ECON 4160, Spring term 2013. Lecture 11

Non-stationarity and co-integration 1/2

Ragnar Nymoen

Department of Economics

6 Nov 2013

Introduction I

Main references:

- Davidson and MacKinnon Ch 14
- Refer also back Ch 2.4 and Ch 2.5 in Davidson and MacKinnon for the Frisch-Waugh-Lovell theorem ,which we often make reference to.

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

$$Y_t = \phi_0 + \delta t + \varepsilon_t,\ \delta \neq 0 \quad (1)$$

where ε_t is white-noise and Gaussian.

Y_t is non-stationary, since

$$E(Y_t) = \phi_0 + \delta t \quad (2)$$

even if the variance does not depend on time:

$$Var(Y_t) = \sigma^2 \quad (3)$$

Deterministic trend—trend stationarity II

- 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:

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

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

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

with

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

Deterministic trend—trend stationarity III

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).

- In the pure DT model, non-stationarity is purged 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.

Estimation and inference in the deterministic trend model I

- Since the deterministic trend model can be placed within the VAR class of models, it represents no new problems of estimation.
- Still, the precise statistical analysis is not trivial, as E5101 will show.
- Can mention 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 standard speed of convergence for stationary variables is $T^{1/2}$
- $\hat{\delta}$ is said to be **super-consistent**. It implies that

$$plim(\hat{\delta}) = \delta \text{ and } plim(T \cdot \hat{\delta}) = \delta. \tag{4}$$

- However, 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 stationary VAR case.

AR model with trend I

A slightly model general DT model:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \delta t + \varepsilon_t,\ |\phi_1| < 1,\ \delta \neq 0 \qquad (5)$$

Conditional on $Y_0 = 0$, the solution is

$$\begin{aligned} Y_t = {} & \phi_0 \sum_{j=0}^{t} \phi_1^j - \delta \sum_{j=1}^{t-1} \left(\phi_1^2\right)^j j \qquad (6) \\ & + \delta \left(\sum_{j=1}^{t} \phi_1^{j-1} \right) \cdot t + \sum_{j=0}^{t} \phi_1^j \varepsilon_t \end{aligned}$$

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:

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

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

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

OLS estimation of models with deterministic trend I

- Above, saw that $Y_t \sim AR(1) + \textit{trend}$ can be transformed to $Y_t^s \sim AR(1)$.
- But how can we estimate $(\phi_0, \phi_1, \delta)'$?
- With reference to Frisch-Waugh-Lovell theorem, $\hat{\phi}_1$from OLS on

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \delta t + \varepsilon_t \tag{7}$$

  is identical to the regression coefficient in the regression between the de-trended residuals for Y_t and Y_{t-1}.
- Most practical to obtain OLS estimates for $(\phi_0, \phi_1, \delta)'$ in *one* step: By estimation of (7).

OLS estimation of models with deterministic trend II

- This extends to $Y_t \sim AR(p) + trend$ and to $Y_t \sim ARDL(p, p) + trend$
- Distribution of OLS estimators:
- For the pure DT model we saw that $\hat{\delta}$ is super-consistent (converge at rate $T^{3/2}$).
- In the $AR(p) + trend$ model 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 III

- 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 variables that converge at different rates.
  - In such situations the slowest convergence rates dominates, it is $\sqrt{T}$.
- The practical implication is that the stationary "asymptotic distribution theory" can be used also for dynamic models that include a DT
- For the $AR(p) + \textit{trend}$ or $ARDL(p, p) + \textit{trend}$ the *conditional* means and variances of course depend on time, just as in the model without trend: Adds flexibility to pure DT model.

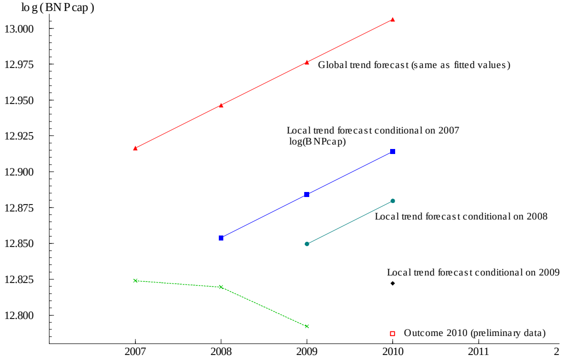
Other forms of deterministic non-stationarity I

- In one important sense, the model with DT is just 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 at a model with constant unconditional mean (with reference to the FWL theorem), this type of non-stationarity has no consequence for the statistical analysis of the model.
- But for forecasts (structural breaks in the forecat period).

Forecasting with pure DT model



Forecasting GNP per capita

- by DT

and

- local trend model

Back-casting may also give strange results!

Døm var ikke høyere enn ei tomflaske

- In 1975 Norwegian recruits averaged 179 cm, and the increase was 0.8 mm a year
- Back-casting with DT model shows that the feared Vikings were no more than 30 cm high!

See Aukrust (1977) for this piece of research!

Stochastic (or local) trend I

$AR(p)$:

$$Y_t = \phi_0 + \phi(L)Y_{t-1} + \varepsilon_t \tag{8}$$

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

Re-writing the model in the (now) usual way:

$$\Delta y_t = \phi_0 + \phi^{\ddagger}(L)\Delta y_{t-1} - \underbrace{(1-\phi(1))}_{=p(1)} y_{t-1} + \varepsilon_t \tag{9}$$

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

$$\phi^{\ddagger}(L) = \phi_1^{\ddagger} L + \phi_2^{\ddagger} L^2 + \ldots + \phi_{p-1}^{\ddagger} L^{p-1} \tag{10}$$

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

$$p(\lambda) = \lambda^p - \phi_1 \lambda^{p-1} - \ldots - \phi_p \lambda \tag{11}$$

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

$$p(1) = 1 - \phi(1) = 0. \tag{12}$$

and (9) becomes

$$\Delta Y_t = \phi_0 + \sum_{i=1}^{p-1} \phi_i^{\ddagger} \Delta Y_{t-i} \ + \varepsilon_t. \tag{13}$$

Stochastic (or local) trend III

Definition

Y_t given by (8) 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 $\Delta Y_t \sim I(1)$.
  - An integrated series y_t is called *difference stationary*.
- With reference to our earlier discussion of stationarity 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.
  - The real root 1 to a root at the the so called *long-run frequency* (E 5101).

Stochastic (or local) trend IV

- In the following, we will abstract from unit roots at the seasonal or business cycle frequencies.
- It implies that $Y_t \sim I(1)$ series are dominated by one very long cycle.
- Can however mention that the analysis can be extended to of variables that are integrated 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 (13):

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

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"	$\rightarrow 0$
4 Forecasting Y_{T+h}	$E(Y_{T+h\|T})$ depends on $Y_T\ \forall h$	$\underset{h\rightarrow\infty}{\rightarrow} E(Y_t)$
4 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:

$$Y_t = \phi_0 + Y_{t-1} + \varepsilon_t, \tag{14}$$

in fact we will show with this in a seminar exercise.

Contrasting I(0) and I(1) II

- # 5 follows from Spectral analysis, E 5101.
- We concentrate considering the inference aspects of models with $I(1)$ variables.
- We start by a demonstrating what turns out to be *the* fundamental problem of standard inference theory, and then make use of the (non-standard) statistical theory that makes it possible to make valid inference in the $I(1)$-case.

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

- In 1974 Clive Granger and Paul Newbold 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 use Pc Naive an 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

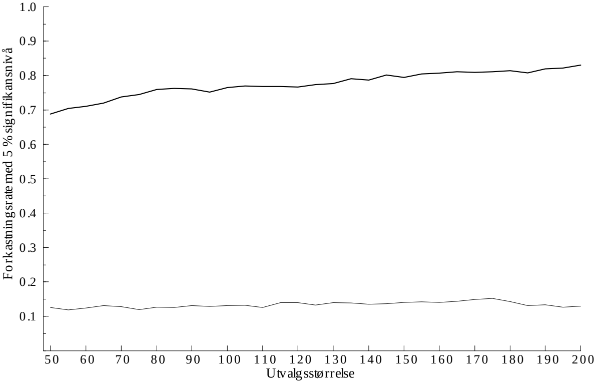
$$YA_t = \alpha + \beta YB_t + e_t$$

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

Spurious regression III

We consider the stationary GDP first, then the non-stationary DGP.

Spurious regression IV



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

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 the ECM from of the ADL:

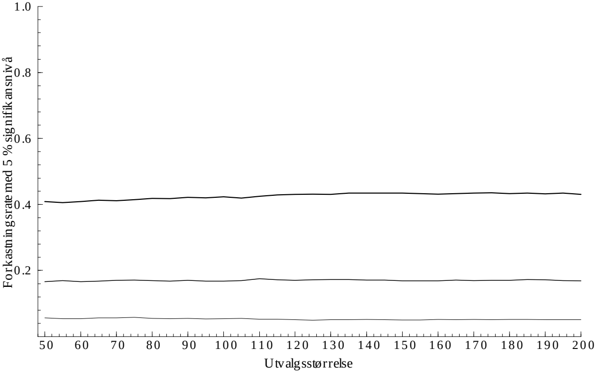
$$\Delta YA_t = \phi_0 + \rho YA_{t-1} + \beta_0 \Delta YB_t + \beta_1 YB_{t-1} + \varepsilon_{At} \quad (15)$$

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:

$$Y_t = Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2) \tag{16}$$

We estimate the model

$$Y_t = \rho Y_{t-1} + u_t, \tag{17}$$

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}.

- 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:

$$T \cdot (\hat{\rho} - 1) = \frac{\frac{1}{T} \sum_{t=1}^{T} Y_{t-1}\varepsilon_t}{\frac{1}{T^2} \sum_{t=1}^{T} Y_{t-1}^2}. \tag{18}$$

when the DGP is (16).

- The asymptotic distribution of $T \cdot (\hat{\rho} - 1)$ is

$$T \cdot (\hat{\rho} - 1) \underset{T \to \infty.}{\xrightarrow{L}} \frac{\frac{1}{2}(X - 1)}{\int_0^1 \left[W(r)\right]^2 dr} \tag{19}$$

The Dickey Fuller distribution IV

- X is distributed $\chi^2(1)$. $W(r)$, $r \in [0,1]$, is a "Standard Brownian motion".
- $\chi^2(1)$ 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 (19) negative have probability 0.68.
- 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 (19) is called an Dickey-Fuller (D-F) distribution.

The Dickey Fuller distribution V

- Under the H_0 of $\rho = 1$, also the “t-statistic” from OLS on (17) has a Dickey-Fuller distribution, which is of course relevant for practical testing of this H_0. We can refer to it at **τ-statistic** as in DM og as t_{DF} to remind us that it is a “t-statistic” but with a Dickey-Fuller distribution under the H_0 of unit-root

$$t_{DF} = \frac{\hat{\rho} - 1}{se(\hat{\rho})} \tag{20}$$

where

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

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

The Dickey Fuller distribution VI

$\hat{\sigma}^2$ is consistent since $\hat{\rho}$ is consistent.

- "Written out", t_{DF} is:

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

- It can be shown that

$$t_{DF} \underset{T\rightarrow\infty.}{\overset{L}{\longrightarrow}} \frac{\frac{1}{2}(X-1)}{\sqrt{\int_0^1 \left[W(r)\right]^2 dr}} \tag{21}$$

- 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.

Dickey-Fuller tables and models I

- The distribution (21) have been tabulated by Monte-Carlo simulation, see reference on page 618 in DM
- The distribution depend on whether the DF regressions constain no constant (**nc**), constant (**c**), constant and trend (**ct**) and constant and trend and squared trend (**ctt**), cf figure 14.2 in DM.
- It is important to choose a relevant DF-regression for your data set. For example we will usually include at least a constant, which implies a liner trend under the H_0 of a unit-root, and a mean different from zero under the alternative of stationarity-

Augmented Dickey-Fuller tests I

Let the DGP be the $AR(p)$

$$Y_t - \sum_{i=1}^{p} \phi_i Y_{t-i} = \varepsilon_t \tag{22}$$

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

$$\Delta Y_t = \sum_{i=1}^{p-1} \phi_i^{\ddagger} \Delta y_{t-i} \; - (1 - \phi(1)) Y_{t-1} + \varepsilon_t \tag{23}$$

$y_t \sim I(1)$ 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

$$\Delta Y_t = \sum_{i=1}^{k-1} b_i \Delta Y_{t-i} + (\rho - 1) y_{t-1} + u_t \tag{24}$$

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
- Davidson and MacKinnon 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

Aukrust, K. (1977) *Ludvik* Helge Erichsens Forlag, Oslo

Granger C.W.J and P. Newbold (1974) Spurious Regressions in Econometrics, *Journal of Econometrics*, 2, 111-120.

Patterson, K. (2011), Unit Root Tests in Time Series. Volume 1: Key Concepts and Problems, Palgrave MacMillan.