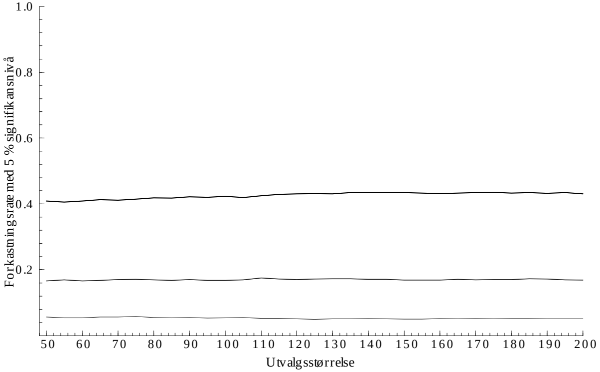
$$\Delta YA_t = \phi_0 + \rho YA_{t-1} + \beta_0 \Delta YB_t + \beta_1 YB_{t-1} + \varepsilon_{At} \tag{13}$$

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 ADL regression model (13) performs better than the static regression,
  - for example, $t_{\hat{\beta}_0}$ seems to behave as in the stationary case.
  - This does hold true in general, since β_0 is a coefficient on a stationary variable.
- But inference based on $t_{\hat{\beta}_0}$ and $t_{\hat{\beta}_1}$ continues to over-reject (the size of the test is wrong) also in the dynamic model.
- Conclude that the spurious regression problem is fundamental.
- We need **non-standard inference theory** before it can be tackled.
- Start with unit-root testing.

The Dickey Fuller(DF) distribution I

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

$$Y_t = Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2) \tag{14}$$

We estimate the model

$$Y_t = \rho Y_{t-1} + u_t, \tag{15}$$

where our choice of OLS estimation is based on an assumption about white-noise disturbances u_t.
Since the model can be re-parameterized as

$$\Delta Y_t = (\rho - 1) Y_{t-1} + u_t$$

The Dickey Fuller(DF) distribution II

we understand intuitively that the OLS estimator $\widehat{(\rho-1)}$ is consistent: The stationary (finite variance) series ΔY_t cannot depend on the infinite variance variable Y_{t-1}.

- However, consistency alone does not guarantee that

$$\sqrt{T}\cdot(\hat{\rho}-1)=\frac{\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t}{\frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}^2} \tag{16}$$

has a normal limiting distribution in this case (when indeed $\rho=1$).In fact, we suspect that the distibution collapses to 0, since $\hat{\rho}$ approaches 1 at a rate faster than $\sqrt{T}$.

The Dickey Fuller(DF) distribution III

- To compensate that we change from $\sqrt{T}$to T. It has been shown that

$$T \cdot (\hat{\rho} - 1) \xrightarrow[T \to \infty.]{L} \frac{\frac{1}{2}(X - 1)}{\int_0^1 \left[W(r)\right]^2 dr} \tag{17}$$

- In the denominator, $W(r)$ represents a (Standard Brownian Motion) 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.
  - But the important thing to note is that the denominator is always positive, meaning that the sign of the bias depens on the numerator.

The Dickey Fuller(DF) distribution IV

- The random variable X in the numerator of (17) is distributed $\chi^2(1)$, and values close to 0 are therefore quite probable.
- As a result, negative $(\hat{\rho} - 1)$ values will be over-represented when the true value of ρ is 1.
- The distribution in (17) is called a **Dickey-Fuller (D-F)** distribution.

Under the H_0 of $\rho = 1$, also the "t-statistic" from OLS on (15) has a Dickey-Fuller distribution, which is of course relevant for practical testing of this H_0.

$$t_{DF} \xrightarrow[T \to \infty.]{L} \frac{\frac{1}{2}(X-1)}{\sqrt{\int_0^1 [W(r)]^2 \, dr}} \tag{18}$$

The Dickey Fuller(DF) distribution V

- 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 useful, 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 critical values of the DF distribution (18) have been tabulated by Monte-Carlo simulation.
- There is however not a single table, but several, since the DF-distribution depends on whether a constant term, or a trend is included in the estimated model.
- See the mentioned paper by Ericsson and MacKinnon (2002).
- PcGive uses the relevant critical values, given the specification of the model.
- The "rule of thumb" is that Type-I error probability is best controlled by over-representing the deterministic terms, rather than under-representing them.

Dickey-Fuller tables and models II

- If a time plot of Y_t shows long-swings around a constant mean, the Dickey-Fuller regression model that we use for testing should still include a deterministic trend.
- If we reject the unit-root, we can test whether the trend is significant by a standard (t-test) conditional on stationarity.
- The cost of this procedure is the Type-II error probability can become large.

Augmented Dickey-Fuller tests I

Let the Data Generating Process (DGP) be the $AR(p)$

$$Y_t - \sum_{i=1}^{p} \phi_i Y_{t-i} = \varepsilon_t \tag{19}$$

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

$$\Delta Y_t = \sum_{i=1}^{p-1} \phi_i^{\ddagger} \Delta Y_{t-i} \; - (1 - \phi(1)) Y_{t-1} + \varepsilon_t \tag{20}$$

$Y_t \sim I(1)$ 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

$$\Delta Y_t = \sum_{i=1}^{k-1} b_i \Delta Y_{t-i} + (\rho - 1) Y_{t-1} + u_t \qquad (21)$$

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
- DM 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.
- 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.