Time Series

tsset year

Twoway (tsline variable_name variable_name)
tsline draws line plots for time-series data.

corrgram variable_name
corrgram tabulates autocorrelations, partial autocorrelations.

ac variable_name
ac produces a correlogram (a graph of autocorrelations) with pointwise confidence intervals that is based on Bartlett's formula for MA(q) processes. By observing ac graph it is seen that it have a declining(or upward) trend it have unitroot

pac variable_name
pac produces a partial correlogram (a graph of partial autocorrelations) with confidence intervals.

varsoc variable_name
varsoc computes four information criteria as well as a sequence of likelihood ratio (LR) tests.
The information criteria include the FPE, AIC, the HQIC, and SBIC. For a given lag p, the LR test compares a VAR with p lags with one with p − 1 lags. The varsoc command is used to select the lag order for time series data. An '*' indicates the optimal lag.

dfuller variable_name, trend regress lag(optimum-1)
dfuller performs the augmented Dickey–Fuller test that a variable follows a unit-root process. The null hypothesis is that the variable contains a unit root, and the alternative is that the variable was generated by a stationary process. Because the data show a clear upward trend, we use the trend option with dfuller to include a constant and time trend in the augmented Dickey–Fuller regression. By getting p-value for Z(t) (0.00) we can reject the null hypothesis of a unit root at all common significance levels.

pperron variable_name , trend regress lag(optimum-1)
The null hypothesis is that the variable contains a unit root, and the alternative is that the variable was generated by a stationary process. By getting p-value for Z(t) (0.00) we can reject the null hypothesis of a unit root at all common significance levels.

dfgls variable_name
dfgls performs a modified Dickey–Fuller t test for a unit root in which the series has been transformed by a generalized least-squares regression. The null hypothesis is that the variable contains a unit root, and the alternative is that the variable was generated by a stationary process. We have to check t-stat for each lag value.

kpss variable_name
This is test for stationarity of a time series. The null hypothesis is that the variable was generated by a stationary process, and the alternative is that the variable contains unit root. This test differs from those in common tests by having a stationary null. We have to check t-stat for each lag value.

Tests for structural break with Known Break Points

a) Two types of dummy variables D1,D2. D1 is the dummy variable equal to 1 if the time variable > break date and 0 if time variable ≤ break date.
   Intercept dummy - **D1= (time>2003).**
   D2 is the dummy variable created by taking interaction term between the intercept dummy and the lagged dependent variable.
   Slope dummy - **D2 = D1*l1. variable_name.**
b) Now we estimate a regression equation – **reg variable_name l1.variable_name D1 D2**.
c) Then I have to perform the F-test by using – test D1 D2. Null hypothesis is no structural break and alternative hypothesis is structural break is present.

estat sbknown, break(year)
This test performs testing for known structural break by using a Wald or a likelihood-ratio (LR) test after estimating a model. Null hypothesis is no structural break and alternative hypothesis is structural break is present.

d) Unitroot Test of residual with Break Points

predict e_hat, residual
varsoc e_hat (to obtain the optimum lag length)
tsline e_hat (to see if this contains trend)
dfuller e_hat trend regress, lag(optimum-1)
If e_hat is stationary then we can infer that the inclusion of structural break have removed the non stationarity of the series.

Test for cointegration

Engle–Granger's Two-Step Method

- In the first step, we need to perform an OLS estimation

reg dependent_variable_name independent_ variable_name
If the value of R2 is very high then that relationship may be spurious.

- **estat dwatson**
  estat dwatson computes the Durbin–Watson d statistic (Durbin and Watson 1950) to test for first-order serial correlation in the disturbance when all the regressors are strictly exogenous. The null hypothesis of the test is that there is no first-order autocorrelation. The Durbin–Watson d statistic can take on values between 0 and 4 and under the null d is equal to 2.
  High R^2 and low Durbin watson statistics value imply spurious regression.
- To test for cointegration we have to carry out unit root test for the residual.
  predict e_hat, residual
  varsoc e_hat (to obtain the optimum lag length)
  tsline e_hat (to see if this contains trend)
  dfuller e_hat
  If we fail to reject null, i.e., e_hat contains unitroot then the original series are not cointegrated.

Vector Autoregression(VAR) Method

Before estimating a VAR model we must specify the order p of a VAR. To find the optimum lag length – **varsoc dependent_variable_name independent_ variable_name**

If we get optimum lag length 2 to specify a model that includes lags 1 and 2

var dependent_variable_name independent_ variable_name, lags(1/2)

If after estimating the VAR we can see that coefficients of own lags of dependent_variable_name (or independent_ variable_name) are significant but cross coefficients are not significant then we infer in this system of equations, dependent_variable does not depend upon independent_ variable; independent_ variable also has no significant relation with dependent_variable . Therefore, the unrestricted VAR of this type is of little use in analysing this relation.

vargranger

To check causality direction of a model we use Granger causality test . For each equation in a VAR, vargranger tests the hypotheses that each of the other endogenous variables has no causal effect on the dependent variable in that equation. Null is no causality

vecrank dependent_variable_name independent_ variable_name (Johensen)

It is to determine the number of cointegrating equations using Johansen's multiple-trace test method. If we can see * appears in our result we can't reject the null hypothesis of no cointegration, that is there is one co-integration relationship between these variables.

If * appears at 2 max no. of cointegration is 2. If we couldnt find any * then there is no cointegration.

Arch Garch

Engle (1982) developed the autoregressive conditional heteroscedastic (ARCH) class of models to analyse risk and uncertainty. The ARCH model estimates a weighted average of past
squared residuals.

reg d.var_name

estat archlm, lags(1)

if we can reject the null we infer presence of arch effect

predict e_hat, residual

ac e_hat

When the squared error in a regression model follows AR process, the error will follow the ARCH process.
To estimate a GARCH (1, 1) process
arch d.var_name, arch(1) garch(1)
To estimate a GARCH (2, 1) process
arch d.var_name, arch(2) garch(1)

Panel Data

xtset airline year
It is used to declare the cross-section and the time series variable. This also shows if the dataset is balanced.

Heteroskedasticity Test
Null hypothesis is no heteroskedasticity.
To test we can use lr test
xtgls total_cost output fuel_price load_factor, igls panels(heteroskedastic)
estimate store hetero
xtgls total_cost output fuel_price load_factor, igls
we will need to tell lrtest how many constraints we have implied.
local df = e(N_g) – 1
lrtest hetero . , df(`df')
If we can reject the null hypothesis then there is heteroskedasticity present in the data.

Autocorrelation Test
Wooldridge (2002, 282–283) derives a simple test for autocorrelation in panel-data models. Null hypothesis is no first order autocorrelation
xtserial total_cost output fuel_price load_factor
If we can reject the null hypothesis there autocorrelation is there in the panel data.

First we will check for cross section dependency. Then if we find CD we will perform second generation unit root test or we will perform first generation. Now suppose after taking d.variable unit root is removed then we will apply different estimating model Pooled, FE, RE of d.variable . After that we will carry out test to check which model is best fitted.

Cross-section Inter-dependence test - Testing for cross-sectional dependence is important in fitting panel-data models. To test for Cross-section dependency we will use the Pesaran CD(2004) test. Null hypothesis is no cross section dependence. This test employs the correlation-coefficients between the time-series for each panel member. To do this test the stata command is **xtcd output fuel_price load_factor**
Or
xtcdf output fuel_price load_factor

If null hypothesis is rejected we infer cross-section units are interdependent.

Unit root Test

Modelling
Fully Restricted Model: Pooled Regression - If we impose a strong restriction that every entity is homogeneous, then we have a purely restrictive model or pooled regression model. Pooled regression model is the multiple linear regression model with panel data.
reg total_cost output fuel_price load_factor
The F statistic tests the null hypothesis that all of the coefficients on the independent variables are equal to zero. If we reject this null hypothesis with extremely high confidence we can't accept that pooled OLS can be a good as it does not consider the heterogeneity across airline firms or years. So, the problem

of endogeneity arises, and the estimate is biased because of unobserved heterogeneity (uit and xit are correlated).
The estimates is stored as pool – **estimate store pool.**

One-Way Error Component Fixed Effects Model – In this model we assume

i) Individual effects are time constant but are not common across the entities.
ii) The idiosyncratic error varies over individuals and time.

In the fixed effects model we can estimate each μ_i along with β.
We can estimate this model by two method –

- The "within" estimation or mean-corrected estimation
- The least squares dummy variable (LSDV)

The "Within" Estimation – This is also know as mean-corrected estimation that uses variation within each individual or entity.

- First we have to generate mean value of dependent and the independent variable over time for each entity by the stata command

egen m_y = mean(total_cost), by (airline)

egen m_x_1 = mean(output), by (airline)

egen m_x_2 = mean(fuel_price), by (airline)

egen m_x_3 = mean(load_factor), by (airline)

- Now to generate time demean variables with the stata command

g w_m_y = total_cost - m_y

g w_m_x_1 = output - m_x_1

g w_m_x_2 = fuel_price - m_x_2

g w_m_x_3 = load_factor - m_x_3

By this step unobserved fixed effects will be eliminated in the mean-corrected model.

- Now we will estimate the demean equation by the stat command

reg w_m_y w_m_x_1 w_m_x_2 w_m_x_3

We store the estimate as we

estimate sore we

The panel regression with mean-corrected estimation can be performed by the stata command –

xtreg total_cost output fuel_price load_factor, fe

We store the estimate as fe - **estimate sore fe**

Goodness of fit of fixed effect model is higher that pooled reg model. Since no dummy variables are used in the within effect model, it has larger degrees of freedom for errors, accordingly, reporting small mean squared errors (MSE).

The least squares dummy variable (LSDV) - To estimate the fixed effects, LSDV assumes the unobserved fixed effects as the coefficients of the binary variables representing the cross-section units.

The LSDV works best when the panel data has relatively fewer cases and more time periods.

reg total_cost output fuel_price load_factor i.airline

Here each airline firm is regarded as a dummy variable for estimation.

estimate store Lsdv

One-way error component random effects model -The loss of degrees of freedom could be avoided if the unobserved effect µi is assumed to be random. If the unobserved effects are random, the error component model will be random effects model.

xtreg total_cost output fuel_price load_factor

estimate store Re

Tabulating the estimates of the model –

estimates table pool fd we fe re LSDV, star stats(N)

Pooled Regression Model Vs Fixed Effect Model

In testing for the validity of the fixed effect, we could test the joint significance of the dummies by performing an F test:

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_N = 0$

$H_1 : \mu_i \neq 0$

The null hypothesis supports the pooled regression with panel data.

testparm i.airline

The rejection of null hypothesis implies that the fixed effects model is better fitted than the pooled OLS.

Pooled Regression Model Vs Random Effect Model

Let, σ_μ^2 is the variance of the distribution of unobserved random effect.

The null hypothesis is given by

$$H0: \sigma_\mu^2 = 0$$

The alternative is

$H1: \sigma_\mu^2 > 0$

xttest0

Fixed Effect Model vs Random Effect model

The Hausman specification test basically tests whether the errors (ui) are correlated with the regressors.

H0: $E(u_{it} | X_{it}) = 0$ (Under null re is consistent)

H1: $E(u_{it} | X_{it}) \neq 0$

hausman fe re

The null hypothesis of this test is that individual effects are uncorrelated with any regressor in the model.

Dynamic Panel

Instrumental Variable Estimation

Anderson and Hsiao propose instrumental variable procedure to solve problem of endogeneity bias in dynamic panel.

xtivreg Dep_var_name Indep_var_name Indep_var_name_2 (l. Dep_var_name =l2. Dep_var_name) , fe

Here we have used 2-period lag value of log of labour (l2.ln_lab) as the instrument for 1-period lag value of the dependent variable (ln_lab).

xtivreg Dep_var_name Indep_var_name Indep_var_name_2 (l. Dep_var_name =l2. Dep_var_name) , re

xtivreg Dep_var_name Indep_var_name Indep_var_name_2 (l. Dep_var_name =l2. Dep_var_name) , fd

Arellano–Bond

Arellano and Bond (1991) developed a dynamic panel data model by utilising

the orthogonality conditions that exist between lagged values of yit and the disturbances

εit . Arellano and Bond (1991) derived GMM estimator for the parameters

of a dynamic panel data model by taking more instruments available.

xtabond Dep_var_name Indep_var_name, lags(1) noconstant

The Wald statistic is used to test the null hypothesis that all the coefficients are zero and the null hypothesis is significantly rejected.

To check over-identifying restrictions

estat sargan

If null is rejected we infer so there is over identifications