Linear models and time series

Gabriela Ciołek

April 5, 2019

Unit roots

- A unit root (also called a unit root process or a difference stationary process) is a stochastic trend in a time series, sometimes called a random walk with drift
- A linear stochastic process has a unit root, if 1 is a root of the process's characteristic equation. Such a process is non-stationary but does not always have a trend.
- If a time series has a unit root, it shows a systematic pattern that is unpredictable

- Consider a discrete-time stochastic process $\{y_t, t = 1, \ldots, \infty\}$ and suppose that it can be written as an autoregressive process of order p:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \varepsilon_t.$$

- Here, $\{\varepsilon_t, t = 0, \ldots, \infty\}$ is a serially uncorrelated, zero-mean stochastic process with constant variance σ^2.
- For convenience, assume $y_0 = 0$. If $m = 1$ is a root of the characteristic equation:

$$m^p - m^{p-1}a_1 - m^{p-2}a_2 - \cdots - a_p = 0$$

  then the stochastic process has a unit root or, alternatively, is integrated of order one, denoted $I(1)$.
- If m $= 1$ is a root of multiplicity r, then the stochastic process is integrated of order r, denoted I(r).

Unit roots-Example AR(1) model

The first order autoregressive model,

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

has a unit root when $a_1 = 1$. In this example, the characteristic equation is $m - a_1 = 0$. The root of the equation is $m = 1$.
If the process has a unit root, then it is a non-stationary time series. That is, the moments of the stochastic process depend on t. To illustrate the effect of a unit root, we can consider the first order case, starting from $y0 = 0$:

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

By repeated substitution, we can write

$$y_t = y_0 + \sum_{j=1}^{t} \varepsilon_j.$$

Example AR(1) model

The variance of y_t is given by:

$$\mathrm{Var}(y_t) = \sum_{j=1}^{t} \sigma^2 = t\sigma^2.$$

The variance depends on t since $\mathrm{Var}(y_1) = \sigma^2$, while $\mathrm{Var}(y_2) = 2\sigma^2$. Note that the variance of the series is diverging to infinity with t.
If $a_1 = 1$ we obtain a whole class of highly persistent time series processes that also have linearly trending means.

Unit roots

- Using time series with strong persistence of the type displayed by a unit root process in a regression equation can lead to very misleading results if the CLM assumptions are violated.
- Fortunately, simple transformations are available that render a unit root process weakly dependent
- Weakly dependent processes are said to be integrated of order zero, [I(0)]. Practically, this means that nothing needs to be done to such series before using them in regression analysis: averages of such sequences already satisfy the standard limit theorems

Unit roots

- Unit root processes, such as a random walk (with or without drift), are said to be integrated of order zero, or I(0). This means that the first difference of the process is weakly dependent (and often stationary).
- This is simple to see for a random walk. With $y_t = y_{t-1} + e_t,\ t = 1, 2, \cdots$, for $t = 1, 2, \cdots$. We consider first-differenced series:

$$\Delta y_t = y_t - y_{t-1} = e_t,\ t = 2, 3, \cdots$$

which is actually an i.i.d. sequence. Thus, when we suspect processes are integrated of order one, we often first difference in order to use them in regression analysis

Unit root tests

- The Dickey Fuller Test (sometimes called a Dickey Pantula test), which is based on linear regression. Serial correlation can be an issue, in which case the Augmented Dickey-Fuller (ADF) test can be used. The ADF handles bigger, more complex models. It does have the downside of a fairly high Type I error rate.
- The Elliott-Rothenberg-Stock Test
- The P-test takes the error terms serial correlation into account,
- The DF-GLS test can be applied to detrended data without intercept.
- The Phillips-Perron (PP) Test is a modification of the Dickey Fuller test, and corrects for autocorrelation and heteroscedasticity in the errors.
- The Zivot-Andrews test allows a break at an unknown point in the intercept or linear trend.

Dickey-Fuller test

In statistics, the Dickey-Fuller test tests the null hypothesis that a unit root is present in an autoregressive model. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity.

Serial correlation (serial correlation is the relationship between a given variable and itself over various time intervals) can be an issue, in which case the Augmented Dickey-Fuller (ADF) test can be used. The ADF handles bigger, more complex models. The simplest approach is to start with a model AR(1):

$$y_t = \alpha + \rho y_{t-1} + e_t, \tag{1}$$

where e_t is a zero mean process, i.e. $y = \mathbb{E}[e_t | y_{t-1}, \cdots, y_0] = 0$. Thus y has a unit root if and only if $\rho = 1$.

We will test $H_0 : \rho = 1$ versus $H_1 : \rho < 1$. The alternative $H_1 : \rho > 1$ is not considered since it would mean that the series y Would be explosive. When $|\rho| < 1$, theb y is a stable $AR(1)$, that is, of weak dependence or asymptotically correlated $Corr(y_t, y_{t+h}) = \rho^h \to 0$ when $|\rho| < 1$. In practice we consider the equation obtained by removing y_{t-1} from both sides of the preceding equation:

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t. \tag{2}$$

We set $\theta = \rho - 1$. Under H_0 y_{t-1} is I(1) (stochastic process has a unit root). We test $H_0 : \theta = 0$. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity.

Augmented Dickey-Fuller test

In statistics and econometrics, an augmented Dickey-Fuller test (ADF) tests the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. It is an augmented version of the Dickey-Fuller test for a larger and more complicated set of time series models. We will also test the Augmented Dickey-Fuller regression model with additional delays for the model AR(p):

$$\Delta y_t = \alpha + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + e_t.$$

Under $H_0 : \theta = 0$ Δy_t is an autoregressive stable model of order p. We can then use the t-test on θ as before, the critical values are the same, and this test is called the augmented Dickey-Fuller test. The intuition behind the test is that if the series is integrated then the lagged level of the series y_{t-1} will provide no relevant information in predicting the change in y_t besides the one obtained in the lagged changes Δy_{t-k}. In this case the $\gamma = 0$ and null hypothesis is not rejected.

Stationarity

A common assumption in many time series techniques is that the data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations

The Ljung-Box test

The Ljung-Box test (named for Greta M. Ljung and George E. P. Box) is a type of statistical test of whether any of a group of autocorrelations of a time series are different from zero. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags, and is therefore a portmanteau test.
The Ljung-Box test may be defined as:
H_0 : The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).
H_1 : The data are not independently distributed; they exhibit serial correlation.

Model selection: AIC and BIC criteria

- in statistics, the Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the Akaike information criterion (AIC).
- When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. Both BIC and AIC attempt to resolve this problem by introducing a penalty term for the number of parameters in the model; the penalty term is larger in BIC than in AIC.

AIC criterium

A formula for least squares regression type analyses for normally distributed errors:

$$AIC = nlog(\widehat{\sigma^2}) + 2K$$

Where:
$\widehat{\sigma^2}$ = Residual Sum of Squares/n,
n = sample size,
K is the number of model parameters.

Testing change of structure

The Chow test is a test of whether the true coefficients in two linear regressions on different data sets are equal. In econometrics, it is most commonly used in time series analysis to test for the presence of a structural break at a period which can be assumed to be known a priori (for instance, a major historical event such as a war). The Chow test is often used to determine whether the independent variables have different impacts on different subgroups of the population

Chow test

Suppose that we model our data as

$$y_t = a + bx_{1t} + cx_{2t} + \varepsilon.$$

If we split our data into two groups, then we have

$$y_t = a_1 + b_1 x_{1t} + c_1 x_{2t} + \varepsilon$$

and

$$y_t = a_2 + b_2 x_{1t} + c_2 x_{2t} + \varepsilon.$$

The null hypothesis of the Chow test asserts that $a_1 = a_2$, $b_1 = b_2$, and $c_1 = c_2$ and there is the assumption that the model errors ε are independent and identically distributed from a normal distribution with unknown variance.

Test of the change of structure

In case when the breaking point is not known, a QLR test can be performed. It is in fact the maximum F-statistic $F(\tau)$ of the Chow test on a certain sample of breaking points:

$$QLR = \max_{\tau \in \{\tau_0, \tau_0+1, \cdots, \tau_1-1, \tau_1\}} F(\tau).$$

A conventional choice for τ_0 and τ_1 are the inner 70% of the sample (exclude the first and last 15%).

Exercise 1

Load the data phillips.raw
Attention for the missing values

- plot(inf)
- Plot the ACF and the PACF
- Perform Ljung-Box test
- Fit the AR(p) model and use the AIC criterium

$$load\ phillips.raw$$

$$y = phillips(:, 3);$$

$$plot(y)$$

Download the qstat2 function from page: *https://ideas.repec.org/c/boc/bocode/t961403.html* (copy the url and download the script qstat2.m). This function takes two parametres: the vector of time series and order of AR(p) model that we want to test.

$$[qstat, pval] = qstat2(y, 1)$$

Do you reject the hypothesis of the lack of correlation based on obtained p-value?

AIC takes into account the residual or unexplained variance of the model, its complexity and the number of observations used. AIC minimization is a criterion for choosing one model over another.
Test AR(1):

$$y\_ = y(2:n)$$

$$y\_lag = y(1:n-1)$$

$$X = y\_lag$$

$$[n,k] = size(X)$$

$$beta = inv(X' * X) * X' * y\_$$

$$u = y\_ - X * beta;$$

$$sig2 = u' * u/(n-k)$$

$$AIC1 = log(sig2) + (2. * (1))./n;$$

Test AR(2):

$$y\_ = y(3:n)$$

$$y\_lag = y(2:n-1)$$

$$y\_lag2 = y(1:n-2)$$

$$X = [y\_lag y\_lag2]$$

$$[n,k] = size(X)$$

$$beta = inv(X' * X) * X' * y\_$$

$$u = y\_ - X * beta;$$

$$sig2 = u' * u/(n-k)$$

$$AIC2 = log(sig2) + (2. * (2))./n;$$

Exercise 2

Stationarity testing:

- Divide the sample into two and three parts
- Calculate the means and variances
- Do the unit root test:DF, DF augmented with 4 delays

$$y1 = y(1 : floor(n/2))$$
$$y2 = y(floor(n/2) : n)$$
$$m1 = mean(y1)$$
$$m2 = mean(y2)$$
$$v1 = var(y1)$$
$$v2 = var(y2)$$
$$y1 = y(1 : floor(n/3))$$
$$y2 = y(floor(n/3) : floor(2 * n/3))$$
$$y3 = y(floor(2 * n/3) : n)$$
$$m1 = mean(y1)$$
$$m2 = mean(y2)$$
$$m3 = mean(y3)$$
$$v1 = var(y1)$$
$$v2 = var(y2)$$
$$v3 = var(y3).$$

The simplest approach is to start with a model AR(1):

$$y_t = \alpha + \rho y_{t-1} + e_t, \tag{3}$$

e will test $H_0 : \rho = 1$ versus $H_1 : \rho < 1$. The alternative $H_1 : \rho > 1$ is not considered since it would mean that the series y Would be explosive. When $|\rho| < 1$, theb y is a stable $AR(1)$, that is, of weak dependence or asymptotically correlated $Corr(y_t, y_{t+h}) = \rho^h \to 0$ when $|\rho| < 1$. In practice we consider the equation obtained by removing y_{t-1} from both sides of the preceding equation:

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t. \tag{4}$$

We set $\theta = \rho - 1$. Under H_0 y_{t-1} is I(1) (stochastic process has a unit root) and the asymptotic properties of the t-statistic change: the asymptotic distribution of t-statistic is known as the Dickey-Fuller distribution. So we will no longer use "usual" critical values, but Dickey-Fuller's, and at that point we use a Dickey-Fuller test. We test $H_0 : \theta = 0$. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity.

$$y\_ = y(2:n)$$

$$y\_lag = y(1:n-1)$$

$$Delta\_y = y\_ - y\_lag$$

$$[n,k] = size(Delta\_y)$$

$$X = [ones(n,1)y\_lag]$$

$$[n,k] = size(X)$$

$$beta = inv(X' * X) * X' * Delta\_y$$

$$u = Delta\_y - X * beta;$$

$$sig2 = u' * u/(n-k);$$

$$std = sqrt(diag(sig2 * inv(X' * X)))$$

$$t = beta./std$$

ADF test

$$y\_ = y(2 : n)$$

$$y\_lag = y(1 : n - 1)$$

$$Delta\_y = y\_ - y\_lag$$

$$[n, k] = size(Delta\_y)$$

$$Delta\_y\_ = Delta\_y(1 : n - 4)$$

$$Delta\_y\_lag = Delta\_y(2 : n - 3)$$

$$Delta\_y\_lag2 = Delta\_y(3 : n - 2)$$

$$Delta\_y\_lag3 = Delta\_y(4 : n - 1)$$

$$Delta\_y\_lag4 = Delta\_y(5 : n)$$

$$y\_lag\_ = y(1 : n - 4)$$

$$[n, k] = size(Delta\_y\_)$$

$$X = [ones(n,1) \;\; y\_lag\_ \;\; Delta\_y\_lag \;\; Delta\_y\_lag2 \; Delta\_y\_lag3 \;\; Delta\_y\_lag4]$$

$$[n,k] = size(X)$$

$$beta = inv(X' * X) * X' * Delta\_y\_$$

$$u = Delta\_y\_ - X * beta;$$

$$sig2 = u' * u/(n-k);$$

$$std = sqrt(diag(sig2 * inv(X' * X)))$$

$$t = beta./std$$

Exercise 3

- Perform Chow test for philips data with a change point for year 1981
- Perform QLR test with 15% trimming

We take $t_0 = 1981$.
Chow test:

$$[n, k] = size(phillips)$$

$$t0 = 34$$

$$D\_tau = [zeros(34, 1);\ ones(n - 34, 1)]$$

$$D\_tau = D_t au(1 : n - 1)$$

$$y\_ = y(2 : n)$$

$$y\_lag = y(1 : n - 1)$$

$$X = [ones(n - 1, 1)\ \ y\_lag\ D\_tau. * ones(n - 1, 1)\ D\_tau. * y\_lag]$$

$$[n, k] = size(X)$$

Unrestricted model:

$$beta0 = inv(X' * X) * X' * y\_$$

$$u0 = y\_ - X * beta0$$

$$SSR0 = u0' * u0$$

Restricted model:

$$X = X(:, [1, 2])$$

$$beta1 = inv(X' * X) * X' * y\_$$

$$u1 = y\_ - X * beta1$$

$$SSR1 = u1' * u1$$

$$F = ((SSR1 - SSR0)/SSR0) * ((n - k)/1)$$

$$p = fdis\_prb(F, 1, n - k)$$

Using 15% trimming is equivalent to take $\tau = 0.15T$ and $\tau_1 = 0.85T$.
QLR statistic

$$[n, k] = size(phillips);$$

$$tau0 = floor(0.15 * n);$$

$$tau1 = floor(0.85 * n);$$

$$all\_chows = [];$$

$$y\_ = y(2 : n)$$

$$y\_lag = y(1 : n - 1)$$

$$k = 4$$

$$for \quad t = tau0 : tau1;$$

$$D\_tau = [zeros(t, 1);\; ones(n - t, 1)]$$

$$D\_tau = D\_tau(1 : n - 1)$$

$$X = [ones(n - 1, 1)\; y\_lag D\_tau. * ones(n - 1, 1)\; D\_tau. * y\_lag]$$

$$beta0 = inv(X' * X) * X' * y\_$$

$$u0 = y\_ - X * beta0$$

$$SSR0 = u0' * u0$$

$$X = X(:, [1, 2])$$

$$beta1 = inv(X' * X) * X' * y\_$$

$$u1 = y\_ - X * beta1$$

$$SSR1 = u1' * u1$$

$$F = ((SSR1 - SSR0)/SSR0) * ((n - k)/1)$$

$$all\_chows(end + 1) = F$$

$$end;$$

$$QLR = max(all\_chows)$$

$$[QLR, tau] = max(all\_chows)$$