Non-Stationary Stochastic Time Series Models

Series de Tiempo

UTDT

January 2020

Non-Stationary Series

Importance of cheking for the existence of Unit Roots:

- Standard assymptotical theory is not valid
- Possibility of Spurrious Regressions.
- Regress series of the same order of integration.

STRUCTURE OF THE LECTURE

- Definition and Properties.
- Estimation
- Testing for Unit Roots.

Time series

- There are three important types of time series which one is likely to find in financial econometrics:

Stationary, $I(0)$
Trend stationary
Non-stationary, $I(1)$

- We have focused until now on stationary processes

Trend stationary

Definition

A trend stationary variable is a variable whose mean grows around a fixed trend. This provides a classical way of describing an economic time series which grows at a constant rate. A trend-stationary series tends to evolve around a steady, upward sloping curve without big swings away from that curve. For simplicity assume that the following process.

$$y_t = \alpha + \mu t + \varepsilon_t \text{ where } \varepsilon_t \tilde{\ } N(0, \sigma^2)$$

Notice that the mean of this process varies with time but the variance is constant.

$$E(y_t) = \alpha + \mu t$$

$$V(y_t) = E(\alpha + \mu t + \varepsilon_t - (\alpha + \mu t))^2 = \sigma^2$$

Remark: $y_t^* = y_t - (\alpha + \mu t)$ then y_t^* is stationary.

A Non stationary Series: I(1) Processes

Definition

An autoregressive process of order p, AR(p), has a **unit root** *if the polynomial in L,* $(1 - \phi_1 L - ...\phi_p L^p)$ *has a root equal to one. The simplest example of a process with a unit root is a random walk, i.e.,*

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

where ε_t *is i.i.d. with zero mean and constant variance.*

- We can easily see that the variance of this processes does not exist: lagging the process one period we can write

$$y_{t-1} = y_{t-2} + \varepsilon_{t-1},$$

and substituting back we get

$$y_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t.$$

Then, repeating this procedure we can easily show that

$$y_t = y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t$$

- The mean can be calculated assuming that y_0 is fixed, then the mean is constant over time

$$E(y_t) = E(y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t) = y_0$$

- The variance of y_t, "conditional" on knowing y_0, can be computed as

$$\begin{aligned} V(y_t) &= V(y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t) \\ &= V(\varepsilon_1) + V(\varepsilon_2) + ... + V(\varepsilon_{t-1}) + V(\varepsilon_t) = t\sigma^2 \end{aligned}$$

  As we move further into the future this expression becomes infinite. We conclude that the variance of a unit root process is infinite.

Comments

- A unit root process will only cross the mean of the sample very infrequently
- A process that has a unit root is also called **integrated of order one**, denoted as $I(1)$
- A stationary process is an **integrated of order zero** process, denoted as $I(0)$
- *Why is this important?*
  - The study of econometric models with non-stationary data has been one of the most important concerns of econometricians in the last 30 years. The topic is very vast and we just will *mention* some of the most important issues.

Spurious Regressions

- Granger and Newbold (1974) have shown that, using $I(1)$, you can obtain an apparently significant regression (say with a high R^2) even if the regressor and the dependent variable are independent.
- They did this by generating independent random walk series against each other
- *Rule of thumb*: whenever you obtain a very high R^2 and a very low DW you should suspect that the result are spurious.
    - the reason for the low DW will be understood in what is to come

Regressing Series that are integrated of the same order

Caution!

You should always check that the variables you are regressing one against the other are of the same order if you are to obtain meaningful results

Testing for unit roots

- We will show that the standard t- test cannot be applied to a process with a unit root
  - under the null of $\alpha = 1$ we get a degenerated distribution
- We will find a distribution for $\widehat{\alpha}$ under this circumstances and see that:
  1. it is not t - student
  2. it is biased to the left

Testing for Unit Roots (II)

Consider the following model:

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

where ε_t is assumed to be N(0,σ^2). It can easily be shown that asymptotically

$$\sqrt{T}(\hat{\alpha}_T - \alpha) \quad \underrightarrow{L} \quad N(0, (1 - \alpha^2))$$

If we want to use this distribution for testing the null hypothesis that $\alpha = 1$, then we find that the distribution under the null "degenerates" (collapses in one point).

- To obtain a non-degenerate asymptotic distribution for $\hat{\alpha}_T$ in the unit root case, it turns out that we have to multiply by T and not by the square root of T. Then the unit root coefficient converges at a faster rate T than for the stationary case.
- To get a better sense of why scaling by T is necessary when the true value of α is unity consider the OLS estimate:

$$\hat{\alpha}_T = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}.$$

Then, substituting y_t by the AR(1) process we get that:

$$\hat{\alpha}_T - \alpha = \frac{\sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

- Multiplying in both sides by T, we get

$$T(\hat{\alpha}_T - \alpha) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}.$$

- Now, under the null that $\alpha = 1$, y_t can be written as:

$$\begin{aligned} y_t &= y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t \\ &= \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t \qquad \textit{if we assume } y_0 = 0. \end{aligned}$$

- Then under the null that $\alpha = 1$, y_t ~N$(0, \sigma^2 t)$.

- Now we proceed to find the distribution of the numerator: under the null,

$$y_t^2 = (y_{t-1} + \varepsilon_t)^2 = y_{t-1}^2 + 2y_{t-1}\varepsilon_t + \varepsilon_t^2$$

and rearranging terms we obtain

$$y_{t-1}\varepsilon_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \varepsilon_t^2).$$

- Then, the sum which appears in the *numerator* can be expressed as

$$\sum_{t=1}^{T} y_{t-1}\varepsilon_t = \sum_{t=1}^{T} \frac{1}{2}(y_t^2 - y_{t-1}^2 - \varepsilon_t^2) = \frac{1}{2}(y_T^2 - y_0^2) - \sum_{t=1}^{T} \frac{1}{2}\varepsilon_t^2.$$

- Then, recalling that $y_0 = 0$ and multiplying by $\left(T^{-1}\right)$ we obtain the expression of the *numerator* as the sum of two terms

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t = (\frac{1}{2T})y_T^2 - \sum_{t=1}^{T}(\frac{1}{2T})\varepsilon_t^2$$

- To find the distribution of this expression we divide each side by σ^2 which yields the following result

$$(\sigma^2 T)^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t = (1/2)(\frac{y_T}{\sigma\sqrt{T}})^2 - \sum_{t=1}^{T}(\frac{1}{2\sigma^2 T})\varepsilon_t^2$$

- Consider the first term of this expression. Since we have shown above that y_t ˜N$(0, \sigma^2 t)$, standardizing we obtain

$$(y_T/\sigma\sqrt{T})\tilde{\ }N(0,1),$$

- and then squaring this expression we find that the first term of the numerator is distributed Chi-square

$$(y_T/\sigma\sqrt{T})^2\tilde{\ }\chi^2(1).$$

- It can be shown using the law of large numbers that the second term converges in probability to σ^2, i.e.

$$(1/T)\sum_{t=1}^{T}\varepsilon_t^2 \quad \underrightarrow{P} \quad \sigma^2, \qquad or \qquad (1/\sigma^2 T)\sum_{t=1}^{T}\varepsilon_t^2 \quad \underrightarrow{P} \quad 1$$

- If we put both results together we can see that the numerator converges to

$$(\sigma^2 T)^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \qquad \underrightarrow{L} \qquad (1/2)(X-1) \quad where \quad X\tilde{}\chi^2(1).$$

- It can also be shown using the law of large numbers that the *denominator* converges in probability to

$$E(T^{-2}\sum_{t=1}^{T} y_{t-1}^2).$$

- Now,as y_{t-1} ~N$(0, \sigma^2(t-1))$, then $E(y_{t-1}^2) = \sigma^2(t-1)$. Therefore the expected value of the denominator can be written as

$$\begin{aligned} E(T^{-2}\sum_{t=1}^{T} y_{t-1}^2) &= T^{-2}\sum_{t=1}^{T} E(y_{t-1}^2) \\ &= T^{-2}\sigma^2 \sum_{t=1}^{T}(t-1) \\ &= \sigma^2 T^{-2}(T-1)T/2 \end{aligned}$$

- Then if we multiply $(\hat{\alpha}_T - 1)$ by T instead than by $\sqrt{T}$, we obtain a non-degenerate asymptotic distribution, but this distribution is not Gaussian.

How do we test for unit roots? Dicky Fuller Test

Consider the following model

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

where ε_t "is assumed" to be N(0,σ^2)

We want to test the Hypothesis of the existence of a unit root therefore we set the following null and alternative hypothesis.

$$\begin{aligned} H_0) \quad \alpha &= 1(UR) \\ H_1) \quad \alpha &< 1(I(0)) \end{aligned}$$

The obvious estimator of is the OLS estimator, $\hat{\beta}$. The problem is that under the null hypothesis there is considerable evidence of the non - adequacy of the asymptotic (approximate in large samples) distribution. Therefore
The latter equation can be rewritten as

$$\Delta y_t = \mu + (\alpha - 1) y_{t-1} + \beta t + \varepsilon_t$$

or

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

For this expression the relevant hypothesis should be written as:

$$\begin{aligned} H_0)\ \lambda &= 0\ (\textit{unit root}) \\ H_1)\ \lambda &< 0\ (I(0)) \end{aligned}$$

Fuller (1976) tabulated, using Monte Carlo methods, critical values for alternative cases, for example for a sample size of 100 the 5 % critical values are

$\mu = 0,\ \beta = 0$	-2.24
$\mu \neq 0,\ \beta = 0$	-3.17
$\mu \neq 0,\ \beta \neq 0$	-3.73

Therefore the method simply consist to check the t-statistic of $\hat{\lambda}$ against the critical values of Fuller (1976). Notice that the critical values depend on

i) the sample size

ii) whether you include a constant and/or a time trend.

This procedure is only valid when there is no evidence of serial correlation in the residuals, $\hat{\varepsilon}_t$ .If there is serial correlation you should need to include additional lags, say $\Delta y_{t-1}, \Delta y_{t-2}$,... until it disappears

Augmented Dickey Fuller

Consider:

$$\Delta y_t = \mu + \lambda y_{t-1} + \beta t + \alpha_1 \Delta y_{t-1} + ... + \alpha_k \Delta y_{t-k} + \varepsilon_t$$

- We chose to augment the regression with k lags. This is usually denoted as ADF(k).
- Choosing the order of augmentation of the DF regression:
  - *choosing k as a function of the number of observations as in Schwert (1989)*

    $$k = INT(12(T/100)^{1/12})$$

  - *information based rules such as AIC and BIC.*
  - *sequential rules*
  - General to specific seems to be preferable to the other methods.

Small sample properties of the Dickey Fuller test

- The power of the test is extremely low
- ADF comes up with conflicting results depending on the order chosen for the regression
- Good practice: start with quite a general model and then delete lags

Phillips - Perron type tests for unit roots

- Alternative approach to DF
- They make a non-parametric correction to the standard deviation which provides a consistent estimator of the variance.

$$S_{Tl}^2 = T^{-1} \sum_{t=1}^{T} (\varepsilon_t^2) + 2T^{-1} \sum_{t=1}^{l} \sum_{t=j+1}^{T} \varepsilon_t \varepsilon_{t-j}$$

$$S_\varepsilon^2 = T^{-1} \sum_{t=1}^{T} (\varepsilon_t^2)$$

where $_l$ is the lag truncation parameter used to ensure that the autocorrelation of the residuals is fully captured.

An asymptotically valid test $\phi = 1$, for

$$\Delta y_t = \mu + (\phi - 1)y_{t-1} + \varepsilon_t \quad \varepsilon_t \tilde{} iiD(0, \sigma^2)$$

when the underlying DGP is not necessarily an AR(1) process, is given by the Phillips Z-test.

$$Z(\tau_\mu) = (S_\varepsilon / S_{Tl})\tau_\mu - (1/2)(S_{Tl}^2 - S_\varepsilon^2)[S_{Tl}[T^2 \sum_{t=2}^{T}(y_t - \overline{y})^2]^{.5}]^{-1}$$

where τ_μ is the t-statistic associated with testing the null hypothesis $\rho = 1$.

Comments The critical values for this test statistic are the same as those used for the same case in the fuller table.
Monte Carlo work suggests that the Phillips-type test has poor size properties (tendency to over reject when is true) when the underlying DGP has large negative MA components.

Structural breaks and unit roots

1. Perron (1988) has shown that an I(0) process with a structural break in the mean will be difficult to distinguish from a I(1) process.
   - If we know where the break takes place, the natural thing to do, is to partial out the break by using dummy variables and test for unit roots once the break has been partialled out.
   - A possible solution to try to identify these breaks is to perform the *ADF* test recursively and to compute recursive t-statistics.

Recursive t-statistics

- The recursive ADF -statistic is computed using sub samples $t = 1..k$ *for* $k = k_0, ...T$, where k is the start up value and T is the sample size of the full sample.
- The most general model (with drift and trend) is estimated for each sub sample
- The minimum value of $\tau_\tau(k/T)$ across all the sub samples is chosen and compared with the table provided by Banjeree, Lumsdaine and Stock

Rolling ADF

- The previous method could also be applied using a (large enough) window (of size k) to see if there are clear changes in the pattern of a series.
- The most general model (with drift and trend) is estimated for each sub sample
- The minimum value of $\tau_\tau(k/T)$ across all the sub samples is chosen and compared with the table found in the class notes.

Tests with stationarity as a null: KPSS test

Consider the following model.

$$y_t = \alpha + \delta t + \xi_t + \varepsilon_t$$

where ε_t is a stationary process and ξ_t is a random walk given by

$$\xi_t = \xi_{t-1} + u_t \quad u_t \tilde{} iid(0, \sigma_u^2)$$

The null of stationarity is formulated as

$$H_0)\sigma_u^2 = 0$$

The test statistic for this hypothesis is given by

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_e^2}$$

where e_t are the residuals of a regression of y_t on a constant and a time trend, $\hat{\sigma}_e^2$ is the residual variance for this regression and S_t is the partial sum of e_t defined by

$$S_t = \sum_{i=t}^{T} e_i \quad t = 1, 2, .., T.$$

Variance Ratio Tests

- This will provide us with another tool to discriminate between (trend) stationary and non-stationary series.
- Consider y_t and assume that it follows follows a random walk, i.e.

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

then by iterative substitution we know that

$$y_t = y_{t-k} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + ... + \varepsilon_{t-k+1}$$

- Denoting the difference between y_t and y_{t-k} as $\Delta_k y_t$, then

$$\Delta_k y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + ... + \varepsilon_{t-k+1}$$

  and clearly the variance of $\Delta_k y_t$, is $\sigma^2 k$, where σ^2 is the variance of ε.
- We can define a "variance ratio" function (a function of k) as

$$\lambda_1(k) = \frac{Var(\Delta_k y_t)}{Var(\Delta_1 y_t)} = k.$$

- Therefore a plot of λ_1 against k should be an increasing straight line

- Alternatively we may define a new function $\lambda_2(k)$ as

$$\lambda_2(k) = \lambda_1(k)/k$$

- If there is a unit root, $\lambda_2(k)$ tends to one when k tends to infinite.
- However if y_t *does not* contain a unit root it can be shown that the lim $\lambda_2(k)$ when k tends to infinite is equal to zero.

Proof.

(idea)Assume the following *AR(1)* process

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \qquad t = 1, ..., T$$

□

we have seen that by iterative substitution we can express this process as

$$y_t = \phi_1^k y_{t-k} + \phi_1^{k-1}\varepsilon_{t-(k-1)} + \phi_1^{k-2}\varepsilon_{t-(k-2)} + ... + \phi_1\varepsilon_{t-1} + \varepsilon_t$$

Then subtracting y_{t-k} in both sides of the equation we get the following expression

$$y_t - y_{t-k} = (\phi_1^k - 1)y_{t-k} + \phi_1^{k-1}\varepsilon_{t-(k-1)} + \phi_1^{k-2}\varepsilon_{t-(k-2)} + ... + \phi_1\varepsilon_{t-1} + \varepsilon_t$$

Proof.

(continues)the variance of $y_t - y_{t-k}$, $Var(\Delta_k y_t)$ is equal to □

$$V(y_t - y_{t-k}) = (\phi_1^k - 1)^2 V(y_{t-k}) + V(\sum_{j=0}^{k-1} \phi_1^j \varepsilon_{t-j})$$

Notice that

$$V(y_{t-k}) = V(y_t) = (1/(1-\phi_1^2))\sigma^2$$

and

$$V(\sum_{j=0}^{k-1} \phi_1^j \varepsilon_{t-j}) = ((1-\phi_1^{2k})/(1-\phi_1^2))\sigma^2,$$

Proof.

We can write the variance of Δy_{t-k} as

$$V(y_t - y_{t-k}) = (\phi_1^k - 1)^2 (1/(1-\phi_1^2))\sigma^2 + ((1-\phi_1^{2k})/(1-\phi_1^2))\sigma^2$$

In the same way we can express for a stationary process the variance of the first difference of y_t, $(y_t - y_{t-1}) = (\phi_1 - 1)y_{t-1} + \varepsilon_t$.

$$V(y_t - y_{t-1}) = (\phi_1 - 1)^2 V(y_{t-1}) + \sigma^2 = (\phi_1 - 1)(1/(1-\phi_1^2))\sigma^2 + \sigma^2$$

then the variance ratio can be written as

$$\lambda_1(k) = \frac{(\phi_1^k - 1)^2 (1/(1-\phi_1^2))\sigma^2 + (\ (1-\phi_1^{2k})/(1-\phi_1^2)\ )\sigma^2}{(\phi_1 - 1)^2 (1/(1-\phi_1^2))\sigma^2 + \sigma^2}$$

□

Proof.

Then the limit of $\lambda_1(k)$ when k tends to infinite for a stationary process is

$$lim_{k \to \infty} \lambda_1(k) = \frac{1}{1-\phi_1}$$

which is a constant provided that $\phi_1 \neq 1$.
It should be clear from the previous result that the limit of $\lambda_2(k)$ equals 0 when k tends to infinite.

$\square$

Trend stationary and difference stationary processes

A trend stationary variable may be written as

$$y_t = \alpha + \mu t + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + ...$$

Then y_{t-k} is simply

$$y_{t-k} = \alpha + \mu(t-k) + \varepsilon_{t-k} + \theta_1 \varepsilon_{t-1-k} + \theta_2 \varepsilon_{t-2-k} + \theta_3 \varepsilon_{t-3-k} + ...$$

The k^{th} difference can be obtained simply by subtracting the two above equations

$$
\begin{aligned}
y_t - y_{t-k} &= \mu k + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_{k-1}\varepsilon_{t-(k-1)} + (\theta_k - 1)\varepsilon_{t-k} + \\
&\quad (\theta_{k+1} - \theta_1)\varepsilon_{t-(K-1)} + ... + (\theta_{k+q} - \theta_q)\varepsilon_{t-q} + ... \\
&= \mu k + \sum_{j=0}^{k-1} \theta_j \varepsilon_{t-j} + \sum_{j=0}^{\infty} (\theta_{k+j} - \theta_j)^2 \varepsilon_{t-j}
\end{aligned}
$$

Then the variance of $y_t - y_{t-k}$ may be written as

$$V(y_t - y_{t-k}) = \sigma^2 \sum_{j=0}^{k-1} \theta_j^2 + \sigma^2 \sum_{j=0}^{\infty} (\theta_{k+j} - \theta_j)^2$$

From the previous equation we can see that when k tends to infinity, the variance of $\Delta_k y_t$ is equal to

$$V(\Delta_k y_t) = 2\sigma^2 \sum_{j=0}^{k-1} \theta_j^2$$

Now the first difference of a trend stationary process, Δy_t is

$$\begin{aligned} y_t - y_{t-1} &= \mu + \varepsilon_t + (\theta_1 - 1)\varepsilon_{t-1} + ... + (\theta_{k+1} - \theta_k)\varepsilon_{t-(k+1)} \\ &\quad + ... + (\theta_{k+q} - \theta_{k+q-1})\varepsilon_{t-q} + ... \end{aligned}$$

Then the variance of the first difference can be written as

$$\begin{aligned} V(y_t - y_{t-1}) &= V(\varepsilon_t + (\theta_1 - 1)\varepsilon_{t-1} + ... + (\theta_{k+1} - \theta_k)\varepsilon_{t-(k+1)} \\ &\quad + ... + (\theta_{k+q} - \theta_{k+q-1})\varepsilon_{t-(k+q)} + ...) \\ &= \sigma^2(1 + \sum_{j=0}^{\infty}(\theta_{j+1} - \theta_j)^2) \end{aligned}$$

The variance ratio should be

$$\lambda_1(k) = \frac{Var(\Delta_k y_t)}{Var(\Delta_1 y_t)} = \frac{\sum_{j=0}^{k-1}\theta_j^2 + \sigma^2\sum_{j=0}^{\infty}(\theta_{k+j} - \theta_j)^2}{(1 + \sum_{j=0}^{\infty}(\theta_{j+1} - \theta_j)^2)},$$

and the limit when k tends to infinity is

$$\lim_{k\to\to\infty}\lambda_1(k) = \lim_{k\to\to\infty}\frac{Var(\Delta_k y_t)}{Var(\Delta_1 y_t)} = \lim_{k\to\to\infty}\frac{2\sum_{j=0}^{k-1}\theta_j^2}{(1 + \sum_{j=0}^{\infty}(\theta_{j+1} - \theta_j)^2)}.$$

- The expression is a constant and might be greater or smaller than one depending on the θ_j values.
- We can distinguish between the two models by simply noting that under the random walk assumption λ_1 increases with k and that under the trend stationary assumption λ_1 tends to a constant.
- Alternatively we can consider λ_2 $Var(\Delta_k y_t)/k$. and note both, that when the model is a random walk this expression tends to 1 (see proof above), and that this ratio should tend to zero when k tends to infinity since $Var(\Delta_k y_t)$ is constant for the trend stationary model.

- **Sampling distribution of λ(k) under the Random Walk Hypothesis.**

$$H_0) \; \alpha = 1 \; or \quad y_t = \mu + y_{t-1} + \varepsilon_t \qquad\qquad \varepsilon_t \tilde{} \; IIDN(0,)$$

It can be shown that asymptotically under the null, $\sqrt{Tk}(\hat{\lambda}_2(k) - 1) \overset{a}{\sim} N(0, 2(k-1))$. Then tests of the null Hypothesis can be carried out on the standardized statistics.