Training Manual

Econometrics for Socio-Economic Research: Tools and Applications

MODELING SIMULATION FORECASTING VISUALIZATION

November 2024

Agricultural Economics and Rural Sociology Division
Bangladesh Agricultural Research Council
Farmgate, Dhaka-1215, Bangladesh
www.barc.gov.bd

Training Manual

Econometrics for Socio-Economic Research: Tools and Applications

Compiled and Edited by:

Dr. Md. Mosharraf Uddin Molla, MD (AERS), BARC
Dr. Md. Shofiqul Islam, PSO (AERS), BARC

November 2024

Agricultural Economics and Rural Sociology Division
Bangladesh Agricultural Research Council (BARC)

Training Manual

Econometrics for Socio-Economic Research: Tools and Applications

10-14 November, 2024
Venue: Computer Training Lab, AIC Building, BARC

Dr. Md. Mosharraf Uddin Molla
Course Director

Dr. Md. Shofiqul Islam
Course Coordinator

The Training Module is Designed for Scientists of NARS Institutes

Organized by

Agricultural Economics and Rural Sociology Division
Bangladesh Agricultural Research Council (BARC)

Program Schedule

Day/Date	Time	Topic/Event	Resource person
10-11-24 Sunday	9:00-9:20	Registration	--
	9:20-9:40	Pre-Evaluation	
	9:40-10:10	Opening	--
	10:10-10.30	Tea Break	--
	10:30-11:20	Fundamentals of econometrics: Essential tools for understanding socio-economic dynamics	Dr. Md. Mosharraf Uddin Molla, MD (AERS), BARC
	11:20-12:10	Linking econometrics with social-economic research: Sampling design, area selection, and model selection	Dr. Ripon Kumar Mondal
	12:10-1:00	Exercise with STATA software: Constructing household dietary diversity score (HDDS) from survey data	Dr. Ripon Kumar Mondal
	1:00-2:00	Lunch	
	2:00-3:00	Exercise with STATA software: Multiple regression analysis and postestimation tests	Dr. Ripon Kumar Mondal
	3:00-4:00	Exercise with STATA software: Technical efficiency and productivity measurement using stochastic frontier analysis	Dr. Ripon Kumar Mondal
	4:00-5:00	Exercise with STATA software: Variable identifications from survey data	Dr. Ripon Kumar Mondal
11-11-24 Monday	9.00-10.00	Socio-economic research priorities in agriculture	Dr. Md. Mosharraf Uddin Molla, MD (AERS), BARC
	10.00-10:30	Tea Break	
	10:30-11:30	Basics of qualitative research design	Dr. Fatema Sarker
	11.30-12.50	Hands-on exercise on different qualitative data collection tools and participatory approach	Dr. Fatema Sarker
	12.50-2.00	Lunch	
	2.00-3.00	Application of Nvivo software: Orientation to data collection and data analysis	Dr. Fatema Sarker
	3.00-4.00	Exercise using Nvivo Software	Dr. Fatema Sarker
	4.00-5.00	Drivation of output supply and input demand elasticities of the agricultural farm	Dr. Md. Abdus Salam
12-11-24 Tuesday	9.00-10.00	Basics of the Choice experiment	Dr. Monoj Kumar Majumder
	10.0-10.30	Tea Break	
	10.30-11.30	Practice on formulating experimental design for choice experiment	Dr. Monoj Kumar Majumder
	11.30-12.30	Exercise on developing choice card	Dr. Monoj Kumar Majumder
	12.30-2.00	Lunch	

	2.00-3.00	Exercise 1: Practicing data analysis for choice experiment	Dr. Monoj Kumar Majumder
	3.00-4.00	Exercise 2: Practicing data analysis for choice experiment	Dr. Monoj Kumar Majumder
	4.00-5.00	Practice with R software: Identifying components of time series data	Md. Sazzadur Rahman Sarkar
13-11-24 Wednesday	9.00-10.00	Time series econometrics: Some basic concepts, sources, processing and transformation	Dr. Shah Johir Rayhan
	10.00-10:30	Tea break	
	10:30-11:30	Practice with Eviews software: Detecting stationarity of time series data	Dr. Shah Johir Rayhan
	11.30-12.30	Practice with Eviews software: Spatial price transmission in the agricultural commodity markets of Bangladesh with NARDL approach	Dr. Shah Johir Rayhan
	12.30-2:00	Lunch	
	2:00-3:00	Practice with Eviews software: Analyzing short-run and long-run asymmetrical effects of climate change on agricultural production	Dr. Shah Johir Rayhan
	3:00-4:00	Concept of impact evaluation and application of Heckman's treatment effect model	Dr. Md. Sadique Rahman
	4:00-5:00	Application of treatment effect models using STATA software	Dr. Md. Sadique Rahman
14-11-24 Thursday	9:00-10:00	Journal article writing tips	Dr. Md. Sadique Rahman
	10:00-10.30	Tea break	
	10:30-11:30	Exercise 1: Data analysis and article writing using hypothetical data	Dr. Md. Sadique Rahman
	11:30-1:00	Exercise 2: Data analysis and article writing using hypothetical data	Dr. Md. Sadique Rahman
	1:00-2:00	Lunch	
	2:00-3:00	An introduction to beta regression: Key concepts and applications	Dr. Md. Shofiqul Islam
	3:00-4:00	Post Evaluation	
	4:00-5:00	Closing & Certificate Awarding Ceremony	

Resource Persons

1. Professor Dr. Ripon Kumar Mondal, Sher-e-Bangla Agricultural University
2. Professor Dr. Md. Sadique Rahman, Sher-e-Bangla Agricultural University
3. Professor Dr. Shah Johir Rayhan, Sher-e-Bangla Agricultural University
4. Dr. Monoj Kumar Majumder, Associate Professor, Sher-e-Bangla Agricultural University
5. Dr. Fatema Sarker, Associate Professor, Sher-e-Bangla Agricultural University
6. Dr. Md. Abdus Salam, Principal Scientific Officer, AERS Division, BARC
7. Dr. Md. Shofiqul Islam, Principal Scientific Officer, AERS Division, BARC
8. Md. Sazzadur Rahman Sarkar, Principal Scientific Officer (A.C.), AERS Division, BARC

Contents

Linking Econometrics with Social-Economic Research: Sampling Design, Area Selection and Model Selection

Ripon Kumar Mondal, PhD
Professor
Department of Agricultural Economics
Sher-e-Bangla Agricultural University, Dhaka-1207

- **Lecture 1: Linking econometrics with social-economic research: Sampling design, area selection and model selection with exercise**

Why is econometrics needed for social-economic research?

Econometrics is the use of statistical and mathematical models to develop theories or test existing hypothesis in economics and to forecast future trends from historical data. In other words, Econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy. In the case of social science research, especially quantitative analysis, econometrics will provide evidence-based proof of any theory or hypothesis rather depending on any ones' particular judgements. However, the validity of the findings of using any econometric methods will depend on proper estimation of any model using appropriate data.

Sampling Design

We know there are many ways of sampling which are classified mainly into two groups: probability sampling and non-probability sampling. However, the main issue is to identify the number of samples required for any kind of analysis and say that the sample should be the best representative of the true population. Calculation of sample numbers largely depends on the objective of the research, and it will be very likely based on the number of parameters of interest. For instance, if we want to identify the determinants of the adoption of artificial insemination (AI) technologies among dairy farmers in Bangladesh, then the target parameter will be the probability of adoption. Now, if we consider whole Bangladesh as a unit, then we can do sampling centrally for Bangladesh and this sample could be divided among different selected representative locations. Furthermore, if our target is to compare the findings among different District, the sampling unit will be the district, and total sample will be calculated sample size * number of districts. Usually, the districts in this case will be called the Primary Sampling Unit. In addition, sample size will increase based on the number of parameters of interest.

Case 1: What will be the sample size if we want to identify the determinants of the adoption of artificial insemination (AI) technologies among dairy farmers in Bangladesh?

In this case, we are considering the single parameter: technology adoption and single unit location, Bangladesh. Suppose we do not know the adoption rate; we will consider the highest probability of adoption (0.50) in this case. Hence, we could apply the well-known Coachran's formula for sample size calculations:

$$n = \frac{Z_{\frac{\alpha}{2}}^2 \, P(1-P)}{e_i^2}$$

Where n is the required sample size, P is the proportion of population which as the attribute that in question, e is the margin of error (desired level of precision), and Z is the normal value at desired level.

If we know the population size, then we could use the following formula proposed by Iarossi, G. (2006):

$$n = \frac{Z_{\frac{\alpha}{2}}^2 \, P(1-P)}{e_i^2 + Z_{\frac{\alpha}{2}}^2 \frac{P(1-P)}{N_i}}$$

Where n is the required sample size, P is the proportion of population which as the attribute that in question, e is the margin of error (desired level of precision), Z is the normal value at desired level, and N is the population size.

Case 2: We want to investigate the effect of any interventions on the GHG emissions in the dairy farming in Bangladesh.

In this case, our target parameter of interest is single: GHG emission per cattle and single location, Bangladesh. Since the parameter of interest is a quantitative and continuous variable, we could not use the formula above. In this situation, we will use the following formula:

$$Sample\ size\ n = \frac{Z_{\frac{\alpha}{2}}^2 \ SD^2}{e_i^2}$$

Where Z is the value of confidence level (at 95% level, 1.96), SD is the standard deviation of the parameter of interest (will get from previous research or pilot study), and e is the precision level which needs to be calculated from previous research. Precision in this case will be calculated in the following way:

Precision $e = z \ \times SE\ of\ mean\ (from\ previous\ study)$ and

$$se = SD/\sqrt{n_0}$$

Where n_0 is the number of samples in previous study (pilot study) and SD is the standard deviation of previous study.

Suppose that from previous pilot study, the mean GHG emissions from 100 cows in Bangladesh was 11.4 Kg per year and standard deviation was 0.40. Now the sample size for present study will be:

$$Sample\ size\ n = \frac{Z_{\frac{\alpha}{2}}^2 \ SD^2}{e_i^2}$$

Precision:

$$e = z \ \times SE\ of\ mean$$

$$n = \frac{1.96 \times (0.40)^2}{(0.0784)^2}$$

$$n = 99.96 \approx 100$$

$$e = 1.96 \times 0.40/\sqrt{100}$$

$$e = 1.96 \times 0.04$$

$$e = 0.0784$$

During final sample calculation, the calculated number of samples could be adjusted by adding 10 % attrition and nearest rounding figure (if needed).

What will happen if there is many strata?

After calculating the total sample required, the sample could be divided into the strata both proportionately and disproportionately.

How many samples are required if we need to compare the effect of treatment and control group of farmers?

If we need to compare the effect of treatment and control group, we need to find the sample size of any one group (treatment or control) first, then we will take the same number of sample from the another group.

Selection of research area

The validity of the research findings largely depends on the selection of the best possible representative area. In this case, multistage sampling, or purposive sampling could be adopted. For instance, if the study cover all over Bangladesh, then area selection could be done following some criteria, such as climatic hotspots, administrative boundaries, market access proximity, cattle density, and so forth. If the study has special focus such as coastal areas, or Barind areas, samples should be selected based on these criteria. For instance, in the case of dairy farming, the following areas could be selected based on climatic hotspots and administrative boundaries. Upazilas could also be selected based on the density of cattle or using random sampling (whichever appropriate):

Table 1: Example of sample area selection

Sl. No.	Division	District	Climate hotspot
1.	Barisal	Patuakhali	Coastal and River estuary
2.	Chittagong	Cox's Bazar	Coastal
3.		Chandpur	Coastal and River estuary
4.	Dhaka	Dhaka	Urban
5.		Tangail	River
6.		Kishoreganj	Haor
7.	Khulna	Satkhira	Coastal
8.		Magura	RLHP*
9.	Mymensingh	Jamalpur	River
10.	Rajshahi	Pabna	Barind and River
11.		Naogaon	Barind
12.	Rangpur	Rangpur	Barind
13.		Kurigram	River
14.	Sylhet	Sunamganj	Haor

Selection of econometric models

Why Model Selection Matters

Model selection is an essential step in econometric analysis. The goal is to avoid common pitfalls such as overspecification, misspecification, and under specification of the model.

Properties of a Multiple Regression Model

MLR 1: Linear in parameter

MLR 2: Random sample of n observations

MLR 3: No perfect collinearity (In the sample (and therefore in the population), none of the independent variables is constant, and there are no exact linear relationships among the independent variables.) eg. Total income = Agricultural income + non-agricultural income. If we take both total income and agricultural income as independent variables, then there is a chance to be perfect or near to perfect collinearity.

MLR 4: The error u has an expected value of zero given any values of the independent variables. In other words, $E(u \mid x_1, x_2, x_3, \dots \dots, x_k) = 0$. When Assumption MLR.4 holds, we often say that we have **exogenous explanatory variables**. If x_j is correlated with u for any reason, then x_j is said to be an **endogenous explanatory variable**. A model will give unbiased estimates if it satisfied MLR.1 – MLR. 4

MLR 5: Homoskedasticity. The error u has the same variance given any values of the explanatory variables. In other words, $Var(u \mid x_1, x_2, x_3, \dots \dots, x_k) = \sigma^2$. Assumptions MLR.1 through MLR.5 are collectively known as the Gauss-Markov assumptions (for cross-sectional regression).

Including Irrelevant Variables in a regression Model (over specifying the model)

One issue that we can dispense with fairly quickly is that of inclusion of an irrelevant variable or over specifying the model in multiple regression analysis. This means that one (or more) of the independent variables is included in the model even though it has no partial effect on y in the population. (That is, its population coefficient is zero.). Including one or more irrelevant variables in a multiple regression model, or over specifying the model, **does not affect the unbiasedness of the OLS estimators**. However, the value of R2 in always increase even if the additional variable is irrelevant. **Does this mean it is harmless to include irrelevant variables?** No. including irrelevant variables can have undesirable effects on the variances of the OLS estimators. A larger variance means a less precise estimator, and this translates into larger confidence intervals and less accurate hypotheses tests.

Omitted Variable Bias (Underspecifying the model)

Now suppose that, rather than including an irrelevant variable, we omit a variable that actually belongs in the true (or population) model. This is often called the problem of **excluding a relevant variable or underspecifying the model**. This problem generally causes the OLS estimators to be biased. Omitted variable bias could raise the endogeneity issue as well.

Deriving the bias caused by omitting an important variable is an example of misspecification analysis. To overcome this misspecification, we need to add the important variables in the model as much as possible.

Selection of regression model: Regression model could be selected based on the AIC and BIC criteria. The model having lower AIC and BIC will be preferable, however, the use of information criteria is subjective. No formal inference can be drawn from the reported values. In a typical approach, a set of potential models is selected, and a superior model is selected from the values of information criteria. A superior model is the model with the lowest value of information criterion. For example, given two models, the model with the lowest AIC fits the data better than the model with the larger AIC.

Practice:

```
. use https://www.stata-press.com/data/r18/sysdsn1
(Health insurance data)
. mlogit insure age male nonwhite
  (output omitted)
. estat ic
Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	615	-555.8545	-545.5833	8	1107.167	1142.54

```
Note: BIC uses N = number of observations. See [R] IC note.
. mlogit insure age male nonwhite i.site
  (output omitted)
. estat ic
Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	615	-555.8545	-534.3616	12	1092.723	1145.783

```
Note: BIC uses N = number of observations. See [R] IC note.
```

The AIC indicates that the model including the site dummies fits the data better, whereas BIC indicates the opposite. As is often the case, different model-selection criteria have led to conflicting conclusions.

❖ **Lecture 2: Exercise with STATA software: Variable identifications from survey data**

For this exercise, we will use data from the IFPRI website: Bangladesh Integrated Household Survey (BIHS) 2018-2019 - IFPRI Dataverse

Please find the data folder in your desktop "barc_training_data_ripon" for this exercise. By following different steps, we will try to find some household and farm level variables using raw data.

❖ Lecture 3: Exercise with STATA software: Constructing household dietary diversity score (HDDS) from survey data

The HDDS are calculated based upon different numbers of food groups because the scores are used for different purposes[1]. The HDDS is meant to provide an indication of household economic access to food, thus items that require household resources to obtain, such as condiments, sugar and sugary foods, and beverages, are included in the score. Twelve food groups are proposed for the HDDS measurement. 24 hours recall basis food consumption data in the household is used to measure the HDDS.

Table 2: Aggregation of food groups from the questionnaire to create HDDS

SL	Food Group	Examples
1	Cereals	• corn/maize, rice, wheat, sorghum, millet or any other grains or foods made from these (e.g. bread, noodles, porridge or other grain products) + insert local foods e.g. ugali, nshima, porridge or paste
2	White tubers and roots	• white potatoes, white yam, white cassava, or other foods made from roots
3	Vegetables	• pumpkin, carrot, squash, or sweet potato that are orange inside + other locally available vitamin A rich vegetables (e.g. red sweet pepper) • forms + locally available vitamin A rich leaves such as amaranth, cassava leaves, kale, spinach • other vegetables (e.g. tomato, onion, eggplant) + other locally available vegetables
4	Fruits	• ripe mango, cantaloupe, apricot (fresh or dried), ripe papaya, dried peach, and 100% fruit juice made from these + other locally available vitamin A rich fruits • other fruits, including wild fruits and 100% fruit juice made from these
5	Meat	• liver, kidney, heart or other organ meats or blood-based foods • beef, pork, lamb, goat, rabbit, game, chicken, duck, other birds, insects
6	Eggs	• eggs from chicken, duck, guinea fowl or any other egg
7	Fish and other seafood	• fresh or dried fish or shellfish
8	Legumes, nuts and seeds	• dried beans, dried peas, lentils, nuts, seeds or foods made from these (eg. hummus, peanut butter)
9	Milk and milk products	• milk, cheese, yogurt or other milk products
10	Oils and fats	• oil, fats or butter added to food or used for cooking
11	Sweets	• sugar, honey, sweetened soda or sweetened juice drinks, sugary foods such as chocolates, candies, cookies and cakes
12	Spices, condiments and beverages	• spices (black pepper, salt), condiments (soy sauce, hot sauce), coffee, tea, alcoholic beverages

[1] FAO (2011). Guidelines for Measuring Household and Individual Dietary Diversity, available at FAO-guidelines-dietary-diversity2011.pdf

How to create HDDS?

Dietary diversity scores are calculated by summing the number of food groups consumed in the household or by the individual respondent over the 24-hour recall period. The following steps are included in creating either the HDDS:

1. Create new food group variables for those food groups that need to be aggregated
2. Create a new variable termed either HDDS
3. Compute values for the dietary diversity variable by summing all food groups included in the dietary diversity score
4. As a check on the creation of the variables, all scores should be within the 0-12 range

Use of HDDS

Table: 3 When to measure HDDS

<table>
<tr><th>Objective</th><th colspan="2">Timing</th></tr>
<tr><td rowspan="2">Assessment of the typical diet of households/ individuals</td><td>In rural, agriculture-based communities</td><td>Any time of the year (if seasonality is not an issue)</td></tr>
<tr><td>When food supplies are still adequate3 (may be up to 4-5 months after the main harvest).
Looking at dietary diversity at different points in the agricultural cycle is one way of investigating seasonality of food security.
In many areas there are important seasonal differences in dietary patterns. For a more complete assessment of usual diet, dietary diversity should be measured during different seasons</td><td>Any time of the year (if seasonality is not an issue)</td></tr>
<tr><td>Assessment of the food security situation in rural, agriculture-based communities</td><td colspan="2">During the period of greatest food shortage, such as immediately prior to the harvest or immediately after emergencies or natural disasters.
This may also serve as a baseline for monitoring change due to an intervention or for investigating seasonality</td></tr>
<tr><td>Assessment of the food security situation in non-agricultural communities</td><td colspan="2">At the moment of concern to identify a possible food security problem.
May also serve as a baseline for monitoring changes due to an intervention</td></tr>
<tr><td>Monitoring of food security/nutrition programmes or agricultural interventions such as crops and livelihood diversification</td><td colspan="2">Repeated measures to assess impact of the intervention on the quality of the diet, conducted at the same time of year as the baseline (to avoid interference due to seasonal differences).</td></tr>
</table>

- ❖ **Lecture 4: Exercise with STATA: Multiple regression analysis and postestimation tests**

Now suppose that we want to investigate the effect of rumen livestock rearing on household food security in Bangladesh. For this analysis, we will use the variables which we already have identified. For instance, we will use the HDDS variable as the proxy for the household food security as the dependent variable. We will consider the number of ruminants (cattle, buffalo, goat and sheep) as the target independent variables along with other control variables such as , cultivable land area, family size, age of the household head, years of schooling of the household head, gender of the household head, and average monthly household income.

We use the following command for the regression:

reg hdds cultivable_land family_size age year_school newgender no_livestock ln_avg_m_income

Post estimation tests:

Now to test the omitted variable bias, we will use the following command in STATA

```
. ovtest

Ramsey RESET test using powers of the fitted values of hdds
       Ho:  model has no omitted variables
                F(3, 4216) =      5.24
                  Prob > F =      0.0013
```

Comment: The result shows that we have omitted variable since we have F-value significant (p-value 0.001). Hence, we have to find more important variable for the model. Since, socioeconomic characteristics across different regions in Bangladesh is not the same, we will use the district level fixed effect in the estimation to solve the omitted variable bias. However, after adding the district level fixed effect we need to check whether still omitted variable exists.

To test the heteroskedasticity, we use the following command:

```
. hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of hdds

         chi2(1)      =    19.85
         Prob > chi2  =   0.0000
```

Comment: Since the p-value is significant (0.000), we have heteroskedasticity in the data. Hence we, need to use the robust standard error.

To test is multicollinearity is an issue, we will use vif command after the regression estimation

. vif

Variable	VIF	1/VIF
ln_avg_m_i~e	1.55	0.644784
newgender	1.54	0.649261
cultivable~d	1.14	0.880923
no_livestock	1.12	0.892091
age	1.11	0.903874
year_school	1.08	0.927276
family_size	1.04	0.965175
Mean VIF	1.22	

Since the cut-off point of VIF is 10 which is much higher than the estimated value 1.22, hence multicollinearity is not an issue in this analysis.

reg hdds cultivable_land family_size age year_school newgender no_livestock ln_avg_m_income i.district , robust

```
. reg hdds cultivable_land family_size age year_school newgender no_livestock ln_avg_m_income i.distr
> ict , robust

Linear regression                               Number of obs     =      4,227
                                                F(70, 4156)       =      14.26
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1733
                                                Root MSE          =     1.4617

------------------------------------------------------------------------------
                |               Robust
           hdds |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
cultivable_land |    .147182    .049082     3.00   0.003     .0509551    .2434089
    family_size |   .0903479   .0101989     8.86   0.000     .0703525    .1103432
            age |  -.0026809   .0017888    -1.50   0.134    -.0061878     .000826
    year_school |   .0608352   .0064569     9.42   0.000     .0481761    .0734942
      newgender |   .0941467   .0734477     1.28   0.200    -.0498501    .2381435
   no_livestock |    .037566   .0109205     3.44   0.001     .0161561     .058976
ln_avg_m_income |   .0637531   .0113059     5.64   0.000     .0415874    .0859188
                |
       district |
             2  |    .448092   .4838744     0.93   0.354    -.5005607    1.396745
             3  |   .3275684   .2481348     1.32   0.187    -.1589087    .8140454
```

Now do the *ovtest* again

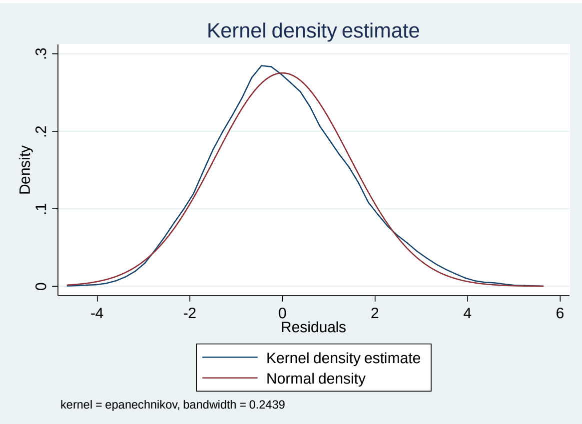
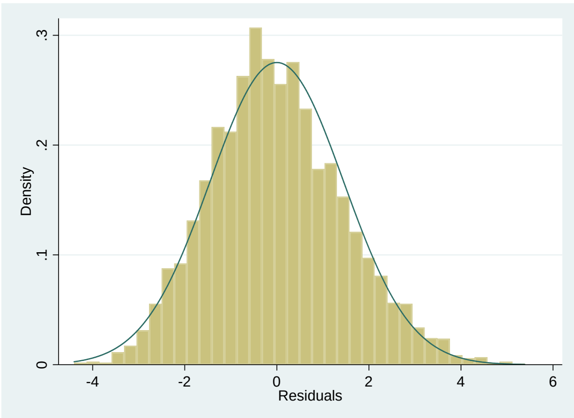
```
. ovtest

Ramsey RESET test using powers of the fitted values of hdds
       Ho:  model has no omitted variables
                 F(3, 4153) =      1.93
```

Comment: No omitted variables

Now we will test the normality of the estimated model with following comment

predict u, residuals

Command: kdensity u, normal	*Command: histogram u, normal*
	

Comment: Both distributions show that the residual of the model is normally distributed

❖ **Lecture 5: Exercise with STATA software: Technical Efficiency and Productivity Measurement using Stochastic Frontier Analysis**

Measuring agricultural productivity:

To keep measures of productivity consistent and aligned with economic theory, production should measure the total output of a specific production process that combines intermediate inputs and factors of production to create a product. It is counted if the product is sold for domestic final consumption, including home consumption by the agricultural household, for export or added to inventories.

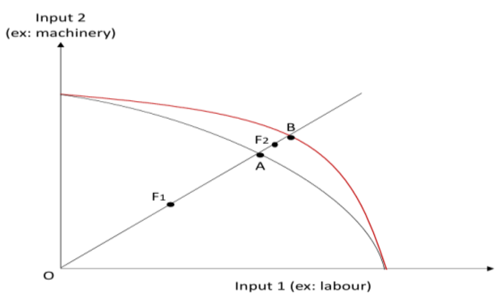
Suppose a farmer sells grain to a feed processing mill that, in turn, sells processed feed to a livestock farmer. Most statistical systems would count the sale from the farm to the mill as a sale from agriculture (part of output) and the purchase of the feed from the mill as an intermediate input. Now consider feed grown on the farm that is used for the farmer's own livestock. It is common and correct not to count own account feed as an output if agriculture productivity is being measured. This holds except if there is an interest in measuring crop productivity or livestock productivity separately.

Following the above example, output can be measured as the sum of sales plus own consumption plus change in inventories. It is also appropriate to measure livestock inventory change in weight gain and not just by the change in the number of heads so that the compositional change in the livestock herd can be better accounted for. As this approach is very data intensive, the number of head method is mostly used. Using auxiliary information and parameters can derive weight estimates. Crop production is measured in the net of harvesting losses and, if possible, net of other on-farm post-harvest losses, to capture the amount that is actually available for use or to be sold. Reducing farm losses would directly translate into higher productivity, as it would lead to higher output with no additional input cost.

Measuring Technical Efficiency

Several methods can be used to quantify technical efficiency. All of them broadly follow the same logic: identifying the share of productivity growth resulting from efficiency changes through the measurement of the distance between observed productivity and a theoretical, optimal or average productivity.

Figure 1. Technical efficiency and productivity: an illustration



Based on figure 1, measuring technical efficiency entails determining the distance between F1 and A, a technically efficient input-output combination. In practice, the ratio OF1/OA is the measure of technical efficiency or, equivalently, OA/OF is a measure of technical inefficiency. The methods to measure technical efficiency differ essentially on the way this distance is defined and estimated and whether auxiliary information is used. Most of these methods can provide farm-level estimates of technical efficiency.

Estimation Practice:

Suppose that we want to find the input productivity and technical efficiency of HYV Boro rice production in Bangladesh. Firstly, we will identify the related variables such as production of HYV Boro rice, inputs used, and other technical efficiency related variables. We will apply the stochastic frontier approach to meet our objective:

The traditional SFP log-linear Cobb-Douglas model takes the following form.

$$\ln(Q_i) = \beta_0 + \sum_{j=1}^{k} \beta_j ln(X_{ji}) + v_i - u_i \qquad (1)$$

where Q denotes the value of the total crop production per hactre i and X_{ji}denotes the *j-th* agricultural inputs used by the *i-th* farm household; β_0 and β_j denote the intercept and parameter coefficient of j-th input respectively. The agricultural inputs $(X_j, j = 1,2, \dots k)$ include family labour man-days per ha $(j = 1)$, hired labour man-days per ha $(j = 2)$, Urea Kg per ha $(j = 3)$, TSP Kg per ha $(j = 4)$, and other chemical fertilizer per ha $(j = 5)$. The term v_i, is the idiosyncratic component which is assumed to be independently $N(0, \sigma_v^2)$ distributed. The term u_i denotes the nonnegative random variable which accounts for the technical efficiency in the production process for the *i*-th household. Therefore $-u_i$ can be interpreted as the technical inefficiency term. Moreover, technical efficiency (u_i) could also be shown as a function of some explanatory variables related with the technical efficiency of production.

$$u_i = \Omega_1 + \sum_{j=1}^{k} \Omega_j m_{ji} + \epsilon_i \qquad (2)$$

where m_j $(j = 1,2, \dots . k)$ denotes the household head's age $(j = 1)$, age^2 $(j = 2)$, education $(j = 3)$, family size $(j = 4)$, distance from nearest local shop $(j = 5)$, Now we will have a single equation while substituting equation (8) into equation (7), which can be estimated applying the maximum likelihood method in SFP using the Cobb-Douglas functional form.

frontier ln_production_ha ln_family_lab_ha ln_hired_lab_ha ln_urea_ha ln_tsp_ha ln_other_ch_fer_ha , uhet (age year_school family_size local_shop_distance)

```
Stoc. frontier normal/half-normal model          Number of obs     =      1,671
                                                 Wald chi2(5)      =      23.45
Log likelihood = -1225.2863                      Prob > chi2       =     0.0003
```

ln_production_ha	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_production_ha						
ln_family_lab_ha	.0137474	.0075739	1.82	0.070	-.0010972	.0285919
ln_hired_lab_ha	.0059607	.0053466	1.11	0.265	-.0045185	.01644
ln_urea_ha	.0608666	.0172503	3.53	0.000	.0270565	.0946766
ln_tsp_ha	.0036544	.0026799	1.36	0.173	-.0015982	.0089069
ln_other_ch_fer_ha	.010219	.0042745	2.39	0.017	.0018411	.0185968
_cons	8.580155	.1035446	82.86	0.000	8.377212	8.783099
lnsig2v						
_cons	-6.452797	.4033136	-16.00	0.000	-7.243277	-5.662317
lnsig2u						
age	-.0368123	.0026399	-13.94	0.000	-.0419864	-.0316382
year_school	-.078394	.0094727	-8.28	0.000	-.0969602	-.0598279
family_size	.0333352	.0158261	2.11	0.035	.0023167	.0643538
local_shop_distance	.0310165	.0226385	1.37	0.171	-.0133542	.0753872
_cons	1.743038	.164179	10.62	0.000	1.421253	2.064822
sigma_v	.0397002	.0080058			.0267388	.0589445

predict tef, te

egen fegrp = cut(te), at(0,.3,.6,.9,1)

tab fegrp

. tab fegrp

fegrp	Freq.	Percent	Cum.
0	72	4.31	4.31
.3	517	30.94	35.25
.6	958	57.33	92.58
.9	124	7.42	100.00
Total	1,671	100.00	

Exploring Qualitative Research: Design, Data Collection Tools, and Participatory Approaches

Dr. Fatema Sarker
Associate Professor
Department of Development and Poverty Studies
Faculty of Agribusiness Management
Sher-e-Bangla Agricultural University

❖ Lecture 1: Basics of qualitative research design

What is qualitative research?

Qualitative research is a process of naturalistic inquiry that seeks an in-depth understanding of social phenomena within their natural setting. It focuses on the "why" rather than the "what" of social phenomena and relies on the direct experiences of human beings as meaning-making agents in their everyday lives.

Rather than by logical and statistical procedures, qualitative researchers use multiple systems of inquiry for the study of human phenomena including biography, case study, historical analysis, discourse analysis, ethnography, grounded theory, and phenomenology.

What distinguishes qualitative from quantitative research methods?

	Quantitative research methods	**Qualitative research methods**
Type of data	Numerical data: Numbers	Non-numerical data: Data that can be presented in textual form
Type of analysis	Statistical - econometric	Interpretative
Conceptual frameworks	Often neo-classical economics Rational choice	Range of social sciences frameworks
Drawing inferences	Typically deductive	Typically inductive
Sample size	Large – goal is statistical representativeness	Small – goal is often “saturation”

Examples of disciplines that use qualitative methods

- Law
- History
- Political science
- Sociology
- Psychology

Some misperceptions about qualitative research

➢ Misperceptions

- Qualitative research means you just interview people.
- Qualitative research is less rigorous than quantitative research.
- Doing qualitative research does not require specific training, everyone can do it.
- Qualitative research requires less preparation than quantitative research.

➢ In reality

- Qualitative research requires different skills from quantitative research.
- Qualitative research requires as much preparation as quantitative research.
- Documenting qualitative findings, analyzing them and writing them up is as challenging as analyzing quantitative data.

Keep in mind

Qualitative research deals with questions that begin with: **Why? How? In what way?**

and not generally with: **how much, how many and to what extent?**

Qualitative research design

Research designs for qualitative studies:

- Case Study Approach;
- Grounded Theory;
- Comparative Historical Research Approaches

Case study

Definitions of the case study approach

- “in-depth, qualitative studies of one or a few illustrative cases” (Hagan, 2006: 240)
- “Attempt to systematically investigate a event or a set of related events with the specific aim of describing and explaining this phenomenon” (see, e.g., Bromley, 1990)
- "a detailed examination of one setting, or a single subject, a single depository of documents or one particular event" (see Gomm, Hammersley & Foster 2000, Yin, 2003)
- "a method involving systematically gathering enough information about a particular person, social setting, event or group to permit the researcher to effectively understand how the subject operates or functions” (Berg, 2009: 317)

What is “a case”?

- Depending on the study, could be an event, a person, an organization, a process, a location

May involve various data collection methods

- Typically, qualitative and taking a holistic approach, but contextualized quantitative studies may also use case study design
- Case Studies may vary in length (could be longitudinal)

What is considered a case study is a matter of definition

- Different disciplinary perspectives

How to select cases?

- In grounded theory: Theoretical sampling
- For rural development studies
- Comparative case study approach often useful
- Strategies for case selection-

✓ If case studies are combined with surveys, or survey data are available
✓ Select cases to represent "types" of interest
✓ Hold as many factors as possible constant, and vary the factor of interest
✓ "Border strategy", comparing extremes
✓ Studying particularly interesting cases (Harvard Business School model)

- **Often useful in preparation of any type of study**

✓ Interdisciplinary team – to find out "what is really going on"

Strategies to select cases

Type	Definition of sampling strategy
Extreme case	The case demonstrates unusual manifestation of the phenomenon, such as outstanding success and notable failures
Intensity case	The case is information rich but not an extreme case.
Maximum variation	Cases, despite having diverse variations, exhibit important common patterns that cut across variations.
Homogeneous	Variation between cases is minimized, analysis is simplified and study is focused.
Typical case	Case illustrates what is typical, normal or average.
Stratified purposeful case	Case illustrates characteristics of a particular subgroup to facilitate comparison and not for generalization or representation.
Critical case	Case that permits logical generalization to other cases because if it is true to this one case, it's likely to be true to all other cases

Snowball	Cases of interest from people who know people who know people who know cases, rich information rich, good examples for study, etc.
Type	Definition of sampling strategy
Criterion	Cases picked because they meet some predetermined criterion.
Theoretical	The cases are manifestation of a theoretical construct and are used to examine and elaborate on it.
Confirming and disconfirming	Cases that elaborate on initial analysis to seek exceptions or test variations.
Opportunistic	Cases that emerge from following leads during field work.
Random purposeful	Cases are randomly selected from a large sample for the purpose of increasing credibility and not for generalization or representation.
Politically important case	Cases are selected or eliminated because they are politically sensitive cases.
Convenience	Cases are selected on the basis of minimum effort, time and money. They are candidate examples of low credibility, information rich cases.
Combination	Cases are flexible and meet different interests and needs

- ❖ **Group exercise**
- ❖ **Lecture 2: Hands-on exercise on different qualitative data collection tools and participatory approach**

Qualitative data collection and analytical methods

- Focus Group Interview
- Participatory rural appraisal (PRA)
- Net-Map and Social Network Analysis
- Content Analysis
- Ethnographic Research Methods
- Discourse analysis

Focus group Interview

A focus group discussion involves **gathering people from similar backgrounds or experiences together to discuss a specific topic of interest**. It is a form of qualitative research where questions are asked about their perceptions attitudes, beliefs, opinion or ideas.

How to conduct a FGD?

1. Formulating the Research Question and Drafting a Discussion Guide
2. Operational Planning
3. Sampling and Recruitment
4. Conduct

Analysis

A concrete question asked during the discussion is not the same as the overall research question. Likewise, what participants say is not equivalent to answering the research question. Before answering the research question, a researcher needs to analyze and interpret the data collected from the FGDs. The analysis of qualitative data is difficult and very time consuming, therefore be sure to reserve enough time for this task. Spending time to conceptualize the entire study process before data collection starts will make it easier to interpret the results later on. Data analysis typically consists of several phases:

1. Transcription: Transcribing recorded statements so that a detailed, written document is available about who said what about a particular question. Transcription of one group discussion takes several hours and generates many pages of text.

2. Coding the transcription: Coding the transcription using 'codes' (and corresponding 'sub-codes' leading to a 'code path' or 'code tree'). Codes are 'labels' that summarize or bookmark short fragments of text, and therefore help to sort and structure the data. Several procedures can be used to established these codes, and it is possible to include different types of codes in one analysis:

a. Deductive codes – those specified before data collection, based on the research question; the Framework Method (Gale et al. 2013) is a valuable and frequently used example.

b. Inductive codes – those that emerge from the analyzed text itself, as in Grounded Theory (Charmaz 2006).

c. Codes referring to the group dynamic, which later help to understand how a group opinion was established in the course of interaction.

3. Reviewing memos: Reviewing memos produced by the researcher and other members of the research team during the course of the study. Such memos often contain reflections on the process of data collection or insights into the research problem. The reality is that qualitative data analysis often begins in the field, because a researcher – exposed to data while collecting them – cannot and should not attempt to refrain from understanding and pre-interpreting data (Pope, Ziebland, and Mays 2000). Such 'interim analysis' is one of the strengths of qualitative research, which allows for refining the research question and instruments when pre-interpreted data suggest the need for it.

4. Analyzing and interpreting qualitative data: Typically, through a two-step approach (Silverman 2006; Wong 2008): a. First, look at what people in the group literally said, remembering that the group, rather than the individual, is the unit of analysis. This part is rather simple and descriptive. A researcher performing this initial step of the analysis will report that, for example, "the consensus achieved by the group was …", "the majority of participants agreed that …", "there were several contradictory opinions about …", "almost no one mentioned ...".

Please note, however, that quantifying findings, although feasible, does not usually add value to scientific research by means of FGDs. b. Second, interpret what people said in an integrated, theoretical way. This often relies on: i. mapping a problem ii. identifying patterns, regularities and themes iii. identifying differences and similarities within the data and between different sources of data iv. making comparisons between different groups involved in the topic (Bromley et al. 2003)

5. Establishing validity and reliability through consensus, coherence, triangulation and reflexivity: Conducting a respondent check is a useful first step towards validating the results. It requires presenting the findings to the discussion participants or to the community (Bromley et al. 2003:16). It does not require that participants support all results and conclusions made by the researcher (and vice versa), but respondent validation can strengthen or weaken the level of trust in the results, and might bring about new insights and motivate the researcher to refine or modify his/her findings.

To successfully establish the reliability of qualitative findings, the researcher is expected to actively think about how his/her own social, economic, ethnic, religious, cultural, personal and scientific background might influence the chosen scientific approach and mode of interpretation. Finally, contrast qualitative FGD results with findings from other techniques used in the same or similar study, or with another data source, such as literature review. This is called 'triangulation' or 'cross-validation' (for instance, through the application of interview, observation, self-reporting and/or meta analysis).

JUST to remember

The technique is based upon the assumption that the group processes activated during an FGD help to identify and clarify shared knowledge among groups and communities, which would otherwise be difficult to obtain with a series of individual interviews.

Yet, this method does not presume that A) all the knowledge is shared equally among a studied group, or that B) in each community there is a common, underlying, homogeneous knowledge.

Rather, an FDG allows the investigator to solicit both the participants' shared narrative as well as their differences in terms of experiences, opinions and worldviews during such 'open' discussion rounds.

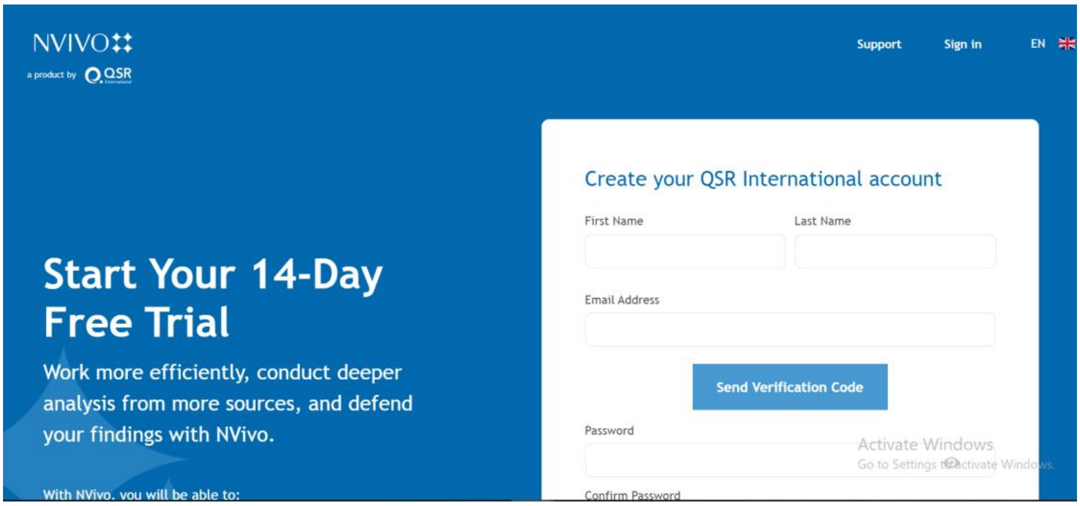
❖ Lecture 3: Application of NVivo software: Orientation to data collection and data analysis

NVivo

Computer program for qualitative data analysis

Developed by the QSR international http://www.qsrinternational.com/

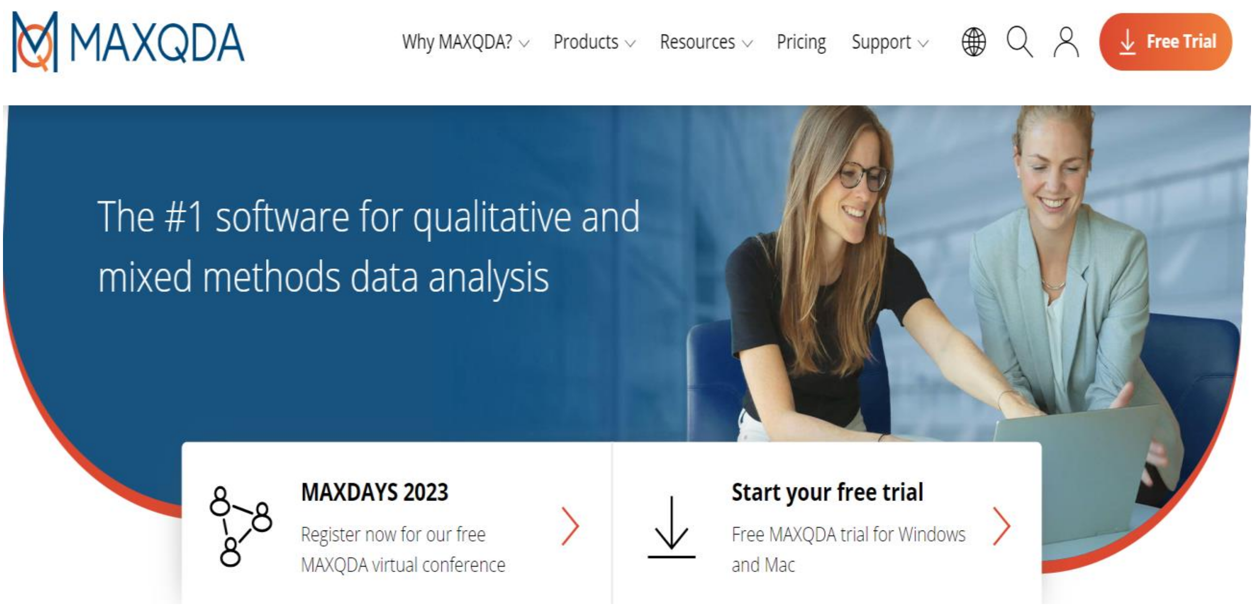
Free 14-day trial versionhttp://www.qsrinternational.com/trial-nvivo



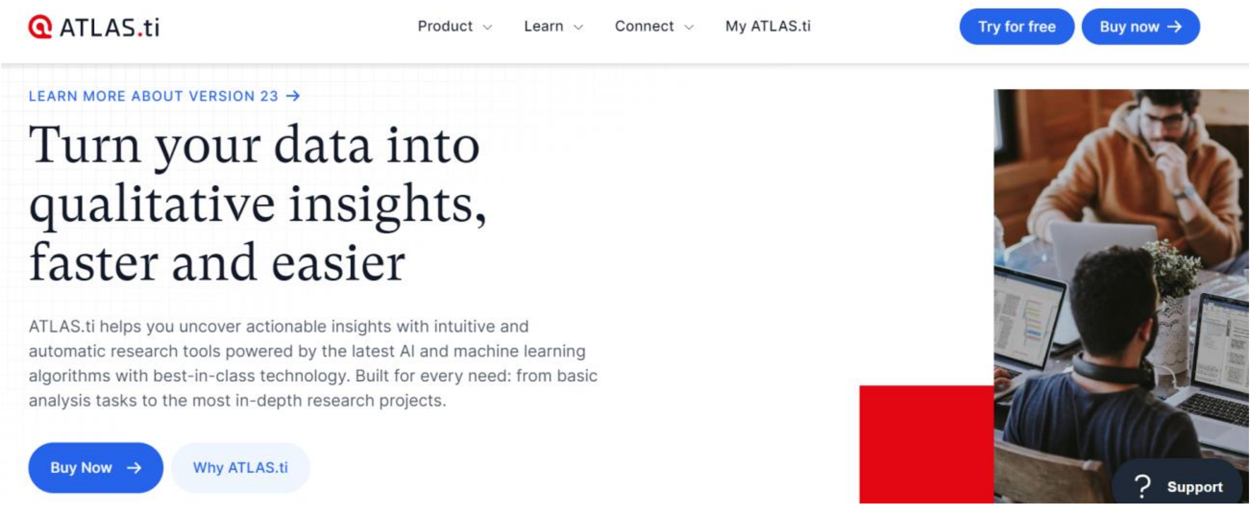
- **A computer program for qualitative data analysis that facilitates**
  - Managing data sources
    - Organizing and keeping track of transcripts, field notes, memos, conceptual maps, etc.
  - Coding text
    - Assigning "labels" ("nodes") to text blocks and managing the codes
  - Querying data:
    - Asking questions, storing results and performing further analysis
  - Displaying data in graphic form
    - Visual display of ideas or concepts from data
- **Watch the introductory tutorial**
  - https://www.youtube.com/watch?v=eXCsA175Ga0&index=1&list=PLNjHMRgHS4Fcx3NfpKsaqXuGdcxI9y-Qa

MAXQDA

MAXQDA is a software program designed for computer-assisted qualitative and mixed methods data, text and multimedia analysis in academic, scientific, and business institutions. It is being developed and distributed by VERBI Software based in Berlin, Germany.



Atlas.ti



Exercise: NVivo/MAXQDA

What we can do in NVivo?

- Import, analyze and manage data
- Code and organize nodes (themes)
- Run queries on data
- Use visualizations to show data connections
- Import and export bibliographic data
- Share projects and data
- Export data and import into other programs

Overview of the NVivo Workspace

- Ribbon (commands organized in groups)
- Navigation view (view to organize folders and access different NVivo components)
    - Can customize view (I like vertical view of nodes and detail pane when coding)

The items available in Navigation View include:

- Sources—the collective term for your research materials including documents, PDFs, datasets (for example, spreadsheets), audio, video and pictures.
- Nodes—containers that let you gather related material in one place so that you can look for emerging patterns and ideas. You can create and organize nodes for themes, people, organizations or other cases. You can also create nodes to gather evidence about the relationships between items in your project.
- Classifications—descriptive information about your sources, nodes and relationships.
- Collections—views (or groupings) of project items that are stored elsewhere in your project—for example a set made up of sources you need to review or Search Folders for frequently used searches.
- Queries—search criteria that can help you to find and explore patterns in source text or coding. You can save queries and rerun them as your project progresses.
- Reports—reports and extracts that you can use to track your progress or make your data available for use in other applications.
- Models—shapes and connectors that provide a way of visually exploring or presenting the data in your project.
- Detail View (shows what you click on in navigation view)
- Status Bar (shows what you are doing)

Process

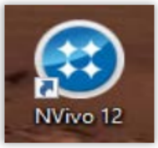
1. Create a new file (button on bottom left of screen (creates a single file .npv) – this is a container for all of your research "sources"
    - Don't store too much in this file – it gets too big and more likely to crash
    - You can link resources
    - Backup copies of this file often – make a _BU with date file
2. Import sources (some default folders there cannot be changed but you can add more folders and subfolders)
    a. Internal (word transcripts, spreadsheets, PDFs (can convert webpages and PowerPoints to PDFs), audio, video, pictures)
        i. To make subfolders (CREATE – COLLECTIONS – FOLDER)
        ii. Can right click and add as well
    b. External (cannot physically be imported such as books and research references)
        i. Can import from Endnote, Zotero or Refworks
    c. Memos – for notes and insights
        i. Can import or just create new as you code

3. Creating Nodes (containers for gathering related materials or a collection of “references” you get by coding)
    a. Under Nodes you have
        i. Nodes
        ii. Relationships
        iii. Node Matrices
    b. You can set up a hierarchy of codes (parent to children)

❖ Lecture 4: Exercise using NVivo software

Exercise 1. *Access NVIVO*

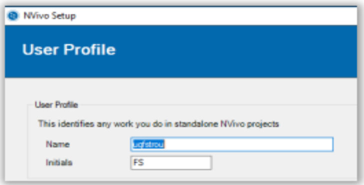
1. Double-click the **Nvivo 12** icon



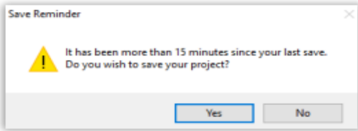
2. Complete profile details, if prompted
3. Add your initials.

These will be used to identifiy your edits as you progress

4. Click on **OK**

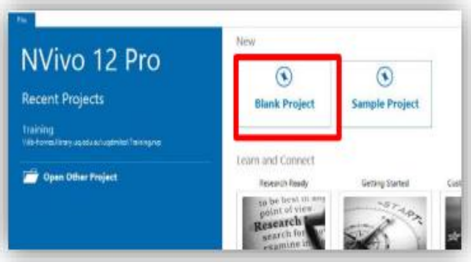


During your session you will receive on-screen prompts to save your progress. The save time can be changed via **File** (tab) – **Options** – **Notifications** (tab)



Exercise 2. *Create a new project*

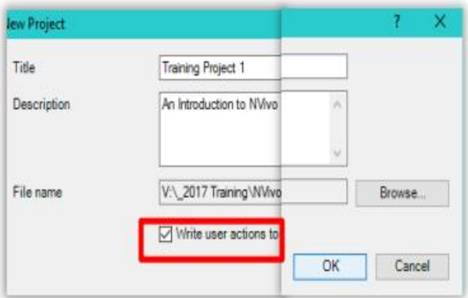
1. Click on the **Blank project** option



2. Complete project details
3. Click **Browse** to save project to your preferred location.

Note: If you are collaborating with other users it is advisable to tick the checkbox to Write user actions to project event log

4. Click on **OK**

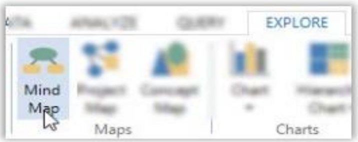


Exercise 3. ***Building a Mind Map***

When considering topics that may be present in your data you can create a mind map to visually explore potential concepts. These can be used as a brainstorming tool for Planning your node hierarchy, during analysis to explore how people talk about a topic or to plan how you will tell the story of your research.

a. **Create a Mind map**

1. Click on the Explore tab and click **Mind Map**



2. Enter a Name: **Mind Map Intro**
3. Add a **Description (**Optional**)**
4. Click on **OK**

Note the location "Maps" found in the Navigation view towards the bottom

A new tab for Mind Map tools will appear in the ribbon

Just to remind

Nvivo is just simple software. We can't expect that we enter the codes, it produces the results. The precision of analysis and correct interpretation depends on our skill and ceaseless practice. Thank you so much for your attention. All the best to you all.

References/Books

- Patton M. Qualitative Research and Evaluation Methods, 3rd edition. Sage Publishers; 2002.
- Curry L, Nembhard I, Bradley E. Qualitative and mixed methods provide unique contributions to outcomes research. Circulation, 2009;119:1442-1452.
- Crabtree, B. & Miller, W. (1999). Doing qualitative research, 2nd edition. Newbury Park, CA: Sage.
- Schensul S, Schensul J. and Lecompte M. 2012 Initiating Ethnographic research: A mixed Methods Approach, Altamira press.
- Barbour, Rosaline S. 2005. "Making Sense of Focus Groups." Medical Education 39: 742–750. ———. 2014. "Analyzing Focus Groups." In The SAGE Handbook of Qualitative Data Analysis, ed. Uwe Flick. London: SAGE Publications, 313–326.
- Bromley, Helene et al. 2003. Glossary of Qualitative Research Terms. London: King's College, University of London

Designing Choice Experiments: Fundamentals and Practical Applications

Dr. Monoj Kumar Majumder
Associate Professor and Chairman
Department of Agricultural Economics
Sher-e-Bangla Agricultural University, Dhaka-1207

❖ Lecture 1: Basics of the Choice Experiment (CE)

What is a choice experiment?

A choice experiment looks to carry out a choice model, which is the decision-making process of an individual or group by highlighting preferences in a given situation. Using surveys, you can estimate the respondents' relative value of the various attributes you provide.

For example, a company that manufactures baby food wants to determine what information should be emphasized on its product labels. They use a choice experiment to show various versions of the proposed labels—one features taste, one emphasizes the product is organic, another is focused on value/price, and one features nutritional values.

The results, collected via survey, will determine what attribute is most important to their target market. This information will be used to choose the label, inform product messaging, and potentially change the direction of their entire marketing strategy.

The most common uses of choice experiments are:

- Making updates to an existing product
- Determine willingness to pay during the price optimization process
- Optimizing product configuration
- Optimizing pricing of a portfolio of products

Applications

- ❖ Marketing researchers use discrete choice models to study consumer demand and to predict competitive business responses, enabling choice modelers to solve a range of business problems, such as pricing, product development, and demand estimation problems. In market research, this is commonly called conjoint analysis.
- ❖ Transportation planners use discrete choice models to predict demand for planned transportation systems, such as which route a driver will take and whether someone will take rapid transit systems. The first applications of discrete choice models were in transportation planning, and much of the most advanced research in discrete choice models is conducted by transportation researchers.
    - ▪ Disaster planners and engineers rely on discrete choice models to predict decision take by householders or building occupants in small-scale and large-scales evacuations, such as building fires, wildfires, hurricanes among others. These models help in the development of reliable disaster managing plans and safer design for the built environment.

- Energy forecasters and policymakers use discrete choice models for households' and firms' choice of heating system, appliance efficiency levels, and fuel efficiency level of vehicles.
- Environmental studies utilize discrete choice models to examine the recreators' choice of, e.g., fishing or skiing site and to infer the value of amenities, such as campgrounds, fish stock, and warming huts, and to estimate the value of water quality improvements.
- Labor economists use discrete choice models to examine participation in the work force, occupation choice, and choice of college and training programs.
- Ecological studies employ discrete choice models to investigate parameters that drive habitat selection in animals.

Common features of discrete choice models

Discrete choice models take many forms, including: Binary Logit, Binary Probit, Multinomial Logit, Conditional Logit, Multinomial Probit, Nested Logit, Generalized Extreme Value Models, Mixed Logit, and Exploded Logit. All of these models have the features described below in common.

Choice set

The choice set is the set of alternatives that are available to the person. For a discrete choice model, the choice set must meet three requirements:

1. The set of alternatives must be collectively exhaustive, meaning that the set includes all possible alternatives. This requirement implies that the person necessarily does choose an alternative from the set.
2. The alternatives must be mutually exclusive, meaning that choosing one alternative means not choosing any other alternatives. This requirement implies that the person chooses only one alternative from the set.
3. The set must contain a *finite* number of alternatives. This third requirement distinguishes discrete choice analysis from forms of regression analysis in which the dependent variable can (theoretically) take an infinite number of values.

As an example, the choice set for a person deciding which mode of transport to take to work includes driving alone, carpooling, taking bus, etc. The choice set is complicated by the fact that a person can use multiple modes for a given trip, such as driving a car to a train station and then taking train to work. In this case, the choice set can include each possible combination of modes. Alternatively, the choice can be defined as the choice of "primary" mode, with the set consisting of car, bus, rail, and other (e.g. walking, bicycles, etc.). Note that the alternative "other" is included in order to make the choice set exhaustive.

Different people may have different choice sets, depending on their circumstances. For instance, the Scion automobile was not sold in Canada as of 2009, so new car buyers in Canada faced different choice sets from those of American consumers. Such considerations are taken into account in the formulation of discrete choice models.

Prominent types of discrete choice models

Discrete choice models can first be classified according to the number of available alternatives.

- ✓ Binomial choice models (dichotomous): 2 available alternatives
- ✓ Multinomial choice models (polytomous): 3 or more available alternatives
- ✓ Multinomial choice models can further be classified according to the model specification:
- ✓ Models, such as standard logit, that assume no correlation in unobserved factors over alternatives
- ✓ Models that allow correlation in unobserved factors among alternatives

In addition, specific forms of the models are available for examining rankings of alternatives (i.e., first choice, second choice, third choice, etc.) and for ratings data.

❖ **Lecture 2: Practice on formulating experimental design for choice experiment**

Steps in Conducting a Choice Experiment

Before deciding to conduct a choice experiment, it is essential to consider whether this method is the most appropriate or whether another technique, such as contingent valuation, would be better. The essence of this decision is whether it makes sense to frame a policy question in terms of the attributes and whether marginal values of the attributes are required for policy analysis. If a policy question, for example, seeks to identify forest management options that will provide the greatest

benefit to moose hunters, then consumer choices between alternative moose hunting sites with different levels of attributes (such as moose abundance, road quality, and travel distance) provide a reasonable framework for analysis (Boxall et al. 1996). In contrast, if the policy question focuses on the value that hunters place on a moose hunting experience given current conditions, then a contingent valuation study may be a better approach (Boyle et al. 1996).

The second issue to consider is the technical composition of alternatives and the perception of attribute bundles by consumers. In the moose hunting example (Boxall et al. 1996), moose abundance, road quality, and travel distance can reasonably be considered to be independent attributes. This may not be the case for a suite of ecological characteristics that are technically linked in production (Boydand Krupnik 2009). If it is decided that a CE is the best approach for conducting policy analysis, then implementation should follow the seven steps outlined in Table 1 (based on Adamowicz et al. 1998). Each step is briefly described following the table.

2.1 Characterize the Decision Problem

The initial step in developing a CE is to clearly identify the dimensions of the problem. This requires thinking about two key issues: (1) the geographic and temporal scope of potential changes in policy attributes, and (2) the types of values that are associated with those changes. The geographic scope of a CE would include consideration of whose values are to be included in the valuation or benefit-cost analysis. If the value of a change in an endangered species management program is being considered, for example, should the CE be applied to people living in the region, province/state, country, or internationally? It is essential to identify who will be impacted by changes in policy attributes as well as to articulate how they will be impacted. In addition, if the policy context is specific to a geographic site, the location of substitute sites will be important in the design, as demonstrated in a tropical rainforest preservation study reported by Rolfe et al. (2000).

Table 1: Steps in implementing a choice experiment

Step 1: Characterize the decision problem
Step 2: Identify and describe the attributes
Step 3: Develop an experimental design
Step 4: Develop the questionnaire
Step 5: Collect data
Step 6: Estimate model
Step 7: Interpret results for policy analysis or decision support

Temporal considerations will also be important. There may be a need to include an attribute for program duration or when the benefits will accrue to the public (e.g., Qin et al. 2011).

The second issue is the type of value arising from the policy under consideration. Is the choice to be examined one that reflects use value or behavior (such as recreation site choice or choices of market goods), or is the choice best represented as a public choice (referendum) on a set of attributes arising from a policy change? The latter may contain both use and passive-use values—or it may reflect total economic value.

2.2 Attribute Identification and Description

Once the decision problem is characterized, it is necessary to identify and describe the relevant attributes, including the levels to be used for each attribute. Holding structured conversations (focus groups) with resource managers, scientists, and people who typify the population that will be sampled will help identify the important attributes. At this stage, it is often challenging to decide how many attributes to include in the experiment as well as the particular levels that each attribute can take.

Focus groups can be very useful in this case. Group members can be asked to describe what attributes they think of when considering the goods and services being affected by the policy. They can provide information on whether attributes and levels are credible, understandable, and clearly presented. Focus groups of policymakers and the public can be useful to identify whether the attributes being considered by policymakers coincide with those being evaluated by members of the public. However, focus groups will often provide long lists of attributes that could result in complex choice tasks. Because not much is known about how people respond to highly complex survey questions (Mazzotta and Opaluch 1995; Swait and Adamowicz 2001a, b), it is a good idea to keep the set of attributes and levels as simple as possible. Overall, focus groups are a very important and effective way to construct attributes, levels, and the appropriate framing of a choice task.

Describing attributes that represent passive-use values (such as the value of biodiversity conservation) can be particularly challenging. Boyd and Krupnick (2009) suggested that attributes should be thought of as endpoints that directly enter the utility functions or household production functions of consumers, or—if intermediate inputs are being considered—the pathway to the endpoint needs to be made clear. Thus, passive-use values associated with forest biodiversity, for example, can be described using indicators of species richness (Horne et al. 2005).

However, because forest biodiversity can be influenced by forest management processes that are under the control of decision-makers, attributes could be described in terms of those processes so long as the linkages between processes and outcomes are made clear. Because individuals might be interested in the processes associated with the endpoint, it is important to clarify the things that people value, what decision-makers can affect, and the description of the attributes during this stage of survey development. In addition to identifying utility endpoints, Schultz et al. (2012) recommended further standards for attributes in stated-preference studies that include measurability (endpoints are quantifiable), interpretability (endpoints can be understood by a nonscientist), and comprehensiveness (all relevant endpoints are described).

Once the attributes have been defined, attribute levels must be specified. In some cases, this is simple, such as the presence or absence of some attribute. In other cases, the assignment of levels is more difficult, such as determining the appropriate levels and ranges used to specify forest species richness (Horne et al. 2005). This issue is also faced when specifying price or cost levels. Because the price/cost attribute provides control over the key factor that determines welfare measures, it is important that this attribute can be estimated precisely in the econometric model and also be reasonable in the policy context. Much as in contingent valuation, we would like low-price/cost alternatives to be frequently purchased and high-price alternatives to be rarely purchased. 4 Price levels should not be so high or low that they do not appear to be credible, but it may be informative for prices/costs to lie outside the range of existing market prices (such as travel costs) or be reasonable costs for the provision of public programs. Pilot studies play an important role in testing price or cost levels, as well as all other attributes and levels, to ensure that they have sufficient variation to identify the parameters and to ensure that welfare measures can be calculated. These first two steps, which are critical to the successful implementation of CEs,

are often not given the due consideration they require. Practitioners are encouraged to spend significant time and effort in scoping the problem, using focus groups and pretests, and making sure the choice context and scenario descriptions are carefully developed.

2.3 Develop an Experimental Design

Once attributes and levels have been determined, the researcher must determine the number of alternatives to present in each choice set (two, three, four, etc.), and the number of choice sets to present to the respondents (one, four, eight, 16, etc.). The number of alternatives could depend on the type of value being measured and/or on the context of the study. At a minimum, choice questions should contain a status quo alternative and an alternative indicating a change from the status quo. A status quo alternative is required in each choice set so that estimated utility functions

represent changes from baseline conditions. Total value (or passive-use value) studies often employ only two alternatives because of the incentive compatibility of a two-alternative choice or referendum (Carson and Groves 2007). The number of alternatives in some studies depends on the number of alternatives that occur in the real world.

The number of choice questions to ask depends in part on the complexity of the task and is often a judgment the researcher must make based on focus groups, pilot tests, and expert judgment. In general, the number of choice sets included in the design depends on the number of degrees of freedom required to identify the model. The use of multiple-choice sets can also have implications for incentive compatibility (Carson and Groves 2007, 2011).

Experimental design procedures are used to assign attribute levels to the alternatives that form the basis for choices and to construct the sets of choices that will be presented to respondents. Alternatives presented to the respondents must provide sufficient variation over the attribute levels to allow one to identify preference parameters associated with the attributes. In most cases, presenting all combinations of attributes and levels will be impossible. Thus, experimental design procedures are used to identify subsets of the possible combinations that best identify attribute

preferences. Because of the importance of this topic to the success of any CE (Scarpa and Rose 2008), it is discussed in detail in Sect. 5.3.

2.4 Questionnaire Development

As with other stated-preference methods, CEs involve surveys, and various questionnaire formats can be used for collecting data (see Chap. 3), including:

- Mail-out, mail-back surveys.
- Telephone recruitment, mail-out, mail-back surveys.
- Telephone recruitment, mail-out, telephone surveys.
- Computer-assisted surveys at centralized facilities or in person.
- Intercept surveys that could be paper and pencil or computer-assisted.
- Internet-based surveys, including Internet panels.

The selection of the questionnaire format is usually based on pragmatic concerns, such as availability of a sample frame and budget limitations. In the case of CEs, Internet modes, particularly Internet panels, are becoming increasingly popular. Because CEs present respondents with complex sets of choice questions and randomization of the order of these questions is desirable, mail and telephone surveys can be more difficult to use relative to Internet or computer-based in-person surveys (e.g., using tablets to collect information from respondents). Also, in some

cases information from early parts of a survey is used in the design of attributes and/or levels in the choice tasks, making computer-based Internet or in-person surveys more convenient.

2.5 Data Collection

Data collection should be carried out using the best survey practices (e.g., Dillman 1978). Chapter 4 outlines a number of issues in data collection for contingent valuation studies that apply as well to the implementation of CEs. One unique feature arising in CEs is that multiple choice sets are presented to individuals with the intent that choice sets be considered independently and without comparing strategically across choice sets. This means that it is desirable to prevent respondents from reading ahead or going back and changing responses. It is also valuable to randomize the order of the presentation of the choice sets so that the first task, in a large enough sample, can be used to estimate values that are not affected by repeated choices. In a mail survey (paper and pencil), this is very difficult to accomplish because respondents can flip through the survey booklet. Computer-based surveys (Internet and in-person) can achieve this through the design of the survey implementation program. Computer-based methods also capture the amount of time spent on each question, which tells researchers if respondents are taking time to consider the choice set carefully.

2.6 Model Estimation

Once data have been collected, the next step is to estimate preference parameters using a random utility model. A growing number of econometric specifications have been used to analyze choice data. These models typically vary over how the error term is interpreted, particularly in the context of heterogeneity in preferences across respondents.

2.7 Policy Analysis and Decision Support

Most CE applications are targeted to generating welfare measures (see Sect. 5.5), predictions of behavior, or both. These models are used to simulate outcomes that can be used in policy analysis or as components of decision support tools. CEs provide the opportunity to evaluate the welfare effects of multiple policy options involving combinations of attributes and levels. They also allow for calibration to actual policies or outcomes when these conditions become known. For example,

choice experiments on park visitation have been calibrated using actual visitation information when measuring nonmarginal welfare impacts (Naidoo and Adamowicz 2005). As such, they can provide a richer set of policy information than most other valuation approaches.

- ❖ **Lecture 3: Exercise on developing choice card**

Theoretical motivation

3.1 The choice experiment method

Individuals are traders. They consciously or sub-consciously make decisions by comparing alternative and selecting an action which is known as a choice outcome. This study will draw on ideas from economics and psychology perspective, starting with the notion that it is an individual's preference for specific alternatives that best determine what alternative is chosen. The overall utility associated with the i^{th} alternative can be divided into the contributions that are observed by the researcher and those that are not observed by the researcher. Suppose these sources of relative utility represent as V_i and ε_i . V_i is the deterministic portion of the utility and ε is common notation which is used to refer to the unobserved influences as error or random error term. In choice analysis, both V_i and ε_i have great relevance. It is assumed that there is a strong relationship between V_i and ε_i . These two components are independent and additive. A utility function is strongly additive if it can be written as

$$U = \sum_{i=1}^{n} f_i(q_i) \tag{1}$$

where f_i are increasing. Additive is a special case of separability[2]. Any utility function that has a monotonic transformation[3] which is additive may be treated as being additive for all theorems applicable to additive functions (Henderson and Quandt 1980). An additive utility function has the property that all cross partials equal zero, i.e.

$$\delta^2 U / \delta q_i \, \delta q_j = 0 \text{ for all } i \neq j \tag{2}$$

It will take the form under the strict quasi-concavity condition and the two-variable case as

$$f_{11} f_2^2 + f_{22} f_1^2 < 0 \tag{3}$$

A behavioral choice rule can be explained by the Lancasterian theory of value and random utility theory (RUT). The following part briefly explains these two issues.

3.1.1 Lancasterian theory of value

Lancaster (1966) asserted that the good does not give utility to the consumer, it possesses characteristics and these characteristics give rise to utility; a good will possess more than one characteristic and these will be shared by more than one good and Goods in combination may possess characteristics different from those pertaining to the goods separately. Assumed that an individual good or a collection of goods as a consumption activity and associate a scalar with it. It is also assumed that the relationship between the level of activity k and y_k and the goods consumed in that activity to be linear and objective, so that, if x_j is the j^{th} commodity

$$x_j = \sum_k a_{jk} y_k \tag{4}$$

---

[2] A utility function is strongly separable in all of its arguments if it can be written as $U = F\left[\sum_{i=1}^{n} f_i(q_i)\right]$

[3] A (positive) monotonic transformation is a way of transforming one set of numbers into another set of numbers so that the rank order of the original set of numbers is preserved. It is thus a function, f , mapping real numbers into real numbers which satisfies the property that if $x > y$, then $f(x) > f(y)$

with activity vector

$$x = Ay \tag{5}$$

Since the relationships are assumed to be objective, the equations are assumed to hold for all individuals, the coefficients a_{jk} being determined by the intrinsic properties of the goods themselves. It is also assumed that each consumption activity produces a fixed vector of characteristics and that relationship is again linear, so that, if z_i is the amount of the i^{th} characteristic

$$z_i = \sum_k b_{ik} y_k \tag{6}$$

with activity vector like equation (5)

$$z = By \tag{7}$$

Again, it is assumed that the coefficients b_{ik} are objectively determined for some arbitrary choice of the units of z_i.

It is assumed that the individual possesses an ordinal utility function on characteristics U_z and that he will choose a situation which maximizes U_z . U_z is provisionally assumed to possess the ordinary convexity properties of a standard utility function. The chief purpose of making the assumption of linearity is to simplify the problem. A viable model could certainly be produced under the more general set of relationships.

$$F_k(z, x) = 0; \quad k = 1 \ldots\ldots\ldots\ldots\ldots m \tag{8}$$

In this model, the relationship between the collections of characteristics available to the consumer-the vectors z-which are the direct ingredients of his preferences and his welfare and the collections of goods available to him-the vector x-which represent his relationship with the rest of the economy, is not direct and one-to-one, as in the traditional model, but indirectly, through the activity vector y (Lancaster 1966).

3.1.2 Random utility theory (RUT)

The concept of random utility theory (RUT) and the random service theory (RST) are almost the same which plays an important role to explain consumer behavior. RUT says that not all of the determinants of utility derived by individuals from their choices is directly observable to the researcher, but that an indirect determinant of preferences is possible (McFadden 1974; Manski 1977). The utility function for a representative consumer can be decomposed into observable and stochastic sections:

$$U_{an} = V_{an} + \varepsilon_{an} \tag{9}$$

where U_{an} is the latent and unobservable utility held by consumer *n* for choice alternative *a*, V_{an} is the systemic or an observable portion of utility that consumer *n* has for choice alternative

a and ε_{an} is the random or unobservable portion of the utility that consumer n has for choice alternative a. Research is focused on a probability function, defined over the alternatives which an individual faces, assuming that the individual will try to maximize his utility (Bennett and Blamey 2001; Louviere et al. 2000). This probability is expressed as

$$P(a/C_n) = P\left[(V_{an} + \varepsilon_{an}) > (V_{jn} + \varepsilon_{jn})\right] \quad \forall_a \neq j \tag{10}$$

for all j options in choice set C_n, a and n are also described as

$$P(a/C_n) = P\left[(V_{an} - V_{jn}) > (\varepsilon_{jn} - \varepsilon_{an})\right] \quad \forall_a \neq j \tag{11}$$

Equation (11) holds the principle of RUT which exhibits the stochastic components are independently and identically distributed (IID) with a Gumbel or Weibull distribution. This leads to the use of multinomial (MNL) or conditional logit (CL) or basic model. It helps to determine the probability of choosing a over j options (Hanley et al. 2001; Alpizar et al. 2001).

The estimated deterministic (indirect) utility function generally will have the following form:

$$P\left(U_{an} > U_{jn}\right) = \frac{\exp(\mu V_a)}{\sum j \exp(\mu V_j)} \quad \forall_a \neq j \tag{12}$$

Here, μ is a scale parameter, inversely related to the standard deviation of the error term and not separately identified in a single data set (Bergmann et al. 2006). The implications of this are that the estimated β values cannot be directly interpreted as to their contribution to utility, since using the MNL model choices must satisfy the independence from irrelevant alternatives (IIA) assumption, meaning that the addition or subtraction of any option from the choice set will not affect the relative probability of individual n choosing any other option (Louviere et al. 2000; Bergmann et al. 2006). Modeling constants known as alternative specific constants (ASCs) are typically included in the MNL model. The ASC accounts for variations in choices that are not explained by the attributes or socio-economic-demographic variables and sometimes for a status quo bias (Ben-Akiva and Lerman 1985).

The random parameter logit or extended model provides a simple way to generalize the multinomial logit model to permit the utilities of each alternative to be correlated and it does not require IIA assumption (Cameron and Trivedi 2005). The random utility function in the random parameter logit model will take the following form (Birol et al. 2005).

$$U_{in} = V_{in} + \varepsilon_{in} \equiv Z_i(\beta + \eta_n) + \varepsilon_{in} \tag{13}$$

As we know utility is decomposed into a non-random component (V) and a stochastic term (ε) and the indirect utility is assumed to be a function of the choice attributes Z with parameters β and SED variables (Agimass and Mekonnen 2011). Hence, the probability of choosing alternative i in each of the choice set will have the following form (Birol et al. 2005):

$$P_{in} = \exp(Z_{in}(\beta + h_n)) \Big/ \sum \exp(Z_{jn}(\beta + h_n)) \tag{14}$$

3.2 Choice of attributes and their levels

The first step in our study was to choose the attributes and their associated levels. The reduction of pesticide use by farmers can have many drivers and consequences, depending on context, e.g., if this re-duction is associated with the adoption of agroecological practices, the conversion to organic farming, or the participation in an agri-environ-mental scheme. Such a change can result in monetary gains due to a reduction of input costs, an increased sales price, or subsidies. It can

produce non-monetary outcomes, such as the improvement of farmers' public image, participation in a network, the improvement of farmers' quality of life and health, and improved quality of the environment. It can also have negative outcomes, such as reduced yields, increased risk, the necessity to train to learn new agricultural techniques. As Hanley et al. (2002) explains, the number of attributes considered in a DCE must be limited in order to avoid the cognitive burden of making choices that are too complicated. The selection of the attributes was based on (i) the literature, (ii) discussions with experts in agronomy, epidemiology, ecology, and agricultural economics, (iii)

focus groups of farmers 6, and (iv) pretests on the choice sets. 7 The focus groups and pretests revealed that pesticides are a sensitive topic among the French farming community; thus, we were careful with the employed terms and their potential interpretations. We were also careful to choose attributes that are adapted to different types of farming systems, while remaining concrete for farmers.

As shown in Table 2, the chosen attributes are as follows:

1. The farmer's yearly profit (or gross margin) per hectare, expressed in comparison with the status quo. This average profit per hectare per year, in euro, is the monetary (or cost) attribute. The profit

varies with changes in agricultural practice, owing to unspecified factors such as the impact on yields, pesticide expenses, public aid (e.g., subsidies), sales price, and so on. Therefore, the farmer's profit can increase or decrease with a reduction of pesticides. Following our discussions with experts and the focus groups, this attribute was given the following possible values: −50 €, +0 €, +50 €, +100 €.

2. The production risk, formalized as the frequency of years (number of years out of 10) in which production is drastically and exceptionally reduced owing to pests (i.e., more than 30% of production is lost or damaged owing to diseases, insects, weeds, and so on). This attribute characterizes the main effect of the reduction of pesticides on the variability of production, independently of the level of production or profit (the mean yearly profit is given by the previous attribute). The production risk attribute is expressed in additional years out of 10 (+0, +1 year, or +2 years), compared with the status quo. These levels were set after discussions with experts (farmers, agronomists, and epidemiologists).

Table 2: Selection of attributes and levels.

Attribute	Description	Levels
Profit	Variation in the average yearly profit per hectare	−50 €; + 0 € (SQ);
Production risk	Variation in the number of years, out of 10 years, with exceptionally large production losses	0 year (SQ); +1 year; +2 years
Administrative commitment	Administrative framework of the change of practice, if any	None (SQ); Charter; Contract; Certification
Health and environmental impacts	Reduction in exposure to harmful substances	−0% (only SQ); −20%; −50%; −80%

SQ: level in the status quo (also possible in the other options). only SQ: level only possible in the status quo option

The administrative framework of the change in practice describes whether the change accompanies an administrative commitment. A change of agricultural practices inducing a reduction in pesticide

use may be included as part of an administrative framework. Such a framework can be perceived positively, because it may imply better-valued products, or integration in a network; however, it may also include an administrative burden and, thus, be perceived negatively. This attribute is qualitative, and is expressed as additional commitment over and above the status quo, as follows: “No additional administrative commitment,” “charter” (inducing no contractual specification and flexible commitment), “agri-environmental contract with public authorities” (with specification, and possibly a subsidy), and a “certification process” (with a specification, controls, and a green label, possibly inducing higher sales prices). The potential subsidy or higher sales prices are included in the level of profit given in the first attribute. Only non-monetary aspects of the administrative commitment are included in the administrative commitment attribute.

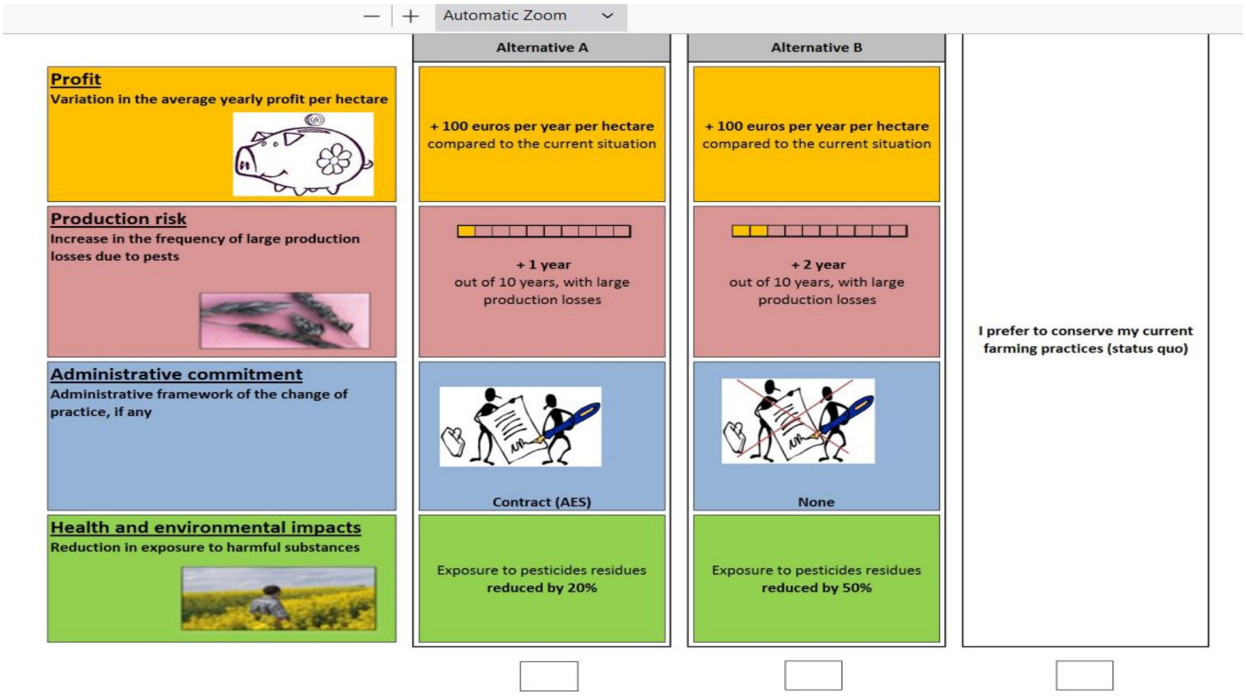
4. The health and environmental impacts indicate the reduction in exposure to harmful substances as a result of the change in practice. This includes the local and global environmental quality (biodiversity, water quality) and the health of farmers, neighbors, and general population. This attribute takes the following values: −0% (status quo only), −20%, −50% −80%, compared with the status quo.

Adding an attribute to encompass production risk helps to increase the credibility of valuation scenarios and reduces hypothetical bias (Rolfe and Windle, 2015). However, the concept of risk is difficult to express as an attribute in a way that is convenient and understandable to respondents. Whereas a mean value expressed as an average is easy to understand by respondents, other scientific terms used to describe a probability distribution, such as variance or standard deviation (or worse, skewness and kurtosis), are poorly understood by the public.

Jaeck and Lifran (2014) expressed their risk attribute as the frequency of below-average yields (zero, one, or three years over five years). This formulation is clear, but it does not allow us to convey the idea of a risk of large production loss due to pests. We wanted to capture the idea that pesticide reduction may induce a larger variability of production, along with an increase in the occurrence of pest attacks resulting in exceptionally large production losses. Discussion within the focus groups confirmed that this was a realistic outcome in the event of low or no pesticide use. We thus opted for the frequency of years with large damages and production losses, for a given mean profit (given by the first attribute). Our production risk attribute is related to the variability of the losses due to pests, but not to the mean yield or mean profit.

Consequently, the profit attribute and the risk attribute are independent. Various tests show that the proposed formulation offers an easy way to express production variability due to an increase of extreme losses. For the "health and environmental impacts" attribute, we first considered having two separate attributes for health and for the environment. We finally chose to group them, because both are highly correlated (Juraske et al., 2007) and we were limited in the number of attributes. In addition, we initially wanted to express this attribute as a reduction of the treatment frequency index (TFI), a crop- and region-normalized indicator of pesticide use, widely used and understood by European farmers. However, pretests revealed that this formulation induced misinterpretations and acceptability problems from farmers who perceived it as a technical objective to be achieved. Whenever farmers believe that achieving the proposed reduction is not possible for their farms, they opt for the status quo. Because we wanted to value here the environmental and health impacts of the agricultural practice, rather than the constraints it implies (captured by other attributes), we

opted for this formulation.



Lecture 4 and 5: Practicing data analysis for choice experiment

Methods – Survey Design

The Survey Instrument

Section 2. Survey – Choice Question 1

Suppose Option A and Option B were the **only** grassland projects you could choose. Which **one** would you choose? Please read **all** the features of **each** option and then **check the box that represents your choice**. If you do not like either option A or option B, then please choose the box marked "No grassland project" which is Option C.

Attribute	Number of Bird Species	Density of Birds	Number of endangered species	Amount of wildflowers.	Use of prescribed fire.	Distance to restored area	Annual cost to your household	I would Choose
Option A	30 different species	10 individuals per acre	3 endangered species	20% covered in wildflowers	Prescribed burning once a year	100 miles	$30	❒ A
Option B	10 different species	15 individuals per acre	6 endangered species	60% covered in wildflowers	No burning	10 miles	$50	❒ B
Option C	No Restoration Project						No cost	❒ C

5

2

Methods – Data Collection/Entry

Data Entry

- When we analyze the data, each choice option should be a row of data.
- All the data from each survey instrument can be entered in this way
  - This is very tedious and time consuming
- Alternative is to enter respondent information and choices per choice question in one row
  - Use the experimental design and the survey version to generate a data sheet

3

Methods – Data Collection/Entry

Data as entered originally (one row per respondent)

Survey No   Survey Version Identifier

Suv No	Sur Ver ID	CEQ1	CEQ2	CEQ3	CEQ4	CEQ5	CEQ6	Q1	Q2	Q3	Q3.1	Q4	Q4.1	Q5	Q6 (age)
1	85	B	A	A	A	A	A	N	N	Y	1	N		NS	22
2	71	A	B	B	A	A	A	Y	N	Y	5	Y	MANY	Y	21
3	85	B	B	B	B	B	B	Y	N	Y	2	Y	5	Y	21
4	85	B	A	A	A	A	B	Y	N	Y	7	Y	10	Y	21

Respondent choices (for each choice question)   Answers to socio-economic questions

Data after converting to fit to mixlogit (STATA) and LimDep (one row per alternative, i*k*q rows)

Survey Version Identifier   Attribute values (that correspond to survey choices)

Suv ver id	resp id	cno	alt	cset	alt1	alt2	alt3	c_id	choice	number_b	density_b	endangered_b	wildflowers	burning	distance	cost	Q1 (Recyc	Q2 (Bird W
85	1	1	1	3	1	0	0	1	0	50	10	0	20	2	100	30	N	N
85	1	1	2	3	0	1	0	1	1	25	20	10	60	0	50	60	N	N
85	1	1	3	3	0	0	1	1	0	0	0	0	0	0	10000	0	N	N
85	1	2	1	3	1	0	0	2	1	50	20	10	60	0	100	0	N	N
85	1	2	2	3	0	1	0	2	0	75	5	0	40	1	50	60	N	N
85	1	2	3	3	0	0	1	2	0	0	0	0	0	0	10000	0	N	N
85	1	3	1	3	1	0	0	3	1	25	20	5	20	2	10	0	N	N
85	1	6	2	3	0	1	0	6	0	50	20	10	40	0	100	60	N	N
85	1	6	3	3	0	0	1	6	0	0	0	0	0	0	10000	0	N	N
71	2	1	1	3	1	0	0	7	1	25	20	10	20	1	50	30	Y	N
71	2	1	2	3	0	1	0	7	0	50	5	5	40	2	10	60	Y	N
71	2	1	3	3	0	0	1	7	0	0	0	0	0	0	10000	0	Y	N
71	2	2	1	3	1	0	0	8	0	25	20	0	60	0	50	30	Y	N

Survey No   Option (alternative) identifier   Respondent choice

Choice No (by survey)   Number of choices   Choice set identifier (across all surveys)   Answers to demographic questions

[1] i = number of respondents, q = number of choice decisions in each survey, k = number of alternatives in each choice decision

Methods –Model- Estimation

CONDITIONAL LOGIT

$$U_j = \sum_{k=1}^{K} \beta_k x_{kj} + \beta_p p_j + \varepsilon_j$$

- **Homogeneous** utility for alternative *j* with *k* attributes
- The marginal value of attribute *k* is the ratio between the parameter β_k and $-\beta_p$.

$$MWTP_k = -\frac{\beta_k}{\beta_p}$$

MIXED MULTINOMIAL LOGIT

$$U_j^i = \sum_{k=1}^{K} \beta_{ki} x_{kj} + \beta_{pi} p_{ij} + \varepsilon_{ij}$$

- Utility for individual *q* choosing alternative *j* with *k* attributes
- Models preference **heterogeneity**; deals well with repeated choices

15

Methods –Model- Estimation

CONDITIONAL LOGIT

$$U_j = \sum_{k=1}^{K} \beta_k x_{kj} + \beta_p p_j + \varepsilon_j$$

- **Homogeneous** utility for alternative *j* with *k* attributes
- The marginal value of attribute *k* is the ratio between the parameter β_k and $-\beta_p$.

$$MWTP_k = -\frac{\beta_k}{\beta_p}$$

MIXED MULTINOMIAL LOGIT

$$U_j^i = \sum_{k=1}^{K} \beta_{ki} x_{kj} + \beta_{pi} p_{ij} + \varepsilon_{ij}$$

- Utility for individual *q* choosing alternative *j* with *k* attributes
- Models preference **heterogeneity**; deals well with repeated choices

15

Methods – Model Estimation

- Model Specification

$$V_{qi} = \beta_1 Z_{richness} + \beta_2 Z_{density} + \beta_3 Z_{endangered} + \beta_4 Z_{wildflowers} + \beta_5 Z_{burning} + \beta_6 Z_{distance} + \beta_7 Z_{cost} + \varepsilon_{qi}$$

Conservation success interaction terms

$$+\beta_8 Z_{richness} * Z_{density} + \beta_9 Z_{density} * Z_{endangered} + \beta_{10} Z_{endangered} * Z_{richness}$$

Public good near? interaction terms

$$+\beta_{11} S_{grassland\ near?} * X_{cost} + \beta_{12} S_{nature\ near?} * X_{cost}$$

Methods – Model Estimation

The Data Set Up

Data after converting to fit to mixlogit (STATA) and LimDep (one row per alternative, i*k*q rows)

Survey Version Identifier

Attribute values (that correspond to survey choices)

Suv ver id	resp id	cno	alt	cset	alt1	alt2	alt3	c_id	choice	number_b	density_b	endangered_b	wildflowers	buming	distance	cost	Q1 (Recyc	Q2 (Bird W
85	1	1	1	3	1	0	0	1	0	50	10	0	20	2	100	30	N	N
85	1	1	2	3	0	1	0	1	1	25	20	10	60	0	50	60	N	N
85	1	1	3	3	0	0	1	1	0	0	0	0	0	0	10000	0	N	N
85	1	2	1	3	1	0	0	2	1	50	20	10	60	0	100	0	N	N
85	1	2	2	3	0	1	0	2	0	75	5	0	40	1	50	60	N	N
85	1	2	3	3	0	0	1	2	0	0	0	0	0	0	10000	0	N	N
85	1	3	1	3	1	0	0	3	1	25	20	5	20	2	10	0	N	N

Survey No

Option (alternative) identifier

Respondent choice

85	1	6	2	3	0	1	0	6	0	50	20	10	40	0	100	60	N	N
85	1	6	3	3	0	0	1	6	0	0	0	0	0	0	10000	0	N	N
71	2	1	1	3	1	0	0	7	1	25	20	10	20	1	50	30	Y	N
71	2	1	2	3	0	1	0	7	0	50	5	5	40	2	10	60	Y	N
71	2	1	3	3	0	0	1	7	0	0	0	0	0	0	10000	0	Y	N
71	2	2	1	3	1	0	0	8	0	25	20	0	60	0	50	30	Y	N

Choice No (by survey)

Number of choices

Choice set identifier (across all surveys)

Answers to demographic questions

[1]i = number of respondents, q = number of choice decisions in each survey, k = number of alternatives in each choice decision

Methods – Model Estimation

```
*CLOGIT

clogit  choice richness density endangered wildflowers burning distance cost, group(ci

. clogit  choice richness density endangered wildflowers burning distance cost,
> group(cid)

Iteration 0:   log likelihood = -1624.6134
Iteration 1:   log likelihood = -1612.6263
Iteration 2:   log likelihood = -1612.5788
Iteration 3:   log likelihood = -1612.5788

Conditional (fixed-effects) logistic regression   Number of obs   =       4950
                                                  LR chi2(7)      =     400.26
                                                  Prob > chi2     =     0.0000
Log likelihood = -1612.5788                       Pseudo R2       =     0.1104
```

Never Copy Stata Tables into a presentation or paper

choice	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
richness	.0161617	.0038973	4.15	0.000	.0085231	.0238004
density	.025141	.0077457	3.25	0.001	.0099596	.0403223
endangered	.1329731	.0136463	9.74	0.000	.1062269	.1597193
wildflowers	.012059	.0018792	6.42	0.000	.0083758	.0157422
burning	-.0099971	.0406831	-0.25	0.806	-.0897344	.0697402
distance	-.0045319	.0009127	-4.97	0.000	-.0063208	-.0027431
cost	-.0147642	.0012305	-12.00	0.000	-.0171759	-.0123524

28

Methods – Model Estimation

- Storing and Saving Estimates

```
estimates store grassland_ce_cllogit
estimates save grassland_ce_cllogit
```

Methods – Model Estimation

```
*MIXLOGIT

global randvars "richness density endangered wildflowers burning distance cost"
mixlogit choice, rand ($randvars) group(cid) id(id) nrep(100)
```

```
Iteration 13:  log likelihood = -1219.6384
Iteration 14:  log likelihood = -1219.6384

Mixed logit model                              Number of obs   =       4950
                                               LR chi2(7)      =     785.88
Log likelihood = -1219.6384                    Prob > chi2     =     0.0000
```

choice	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Mean						
richness	.0307242	.0095429	3.22	0.001	.0120205	.0494279
density	.0938939	.0174795	5.37	0.000	.0596346	.1281532
endangered	.3058226	.0377027	8.11	0.000	.2319266	.3797186
wildflowers	.0314274	.0039477	7.96	0.000	.0236901	.0391648
burning	.1379765	.0881055	1.57	0.117	-.0347071	.3106602
distance	-.0097515	.0022722	-4.29	0.000	-.014205	-.0052981
cost	-.041074	.0043121	-9.53	0.000	-.0495256	-.0326224
SD						
richness	.0898179	.0126366	7.11	0.000	.0650506	.1145853
density	-.1010069	.0210981	-4.79	0.000	-.1423584	-.0596554
endangered	.3983721	.0462853	8.61	0.000	.3076546	.4890896
wildflowers	-.0110484	.0064812	-1.70	0.088	-.0237513	.0016544
burning	.7257335	.1231887	5.89	0.000	.4842881	.967179
distance	.0241212	.0029294	8.23	0.000	.0183797	.0298627
cost	.0540639	.0052452	10.31	0.000	.0437836	.0643442

The sign of the estimated standard deviations is irrelevant: interpret them as

Never Copy Stata Tables into a presentation or paper

30

Methods – Model Estimation

- Storing and Saving Estimates

```
estimates store grassland_ce_mixlogit
estimates save grassland_ce_mixlogit
```

- Creating results table with *esttab*

```
*esttab command needs to be installed
* findit esttab
* then click install for st0085_2

esttab grassland_ce_cllogit grassland_ce_mixlogit using grassland_estimates.rtf,
replace se label nogaps onecell star(* 0.1 ** 0.05 *** 0.01) scalars("ll Log lik."
"chi2 Chi-squared")
```

Methods – Model Estimation

- Generating coefficient plots

You can use the saved estimates to generate plots of the coefficients estimates using coefplot command in Stata.

Needs to be installed with “ssc install coefplot”

```
coefplot (grassland_ce_clWTP, label(CLOGIT) pstyle(p2)) (grassland_ce_mixWTP, label(
MIXLOGIT) pstyle(p3))  , msymbol(S) yline(0) levels (90) vertical recast(bar) barwidth(0.25
) fcolor(*.5) ciopts(recast(rcap)) citop coeflabel (richness = "Richness" density="Density"
 endangered="Endangered" wildflowers="Wildflowers" burning="Burning" distance="Distance",
wrap (9) nobreak) ytitle (WTP Values ($)) title(WTP for Grassland Attributes)
graph save Graph "grasslandWTP.gph", replace
graph export "grasslandWTP.png",  as(png) replace
```

Navigating Impact: Basic Concepts, Treatment Effects, and Academic Writing Essentials

Dr. Md. Sadique Rahman
Professor
Department of Agricultural Finance and Management
Sher-e-Bangla Agricultural University, Dhaka-1207

❖ **Lecture 1: Concept of impact evaluation and application of Heckman's treatment effect model**

What is impact evaluation?

- An impact evaluation relies on rigorous methods to determine the changes in the well-being of individuals, households, or communities which can be attributed to a specific intervention based on cause-and-effect analysis.
- These observed changes can be positive and negative, intended and unintended, direct and indirect.

Impact Evaluation Answers 3 Questions

- ▶ What was the effect of the program on outcomes?
- ▶ How much better off are the beneficiaries because of the program/policy?
- ▶ How would outcomes change if changed program design?

3 points need to be considered:

- Proper counterfactual
- Sample selection bias
- Endogeneity

Counterfactual: what would have happened without the program

- ➢ Counterfactual is key to impact evaluation
- ➢ Treated & counterfactual characteristics:
    - have identical characteristics,
    - except for benefiting from the intervention

Sample Selection bias: People choose to participate for specific reasons.

- ▶ Purposive program placement: Ex. Govt programs
- ▶ Self-selection into the program: self-selection could be based on observed characteristics, unobserved characteristics or both.

Endogeneity:

- ❑ Omitted variable bias from a variable that is correlated with but is unobserved, so cannot be included in the regression;
- ❑ Simultaneous causality bias;

Example: Police →crime; Crime → police

- ❑ Errors-in-variables bias (is measured with error)

Example: (e.g. salary) depends on X* (e.g. intelligence)

Example:

In general, the benefit from a program can be estimated as follows:

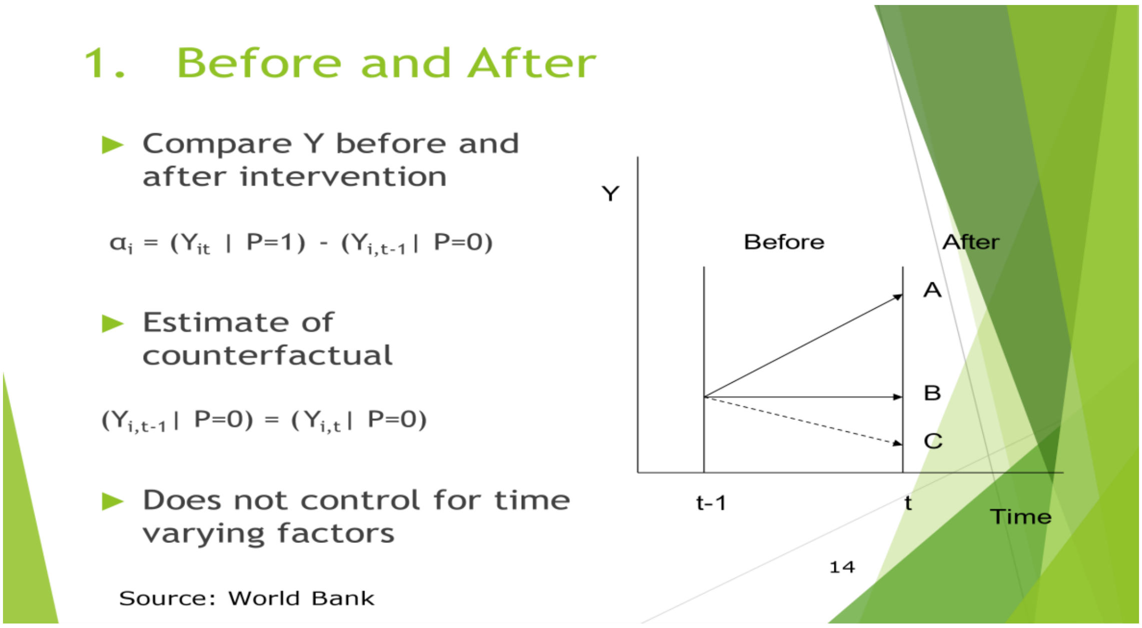
$Y_i = \beta X_i + \gamma T_i + v_i$

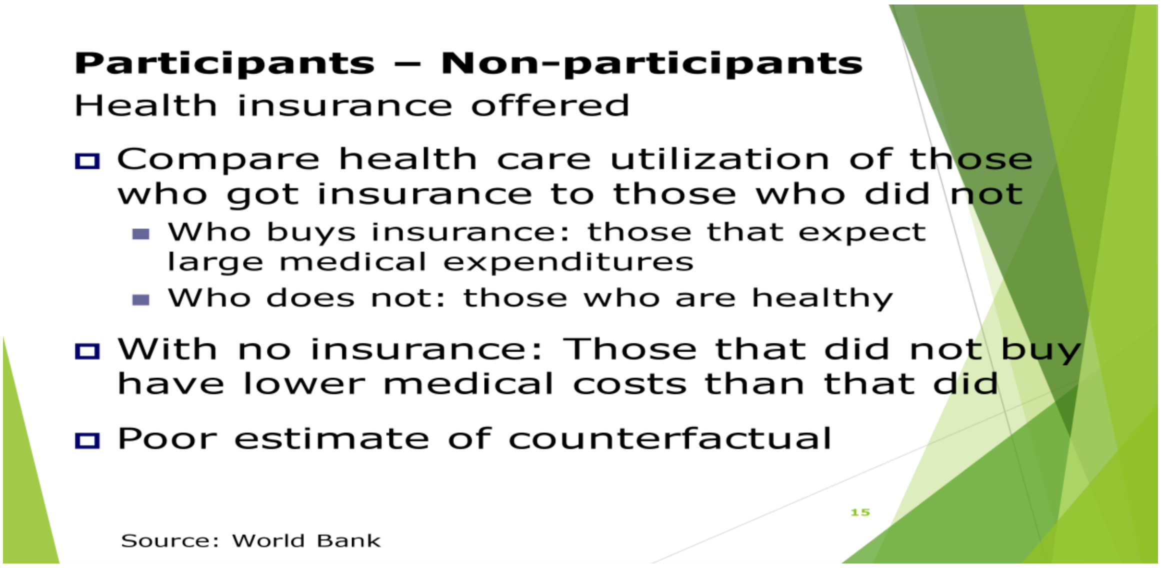
Y_i is outcome variable (for example income), Ti = Treatment variable representing technology adoption status (1 = adopters, 0 otherwise), X_i is the independent variables, v_i is the error term.

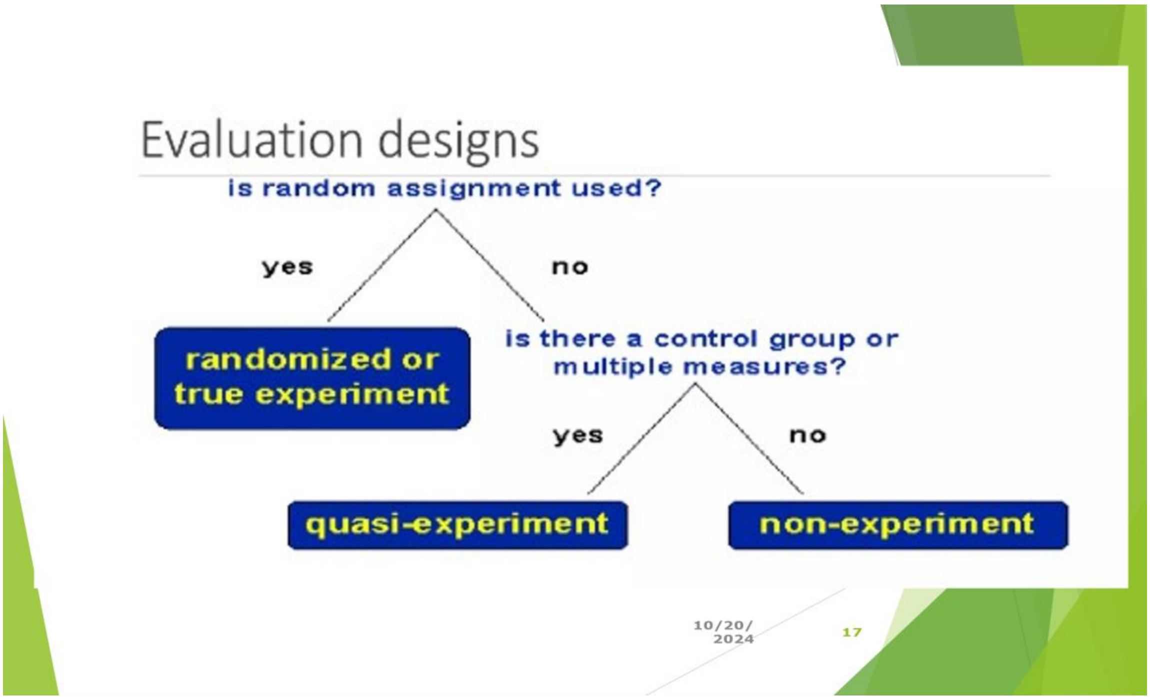
OLS can give unbiased estimates if there is no selection bias/ Endogeneity. But most cases it is not very straight forward.

Analytical approaches:

1. Before – After
2. Participant – non-participants







Treatment Effect model:

(https://economics.mit.edu/sites/default/files/publications/Treatment%20Effects.pdf)

treatment effects

The term 'treatment effect' refers to the causal effect of a binary (0–1) variable on an outcome variable of scientific or policy interest. Economics examples include the effects of government programmes and policies, such as those that subsidize training for disadvantaged workers, and the effects of individual choices like college attendance. The principal econometric problem in the estimation of treatment effects is selection bias, which arises from the fact that treated individuals differ from the non-treated for reasons other than treatment status per se. Treatment effects can be estimated using social experiments, regression models, matching estimators, and instrumental variables.

average causal effects in the treatment effects context are the average treatment effect (ATE), $E[Y_{1i} - Y_{0i}]$, and the average treatment effect on the treated (ATET), $E[Y_{1i} - Y_{0i} | D_i = 1]$. Note that the ATET can be rewritten

$$E[Y_{1i} - Y_{0i} \mid D_i = 1] = E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 1].$$

Heckman's endogenous treatment effect model

Assume, a continuous outcome variable (Y_i), which is the function of a binary treatment dummy variable, T_i, representing the adoption status of household i (T_i = 1 if solar PV adopted, 0 otherwise).

$Y_i = \beta X_i + \gamma T_i + v_i$

Adoption status might not be random, and the treatment variable is considered endogenous—it can be influenced by both simultaneous causality and unobservable individual characteristics, which, in turn, affect the treatment and outcome variables, T_i and Y_i.

❖ **Lecture 2: Application of treatment effect models using STATA software**

Heckman's endogenous treatment effect model (Heckman 1976, 1978), which is an extension of the Heckman two-stage model.

The main difference is that in the extended model, the dependent variable in the selection equation becomes an explanatory variable in the outcome equation.

Heckman's endogenous treatment effect model can only be used when the correlation between the two error terms is greater than zero.

Table 4. Impact of solar PV adoption via second stage of endogenous treatment effect model

Variable	Income		Poverty gap	
	ATT	SE	ATT	SE
Adoption of solar PV	424*	242	−0.157***	0.052
Control variable	Yes		Yes	
Rho	−0.112*	0.06	0.030*	0.017
Wald test	3.03*		2.32*	
Mean variance inflation factor	1.24			

The statistically significant rho values indicate the presence of selection bias in the sample. The Wald test result indicates that we can reject the null hypothesis of no correlation between the selection and outcome equation error terms and, thus, justify the use of Heckman's endogenous TE model .

Command

etregress outcome variable independent variables, **treat** (**treantment variable** = independent variables with instrument)

Example

etregress loginc Age i.hedu i.sedu occupa saving houseown Tv farmha internet electricity dis_Bazar credit, treat(solaradopt = Age i.hedu i.sedu occupa saving houseown Tv farmha mobile internet electricity dis_Bazar credi eleccostUSD)

❖ **Lecture 3: Journal article writing tips**

Manuscript Writing

Think about why you want to publish your work and whether it's publishable?

Ask yourself;

Have I done something new and interesting?
Is my work related directly to a current hot topic?

If all answers are "yes" then you have a difficult job in your hand.....
If any of the responses are "no" you can probably submit your paper to a **local journal but not predatory journals**

General Principles

The article text follows the IMRaD format, which responds to the questions below:

Introduction: What did you/others do? Why did you do it?

Methods: How did you do it?

Results: What did you find?

And

Discussion: What does it all mean?

The main text is followed by the Conclusion, Acknowledgements, References and Supporting Materials.

Length of your manuscript

Generally 25-35 pages, double spaced, auto line numbering ON

Title Page: Title and authors information

Abstract: 1 paragraph (<250 words)

Introduction: including review 1.5-2 pages

Methods: 2-3 pages

Results: 6-8 pages

Discussion: 4-6 pages

Conclusion: 1 paragraph

Figures: 1- 5

Tables: 1- 6

References: 20-50 papers (2-4 pages)

Literature Review

Without proper review you can’t anticipate acceptance.
A through, sophisticated literature review is the foundation for substantial, useful research – **Why?**

1. Distinguishing what has been done from what need to be done.
2. Discovering important variables relevant to the topic
3. Identifying main methodologies that have been used

Write a convincing Introduction

Tell your story……This is your opportunity to convince readers that you clearly know why your work is useful

Additional tips:

- Never use more words than necessary (be concise and to-the-point). **Don't make this section into a history lesson. Be specific what is the problem to be solve.**
- The introduction must be **organized from the global to the particular point of view**, guiding the readers to your objectives when writing this paper.
- **Hypothesis and objectives** must be clearly remarked at the **end of the introduction**.
- **Expressions** such as "novel," "first time," "first ever,“ are **not preferred.**
- **Paragraphs must be connected.**

<u>Selection of Suitable Journal</u>

<u>Few Tips:</u>

- Journal website
- Journals Aims and scope
- ***Turnaround***: A chief factor governing journal selection is turnaround time. This "speed" metrics might be listed by some journals (in days or weeks), as follows:

Average time for an article to be reviewed

Average time from submission to first post-review decision

Average time from acceptance to first online appearance/publication

10/20/2024 38

Other specifics: Miscellaneous journal metrics that help an author in identifying a target journal are as follows:

- Acceptance rate
- Number of journal issues per year
- Number of papers published per year
- Provision of online submission
- Facility to track the manuscript status

10/20/2024 39

Time Series Econometrics: Some Basic Concepts, Sources, Processing and Transformation

Prof. Dr. Shah Johir Rayhan
Department of Agricultural Finance and Management
Faculty of Agribusiness Management
Sher-e-Bangla Agricultural University, Dhaka-1207

❖ Lecture 1: Some Basic Concepts of Time Series Econometrics

What is a time series?

A set of data depending on the time.

A **time series** is a set of observations measured sequentially throughout time.

These measurements may be made **continuously** or taken at discrete time points.

Collection of magnitudes belonging to different periods of some variable or composite of variables such as the production of rice, per capita income, gross national income, price of tobacco, and index of the industrial output.

One of the important and frequent types of data used in empirical analysis. However, it poses several challenges to econometricians/practitioners. e.g.

1) Empirical work based on time series data assumes that the underlying time series is stationary.
2) Autocorrelation: because the underlying time series data is non-stationary.
3) Spurious/nonsense regression: a very high R2 and significant regression coefficients (though there is no meaningful relationship between the two variables)

Key Concepts

1) Stochastic Processes

i) Stationarity Processes

ii) Purely Random Processes

iii) Non-stationary Processes

2) Random Walk Models
i) Random Walk with Drift

ii) Random Walk without Drift

3. Unit Root Stochastic Processes

4. Deterministic and Stochastic Trends

5. The Phenomenon of Spurious Regression

6. Tests of Stationarity/non-stationarity

i) Graphical Method

ii) Unit Root Tests

(1) Stochastic Processes

Stochastic (Random) Process: collection of random variables ordered in time.

- NOTATIONS: Let Y a random variable, Y(t) if continuous (e.g. electrocardiogram), and Y_t if discrete (e.g. GDP, PDI, etc.).
- Now, If we let Y represent GDP, then we can have Y_1, Y_2, Y_3, ... , Y_{20} where the subscript 1 denotes the 1st observation (i.e. GDP for the 1st quarter of 1st year) and the subscript 20 denotes the last observation (i.e. GDP for the 4th quarter of 5th year).

Stationary Stochastic Processes: A stochastic process is said to be stationary/ weakly /covariance/2nd-order stationary if:

- Its mean and variance are constant over time, and
- The value of the covariance between the two time periods depends only on the distance/lag between the two time periods and not the actual time at which the covariance is computed.
- e.g. let's Y_t be a stochastic process, then;

Mean: $E(Y_t) = \mu$ ………………………………………… (1)

Variance: $var(Y_t) = E(Y_t - \mu)^2 = \sigma 2$ ……………………..(2)

Covariance: $\gamma k = E[(Yt - \mu)(Yt + k - \mu)]$ ……………..… (3)

Where γk, the covariance (or auto-covariance) at lag k, If k = 0, we obtain γ_0, which is simply the variance of Y (= σ2); if k = 1, γ1 is the covariance between two adjacent values of Y

Why are Stationary Time Series So Important?

Because if a time series is non-stationary, we can study its behaviour only for the period under consideration, and as a consequence, it is not possible to generalize it to other time periods.

- Therefore, for the purpose of forecasting, such (non-stationary) time series may be of little practical value.
- Non-stationary Stochastic Processes: Although our interest is in stationary time series, one often encounters non-stationary time series.
- A non-stationary time series will have a time-varying mean or a time-varying variance or both.

White Noise Processes

- We call a stochastic process (time series) a purely random/white noise process if it has zero mean, constant variance σ^2, and is serially uncorrelated i.e. [$u_t \sim IIDN(0, \sigma^2)$].
- Note: Here onward, in all equations, the assumption of "white noise" will be applicable on u_t

2) Random Walk Model (RWM)

The classic example of a non-stationary time series is the Random Walk Model (RWM).

It is often said that asset prices, such as stock prices or exchange rates, follow a random walk (i.e. non-stationary).

Types of Random Walks:

a) Random Walk Without Drift: i.e. no constant/intercept term and

$$\hat{GDP} = bGDPt\text{-}1 + ut$$

b) Random Walk with Drift i.e. a constant term is present

$$GDP = a + bGDPt\text{-}1 + ut$$

a) Random Walk without Drift

The time series Y_t is said to be a random walk without drift, if

$Y_t = Y_{t-1} + u_t$ ……. (4)

Here, the value of Y at time (t) is equal to its value at time (t − 1) plus a random shock; thus, it is an AR (1) model.

Believers in the Efficient Capital Market Hypothesis argue that stock prices are essentially random and therefore there is no scope for profitable speculation in the stock market:

If one could predict tomorrow's price based on today's price, we would all be millionaires.

Now from $Y_t = Y_{t-1} + ut$……. (4) we can write:

$Y1 = Y0 + u1$

=> $Y2 = Y1 + u2 = Y0 + u1 + u2$

=> $Y3 = Y2 + u3 = Y0 + u1 + u2 + u3$ and so on…

In general, if the process started at some time 0 with a value of Y_0, we have:

$Yt = Y0 + \sum ut$ ……... ………………………. (5)

Therefore, $E(Yt) = E(Y0 + \sum ut) = Y0$ (why?) ………… (6)

Because ut is "white noise"

In like fashion, it can be shown that: $var(Yt) = t\sigma 2$ …… (7)

Thus, the mean of Y is equal to its starting value, which is constant, but as t increases, its variance increases indefinitely (thus violating the condition of stationarity).

In short, the RWM without drift is a non-stationary stochastic process.

Now, if you write $Yt = Yt - 1 + ut$…... (4) as

$(Yt - Yt{-}1) = \Delta Yt = ut$ ……. (8)

It shows that, while Yt is non-stationary, its 1st difference is stationary.

In other words, the 1st differences of a random walk time series are stationary.

b) Random Walk with Drift

Let's modify, $Y_t = Y_{t-1} + u_t$…………….(4) as follows:

$Y_t = \delta + Y_{t-1} + u_t$…......………….(9)

where δ is the drift parameter.

The name drift comes from the fact that if we write the preceding equation as:

$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t$ …………. (10)

It shows that Y_t drifts upward/downward, depending on δ being positive/negative.

Note that model $Y_t = \delta + Y_{t-1} + u_t$ …………..(9) is also an AR (1) model.

Following the procedure discussed for Random Walk Without Drift, it can be shown that for the random walk with drift model (9),

$E(Y_t) = Y_0 + t \cdot \delta$ ……….. (11)

$\text{var}(Y_t) = t\sigma^2$ …………….... (12)

Here, again for RWM with drift, the mean as well as the variance increases over time, again violating the conditions of stationarity. In short, RWM, with or without drift, is a non-stationary stochastic process.

The random walk model is an example of what is known in the literature as a Unit Root Process.

3) Unit Root Stochastic Process

Let's write the RWM $Y_t = Y_{t-1} + u_t$….. (4) as:

$Y_t = \rho Y_{t-1} + u_t \qquad -1 \leq \rho \leq 1$ ……. (13)

If ρ = 1, (13) becomes an RWM (without drift).

If ρ is in fact 1, we face what is known as the unit root problem (non-stationarity); as the variance of Y_t is not stationary.

The name unit root is due to the fact that ρ = 1.

Thus, the terms non-stationarity, random walk, and unit root can be treated as synonymous.

If, however, $|\rho| \leq 1$, then the time series Y_t is stationary in the sense we have defined it.

Note: Unit Root Stochastic Process will be further explained in the Unit Root Test of Stationarity.

4) Trend Stationary (TS) and Difference Stationary (DS) Stochastic Processes

Deterministic Trend: if the trend in a time series is completely predictable

Stochastic Trend: if it is not predictable

5) Spurious Regression

Stationary Time Series are important, consider the following two random walk models:

$Yt = Yt-1 + ut$ …………... (20)

$Xt = Xt-1 + vt$ …………. (21)

Where we generated 500 observations of ut from ut ∼ N(0, 1) and 500 observations of vt from vt ∼ N(0, 1) and assumed that the initial values of both Y and X were zero. We also assumed that ut and vt are serially uncorrelated as well as mutually uncorrelated. Both these time series are non-stationary; i.e. they are I(1) or exhibit stochastic trends.

- ❖ If the link between the variables is not stationary, we are in the presence of a spurious regression. (i.e, residuals are non-stationary).
- ❖ Spurious regressions are those between variables with a similar trend but don't have an economic sense. Can have statistical significance, but no economic interpretation.
- ❖ According to Newbold and Granger (1974), Spurious regressions signs include:
  - High R2 and low Durbin Watson statistic. (Rule of thumb: R2 > dw)
  - T-Statistics are very high: Variables are highly significant.
  - Residuals are not stationary
- ❖ Non-stationary variables can have a stationary linear combination. Long-run equilibrium.

Sources of Time series data

1: https://data.worldbank.org/country/bangladesh?view=chart

2. https://databank.worldbank.org/source/world-development-indicators

3. https://bbs.gov.bd/

4. https://openknowledge.worldbank.org/home

5. https://www.fao.org/geospatial/resources/data-portals/en/

6. https://dam.portal.gov.bd/

7. https://www.ifpri.org/publications/datasets/

Lecture 2: Practice with EViews software: Detecting Stationarity of Time Series Data

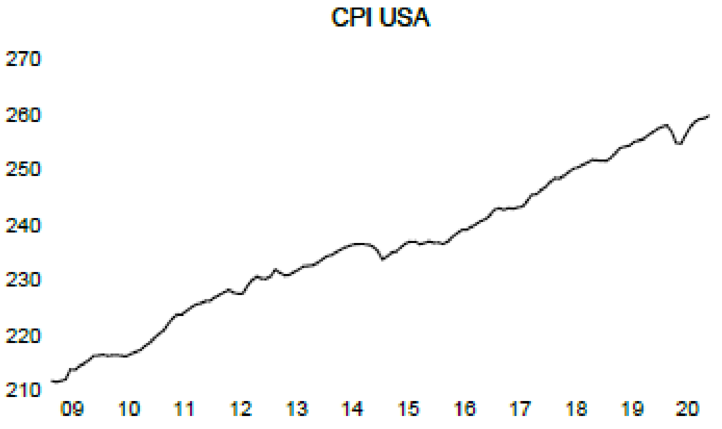
Stationarity

The model estimation method is based on stationary time series. A series is covariance stationary if the mean and covariance of the series do not depend on time. A stationary series will have no trend, its variations around its mean have a constant amplitude, and it wiggles consistently. Since we are trying to predict future values of our variable of interest, we should first ensure that our variable is stationary.

The most common approach to test for stationarity is checking the graph and the correlogram, and finally complementing the analysis with formal unit root tests. Mainly, Augmented-Dickey Fuller and Phillips Perron test.

Graph Analysis

Figure 1: Consumer Price Index USA



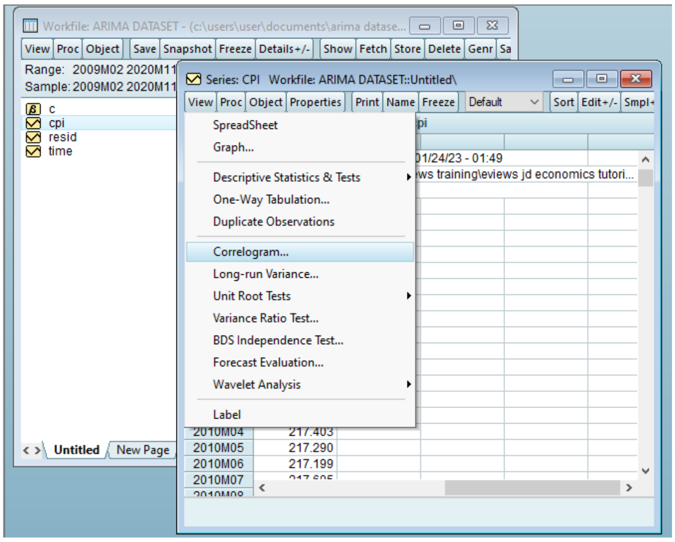
Notes: Consumer Price Index for USA. Source: FRED.

Frequency: Monthly. Time Range: 2009M02-2020M11.

Graph 1 clearly shows an upward trend which is an indicator of non-stationarity. We will continue the analysis with the correlogram.

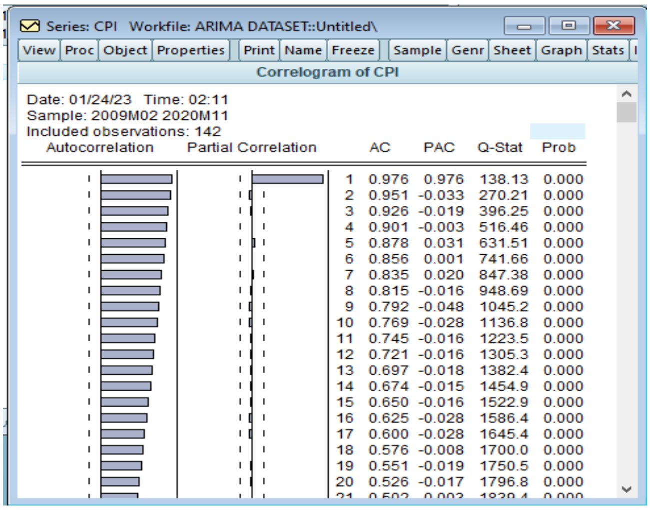
2.2 Correlogram

The next step is displaying the correlogram. A slow decay in the autocorrelation function will indicate that the variable "CPI" is non-stationary.



To display the correlogram, double-click on our variable of interest and select → View → Correlogram → option "Levels".

Figure 2: Correlogram of CPI



The autocorrelation column of our variable CPI in levels suggests that our variable is non-stationary as the bars in the first column do not decay in a fast pattern. We will finish our analysis with the formal unit root tests.

2.3 Unit Root Tests

The last step of the stationarity analysis is to conduct some formal tests. Both the Augmented Dickey-Fuller Test and Phillips-Perron test will help us confirm if our variable of interest is non-stationary.

To conduct the augmented Dickey-Fuller test double-click on our variable of interest and select: ⟶ View ⟶ Unit Root Tests ⟶ Standard Unit Root Tests.

Figure 3: Augmented Dickey-Fuller Test

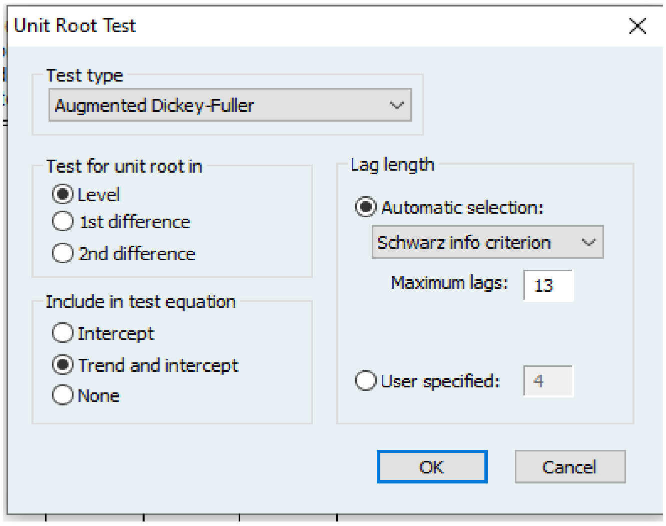


Figure 3 shows the diverse unit root test options we can select. To conduct the Augmented Dickey-Fuller test, we choose it from the "Test Type" dropdown. The next step is to select if we are testing for the existence of a unit root in our variable in levels, first differences or second differences. We will begin by checking the box "In levels". If our variable is not stationary in levels, we will proceed to check in first differences. Next, we should select if we want to include an intercept, trend or intercept, or none. In our case (looking at Figure 1) we select "Trend and Intercept". Finally, we need to specify the number of lags to include in the test. You can set it manually by using the "user-specified" option, or allow for the "Automatic Selection" criteria. The dropdown displays the diverse lag selection criteria (i.e., Schwarz, Akaike, etc.). For our case, we leave the default option: "Schwarz".

Figure 4 displays the test results. The null hypothesis is: "CPI has a unit root". Because the "p-value" is 0.4184 which is bigger than 0.05, we cannot reject the null hypothesis (box 1 in the graph). Consequently, our variable is non-stationary in levels. Box number 2 in the graph reflects the significance of the trend and the intercept. Since the "p-value" is smaller than 0.05, we can conclude that both the trend and the intercept belong to the series. Incorporating the trend and the intercept in the test specification is appropriate.

Lag selection criteria: Schwarz.

The last step is to complement the Augmented Dickey-Fuller Test with the Phillips-Perron test. A significant advantage of the P-P test is that it is non-parametric (i.e., it does not require selecting the lag length as in ADF). As an observation, the P-P test works better with large datasets.

To conduct the Phillips-Perron test double-click on our variable of interest and select: → View → Unit Root Tests → Standard Unit Root Tests → Phillips-Perron Test.

Figure 4: Augmented Dickey-Fuller Test Results

Null Hypothesis: CPI has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic - based on SIC, maxlag=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.323145	0.4184
Test critical values:	1% level	-4.025426	
	5% level	-3.442474	
	10% level	-3.145882	

*MacKinnon (1996) one-sided p-values.

1

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CPI)
Method: Least Squares
Date: 01/24/23 Time: 02:41
Sample (adjusted): 2009M05 2020M11
Included observations: 139 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CPI(-1)	-0.052419	0.022564	-2.323145	0.0217
D(CPI(-1))	0.472051	0.083302	5.666749	0.0000
D(CPI(-2))	-0.144664	0.085014	-1.701660	0.0911
C	11.47553	4.825418	2.378143	0.0188
@TREND("2009M02")	0.016566	0.007301	2.269087	0.0249

R-squared	0.215825	Mean dependent var	0.346101
Adjusted R-squared	0.192417	S.D. dependent var	0.518864
S.E. of regression	0.466280	Akaike info criterion	1.347249
Sum squared resid	29.13393	Schwarz criterion	1.452806
Log likelihood	-88.63380	Hannan-Quinn criter.	1.390144
F-statistic	9.220043	Durbin-Watson stat	2.031186
Prob(F-statistic)	0.000001		

Notes: Augmented Dickey-Fuller Test with trend and intercept.

Similarly to the ADF test, the null hypothesis "CPI has a unit root" cannot be rejected. The "p-value" is 0.5382, which is bigger than 0.05. Therefore, we can confirm that our series is non-stationary. At the 10% level, both the trend and the intercept correspond to the series.

In conclusion, our series is non-stationary and needs to be differentiated to remove the trend. We will conduct the "Augmented Dickey-Fuller" test again, but this time selecting "First Differences" in the test options (See Figure 3.).

Figure 5: Phillips-Perron Test Results

Null Hypothesis: CPI has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 4 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-2.105046	0.5382
Test critical values:	1% level	-4.024452	
	5% level	-3.442006	
	10% level	-3.145608	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.259901
HAC corrected variance (Bartlett kernel)	0.384654

Phillips-Perron Test Equation
Dependent Variable: D(CPI)
Method: Least Squares
Date: 01/24/23 Time: 03:01
Sample (adjusted): 2009M03 2020M11
Included observations: 141 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CPI(-1)	-0.042107	0.023925	-1.759927	0.0806
C	9.366209	5.121771	1.828705	0.0696
@TREND("2009M02")	0.013369	0.007769	1.720878	0.0875

R-squared	0.022136	Mean dependent var	0.341220
Adjusted R-squared	0.007964	S.D. dependent var	0.517381
S.E. of regression	0.515317	Akaike info criterion	1.532977
Sum squared resid	36.64608	Schwarz criterion	1.595716
Log likelihood	-105.0749	Hannan-Quinn criter.	1.558472
F-statistic	1.561985	Durbin-Watson stat	1.186639
Prob(F-statistic)	0.213404		

Notes: Phillips Perron test with trend and intercept.

Figure 6 shows the Augmented-Dickey Fuller test results in the first differences. In the test specifications, we include trend and intercept. As we can see in box 1, the "p-value" is smaller than 0.05, so we can reject the null hypothesis (H0: CPI has a unit root). Consequently, "CPI" is stationary in the first differences.

Observing box 2, we can confirm that the trend is gone as the trend is not significant (p-value is bigger than 0.05). Finally, including a constant "c" in the model is appropriate (p-value smaller than 0.05).

We can confirm that our series is non-stationary in levels, but we achieve stationarity by applying the first differences. The order of the "d" component in our ARIMA (p, d, q) is 1. Now is the time to determine the values of "p" and "q".

Figure 6: Augmented Dickey-Fuller Test - First Differences

Null Hypothesis: D(CPI) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.764199	0.0000
Test critical values:	1% level	-4.025426	
	5% level	-3.442474	
	10% level	-3.145882	

*MacKinnon (1996) one-sided p-values.

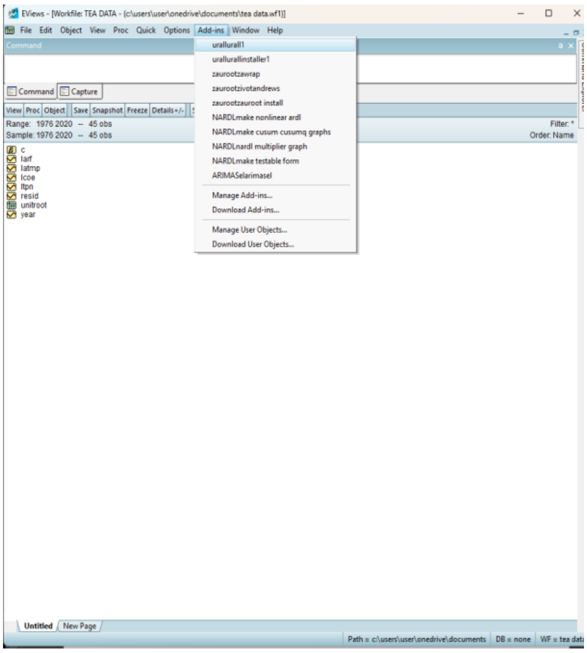
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CPI,2)
Method: Least Squares
Date: 01/24/23 Time: 03:11
Sample (adjusted): 2009M05 2020M11
Included observations: 139 after adjustments

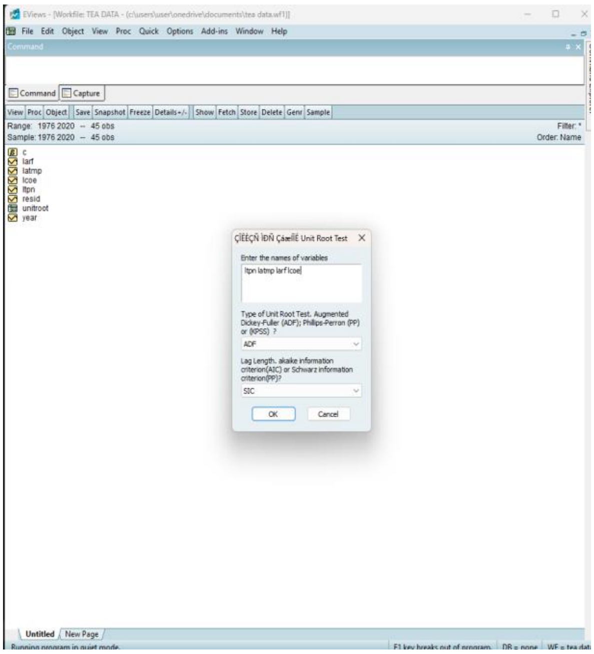
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(CPI(-1))	-0.722052	0.092998	-7.764199	0.0000
D(CPI(-1),2)	0.187989	0.084283	2.230448	0.0274
C	0.267243	0.089208	2.995731	0.0033
@TREND("2009M02")	-0.000239	0.001002	-0.238743	0.8117

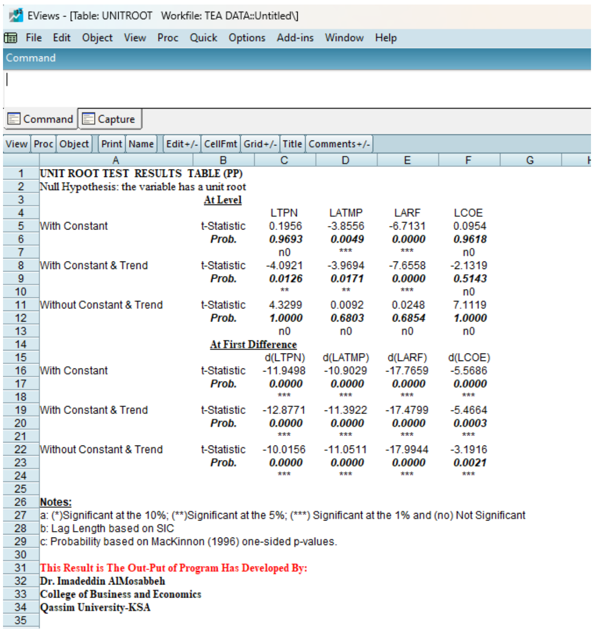
R-squared	0.328865	Mean dependent var	0.002000
Adjusted R-squared	0.313951	S.D. dependent var	0.572044
S.E. of regression	0.473813	Akaike info criterion	1.372347
Sum squared resid	30.30733	Schwarz criterion	1.456792
Log likelihood	-91.37809	Hannan-Quinn criter.	1.406663
F-statistic	22.05057	Durbin-Watson stat	2.054707
Prob(F-statistic)	0.000000		

Notes: Augment Dickey-Fuller test with trend and intercept.

Variable in first differences.







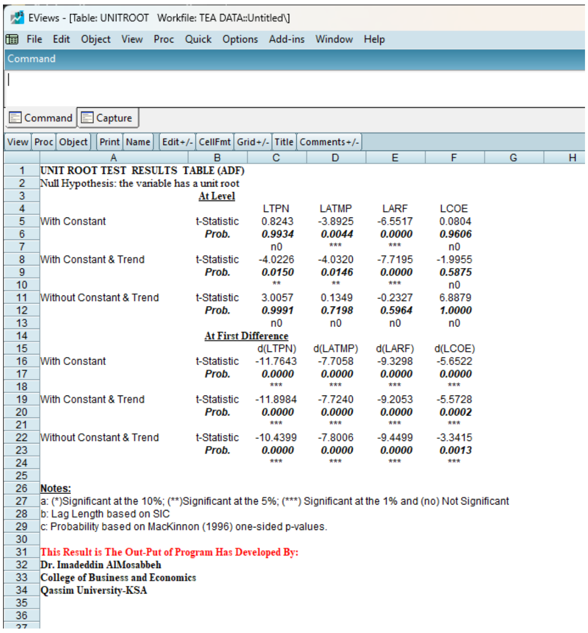
UNIT ROOT TEST RESULTS TABLE (PP)

Null Hypothesis: the variable has a unit root

		LTPN	LATMP	LARF	LCOE
At Level					
With Constant	t-Statistic	0.1956	-3.8556	-6.7131	0.0954
	Prob.	*0.9693*	*0.0049*	*0.0000*	*0.9618*
		n0	***	***	n0
With Constant & Trend	t-Statistic	-4.0921	-3.9694	-7.6558	-2.1319
	Prob.	*0.0126*	*0.0171*	*0.0000*	*0.5143*
		**	**	***	n0
Without Constant & Trend	t-Statistic	4.3299	0.0092	0.0248	7.1119
	Prob.	*1.0000*	*0.6803*	*0.6854*	*1.0000*
		n0	n0	n0	n0
At First Difference					
		d(LTPN)	d(LATMP)	d(LARF)	d(LCOE)
With Constant	t-Statistic	-11.9498	-10.9029	-17.7659	-5.5686
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0000*
		***	***	***	***
With Constant & Trend	t-Statistic	-12.8771	-11.3922	-17.4799	-5.4664
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0003*
		***	***	***	***
Without Constant & Trend	t-Statistic	-10.0156	-11.0511	-17.9944	-3.1916
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0021*
		***	***	***	***

Notes:
a: (*)Significant at the 10%; (**)Significant at the 5%; (***) Significant at the 1% and (no) Not Significant
b: Lag Length based on SIC
c: Probability based on MacKinnon (1996) one-sided p-values.

This Result is The Out-Put of Program Has Developed By:
Dr. Imadeddin AlMosabbeh
College of Business and Economics
Qassim University-KSA



UNIT ROOT TEST RESULTS TABLE (ADF)

Null Hypothesis: the variable has a unit root

		LTPN	LATMP	LARF	LCOE
At Level					
With Constant	t-Statistic	0.8243	-3.8925	-6.5517	0.0804
	Prob.	*0.9934*	*0.0044*	*0.0000*	*0.9606*
		n0	***	***	n0
With Constant & Trend	t-Statistic	-4.0226	-4.0320	-7.7195	-1.9955
	Prob.	*0.0150*	*0.0146*	*0.0000*	*0.5875*
		**	**	***	n0
Without Constant & Trend	t-Statistic	3.0057	0.1349	-0.2327	6.8879
	Prob.	*0.9991*	*0.7198*	*0.5964*	*1.0000*
		n0	n0	n0	n0
At First Difference					
		d(LTPN)	d(LATMP)	d(LARF)	d(LCOE)
With Constant	t-Statistic	-11.7643	-7.7058	-9.3298	-5.6522
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0000*
		***	***	***	***
With Constant & Trend	t-Statistic	-11.8984	-7.7240	-9.2053	-5.5728
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0002*
		***	***	***	***
Without Constant & Trend	t-Statistic	-10.4399	-7.8006	-9.4499	-3.3415
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0013*
		***	***	***	***

Notes:
a: (*)Significant at the 10%; (**)Significant at the 5%; (***) Significant at the 1% and (no) Not Significant
b: Lag Length based on SIC
c: Probability based on MacKinnon (1996) one-sided p-values.

This Result is The Out-Put of Program Has Developed By:
Dr. Imadeddin AlMosabbeh
College of Business and Economics
Qassim University-KSA

❖ Lecture 3: Practice with EViews software: Spatial price transmission in the agricultural commodity markets of Bangladesh with NARDL approach

Non-linear ARDL (NARDL) Model

The Nonlinear Autoregressive Distributed Lag (NARDL) model is a method for modelling both short- and long-run asymmetries. It was developed by Shin, Yu, and Greenwood-Nimmo in 2014. The NARDL model can be applied to stationary and non-stationary time series vectors or combinations of both. The ARDL model is used to examine the symmetric (linear) relationship. This model is extended to include the asymmetric variables and is termed as nonlinear ARDL or NARDL model.

Consider two variables: GDP (Y) and Domestic credit to the private sector by banks (X).

The question is, is the magnitude of GDP change the same in both cases (in both directions)?

Symmetric relationship means the degree of impact of Domestic credit (X) on GDP (Y) is the same when Domestic credit increases (X^+) as when Domestic Credit decreases (X^-).

It might be that an increase in Domestic credit has a stronger impact on GDP than a decrease in Domestic Credit, or perhaps vice versa.

If we find that the magnitude of impact is not the same on both sides of the changes, then we conclude Domestic Credit has asymmetric impact on GDP.

Consider the following (long-run) OLS time series model:

$Y_t = \beta_0 + \beta_1 X_t + U_t$

where Y Target variable

X = Regressor B, Change in Y per unit change in X (captures the direction and magnitude Y's reaction to changes in X)

- Suppose the relationship between X and Y is positive so that ß, > 0 (e.g. $B_1 = 2$)
- This means that if X increases, Y increases 2x as much; and if X decreases, Y decreases 2x as much
- But this assumption of symmetric impact may not hold up in reality

About NARDL

NARDL separates the reactions of Y to negative and positive changes in X

$Y_t = \beta_0 + \beta_1 X_t + U_t$

To capture the effects of asymmetry, NARDL decomposes X into two parts:

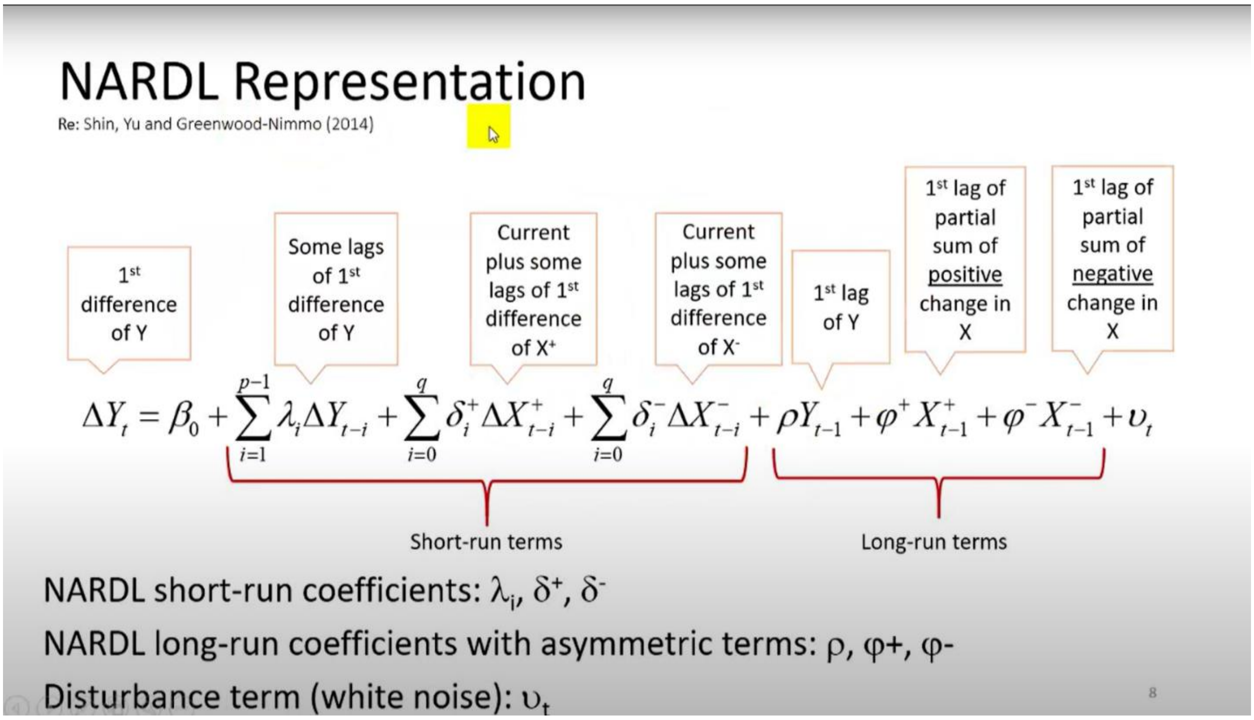
(1) Partial sum of positive change in X, denoted by X+

(2) Partial sum of negative change in X, denoted by X^{--}

Both X+ and X^{--} are included as separate regressors in the NARDL model

Asymmetric long-run regression model:

$Y_t = \beta_0 + \beta_1 X_t^{+} + \beta_1 X_t^{-} + U_t$



Intuitive view of the data structure

$$\boxed{Y_t = \beta_0 + \beta_1 X_t^+ + \beta_2 X_t^- + \upsilon_t}$$

Y	X	X^+	X^-
35	5	5	5
32	1	0	-4
47	8	7	0
54	9	1	0
10	6	0	-3
15	3	0	-3
20	7	4	0
slope	**2.67**	**1.51**	**1.23**

In second row, data in X is 1 which means that X value has decrease from 5 to 1 i.e -4
In third row, data in X is 7 which means that X value has Increase from 1 to 8 i.e 7

Here, the positive slope is 1.51 and negative slope is 1.23
So we want to test is there any significant difference
In positive and negative slope.

$$\boxed{\Delta Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_{1i} \Delta Y_{t-i} + \sum_{i=0}^{q} \alpha_{2i} \Delta X_{t-i}^{+} + \sum_{i=0}^{q} \alpha_{3i} \Delta X_{t-i}^{-} + \rho Y_{t-1} + \varphi^{+} X_{t-1}^{+} + \varphi^{-} X_{t-1}^{-} + \mu_t}$$

Bounds Test for Asymmetric Long-Run Cointegration

- Similar to ARDL bounds test, NARDL bounds test is also a joint test of all lagged one-period levels of x^+, x^-, and y.
- F-test of Pesaran *et al* (2001) or Narayan (2004), if using small n

$$H_0: \rho = \varphi^+ = \varphi^- = 0$$

- t-test of Banerjee et al (1998)

$$H_0: \varphi = 0$$
$$H_A: \varphi < 0$$

How should we conclude?
If we reject H_0 (of no cointegration), we conclude the variables are cointegrated in the presence of asymmetry.

$$\boxed{\Delta Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_{1i} \Delta Y_{t-i} + \sum_{i=0}^{q} \alpha_{2i} \Delta X^{+}_{t-i} + \sum_{i=0}^{q} \alpha_{3i} \Delta X^{-}_{t-i} + \rho Y_{t-1} + \varphi^{+} X^{+}_{t-1} + \varphi^{-} X^{-}_{t-1} + \mu_t}$$

NARDL Long-run Asymmetric Coefficients

We calculate the NARDL long-run levels asymmetric coefficients by:

- dividing the negative of the coefficient of X^{+}_{t} (i.e. φ^{+}) by the coefficient of Y_{t-1} (i.e. ρ):

$$\frac{-\varphi^{+}}{\rho}$$

- and also, by dividing the negative of the coefficient of X^{-}_{t} (i.e. φ^{-}) by the coefficient of Y_{t-1} (i.e. ρ):

$$\frac{-\varphi^{-}}{\rho}$$

11

Wald Test for Long-Run Asymmetry

- If a long-run relationship exists (Bounds Test), we proceed to test if the difference in the asymmetric coefficients is statistically significant:

$$H_0 : \frac{-\varphi^{+}}{\rho} = \frac{-\varphi^{-}}{\rho} \qquad H_A : \frac{-\varphi^{+}}{\rho} \neq \frac{-\varphi^{-}}{\rho}$$

- If we reject H_0, it means we have *long-run asymmetry*. In other words, the magnitude of the change in Y when X increases is NOT THE SAME as when X decreases.

Assumption of NARDL Model

Check all the variables are stationary at level or first difference but neither of the variable should be I(2) i.e. stationary at 2^{nd} difference.

Variables in the study:

Y_t = Occupancy rate (LOR)
X_{1t} = Real GDP growth rate (LGDP)
X_{2t} = Average daily rate (LADR)

Linear functional form: $Y = f(x_{1t}, x_{2t})$

Both regressors are decomposed into their positive and negative shocks so that

Nonlinear functional form: $Y = f(x_{1t}^{+}, x_{1t}^{-}, x_{2t}^{+}, x_{2t}^{-})$

On EViews: LOR = f(LGDP_POS, LGDP_NEG, LADR_POS, LADR_NEG)

Consider the long-run regression model: $y_t = \beta_0 + \beta_1 x_t + \upsilon_t$

The following asymmetric long-run regression describes the empirical model:

$$y_t = \beta_0 + \beta_1 x_t^{+} + \beta_2 x_t^{-} + \upsilon_t$$

where x_t^{+} and x_t^{-} are the partial sums of positive (+) and negative (-) changes in x_t :

$$x_t^{+} = \sum_{j=1}^{t} \Delta x_j^{+} = \sum_{j=1}^{t} \max(\Delta x_j, 0) \qquad x_t^{-} = \sum_{j=1}^{t} \Delta x_j^{-} = \sum_{j=1}^{t} \min(\Delta x_j, 0)$$

18

- NARDL cointegration approach assumes the response of the dependent variable, Y_t, to increases (+) and decreases (-) of each independent variable (x_{it}) is asymmetric (Re: Shin, Yu, and Greenwood-Nimmo, 2014)
- Accordingly, the nonlinear model for this study takes the following form:

$$\Delta y_t = \beta_0 + \sum_{i=1}^{p-1} \lambda_i \Delta y_{t-i} + \sum_{i=0}^{q} \delta_i^+ \Delta x_{1t-i}^+ + \sum_{i=0}^{q} \delta_i^- \Delta x_{1t-i}^- + \sum_{i=0}^{q} \lambda_i^+ \Delta x_{2t-i}^+ + \sum_{i=0}^{q} \lambda_i^- \Delta x_{2t-i}^-$$
$$+\rho y_{t-1} + \varphi_1^+ x_{1t-1}^+ + \varphi_1^- x_{1t-1}^- + \varphi_2^+ x_{2t-1}^+ + \varphi_2^- x_{2t-1}^- + \upsilon_t$$

Long-run asymmetric effects of x_1 on y is calculated as $L_{M1+} = \frac{-\varphi_1^+}{\rho}$ and $L_{M1-} = \frac{-\varphi_1^-}{\rho}$

Short-run asymmetric effects of x_1 on y is represented by $\sum_{i=0}^{q} \delta_i^+$ and $\sum_{i=0}^{q} \delta_i^-$

Using Wald test, if the null hypotheses $\frac{-\varphi_1^+}{\rho} = \frac{-\varphi_1^-}{\rho}$ for long-run symmetry and $\sum_{i=0}^{q} \delta_i^+ = \sum_{i=0}^{q} \delta_i^-$ for short-run symmetry are rejected, we conclude the impact of x on y is asymmetric.

19

Download NARDL Add-inns on EViews 12

- Click ON Add-ins
- Download Add-ins
- Wait for the download to complete
- Scroll down to NARDL and INSTALL

Steps for NARDL

Specify NARDL model

1. Check for Unit Root Test
2. Select Optimum Lag Length
3. Specify the NARDL model with optimum Lag
4. Estimate Bound Test for cointegration
5. Estimate Long-Run Results
6. Estimate Short-run Results and Error Correction Term
7. Perform Some Diagnostic Tests
8. Checking Asymmetric relationship (Wald test)
9. Constructing NARDL Multiplayer Graph

Explanation of NARDL results (most important)

a. Normality Test

b. Serial Correlation Test

c. Heteroskedasticity test

d. Functionality Test (REMSE Reset test)

e. CUSUM and CUSUMSQRT test

The secondary data used in this study will be collected from the Department of Agricultural Marketing, Bangladesh. The analysis of time series data entails several procedures, one of which is to ensure that the research variables are stationary and that none of them are integrated beyond the second order. As a result, the unit root characteristics of the variables employed in the study must be scrutinized. When a variable has more than one order of integration, it produces erroneous results. To verify the order integration of variables, we used Augmented Dickey-Fuller (ADF) and Phillips–Perron (PP) unit root tests before employing the time series econometric model. Before testing the causality, it's imperative to consider the cointegrating properties of the variables. We used the Bounds cointegration test to see if there was any cointegration between variables.

- **Lecture 4: Practice with EViews software: Analyzing Short-Run and Long-Run Asymmetrical Effects of Climate Change on Agricultural Production in Bangladesh**

The newly formed and cutting-edge technique known as the NARDL model will be utilized to evaluate the asymmetrical effect of climate change on agricultural production in Bangladesh. Shin et al. (2014) modify the ARDL model to an asymmetric ARDL or NARDL to examine dynamic adjustment and asymmetries among variables in the short and long run.

The following model can be used to investigate variable relationships:

$$LTPN_t = \alpha_0 + \beta_1 LATMP_t + \beta_2 LARF_t + \beta_3 LCOE_t + \mu_t$$

where LTPN is agricultural production, LATMP donates Bangladesh's average annual temperature, LARF indicates Bangladesh's average yearly temperature, LAHU indicates Bangladesh's average annual humidity, and LCOE is Bangladesh's carbon dioxide emission. Logarithmic transformations were applied to all of the variables. In addition, *t* denotes the time period, and α and β are the parameters to be estimated, whereas μ is the error correction term.

The asymmetric cointegration equation of the above model is as follows:

$$LTPN_t = \beta_0 + \beta_1 LATMP_t^+ + \beta_2 LATMP_t^- + \beta_3 LARF_t^+ + \beta_4 LARF_t^- + \beta_5 LCOE_t^+ + \beta_6 LCOE_t^- + \mu_t$$

Where $LATMP_t^+$, $LATMP_t^-$, $LARF_t^+$, $LARF_t^-$, $LCOE_t^+$, $LCOE_t^-$ indicate the partial sum of positive and negative changes in LATMP, LARF, and LCOE at time *t*, respectively. Here, β ($\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$) are the associated asymmetric long-run parameters. All the variables have decomposed into two new variables that represent positive and negative shocks.

For calculating the asymmetric long-run and short-run relationship among study variables, the equation is as follows:

$$\Delta LTPN_t = \beta_0 + ELTPN_{t-1} + \beta_1^{+}LATMP_t^{+} + \beta_2^{-}LATMP_t^{-} + \beta_3^{+}LARF_t^{+} + \beta_4^{-}LARF_t^{-} + \beta_5^{+}LCOE_t^{+} + \beta_6^{-}LCOE_t^{-} + \sum_{i=1}^{o} \gamma o \Delta LTPNt - 1 + \sum_{i=1}^{p} (\lambda_{1i}{}^{+}\Delta LATMP^{+}_{t-1}\ \lambda_{2i}{}^{-}\Delta LATMP^{-}_{t-1}) + \sum_{i=1}^{q} (\psi_{1i}{}^{+}\Delta LARF^{+}_{t-1}\ \psi_{2i}{}^{-}\Delta LARF^{-}_{t-1}) + \sum_{i=1}^{s} (\eta_{1i}{}^{+}\Delta LACOE^{+}_{t-1}\ \eta_{2i}{}^{-}\Delta LACOE^{-}_{t-1}) + \varphi ECT_{-1} + \mu_t$$

In the equation presented above, (βi) represents long-run coefficients, whereas (λ_i), (ψ_i), and (η_i) represent short-run coefficients.

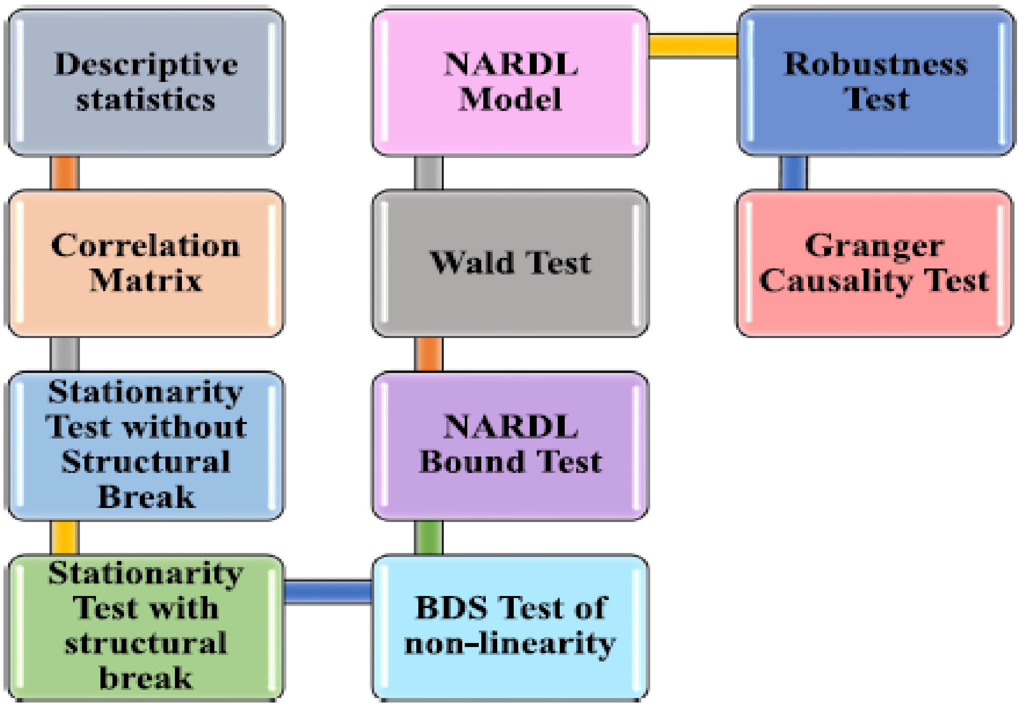
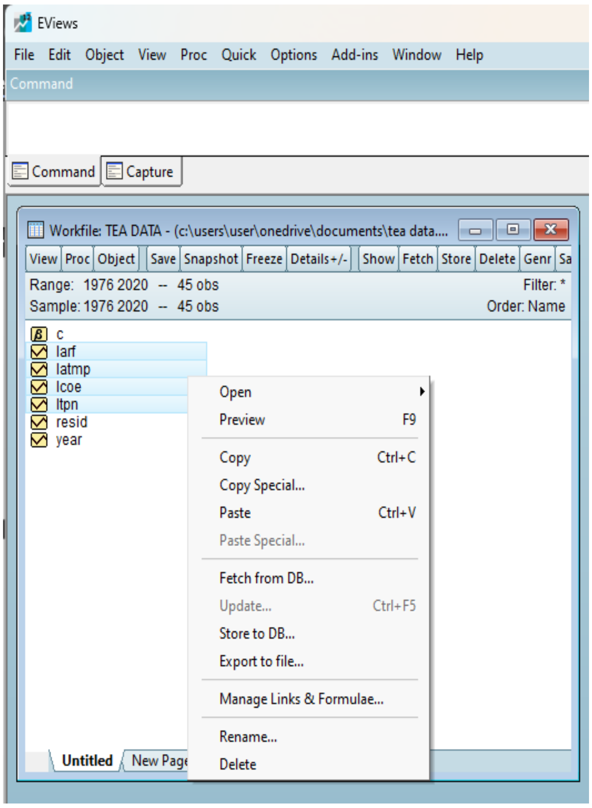
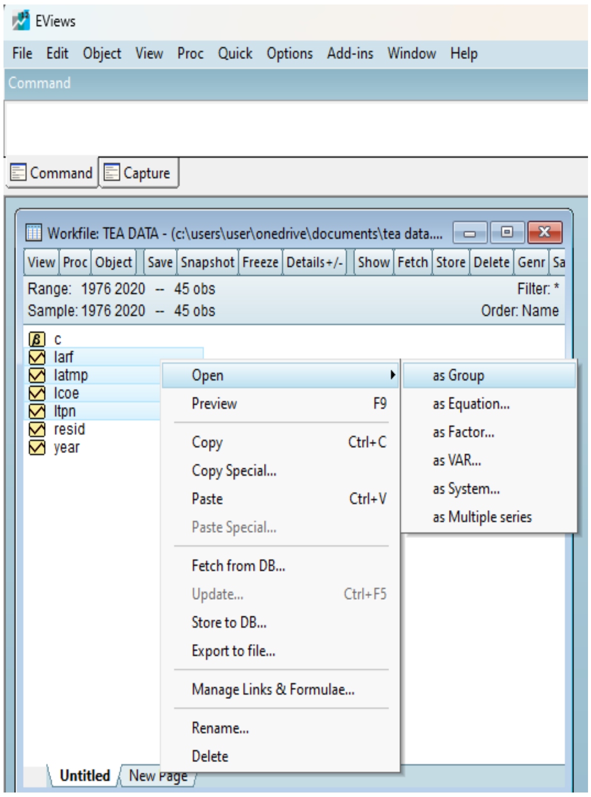
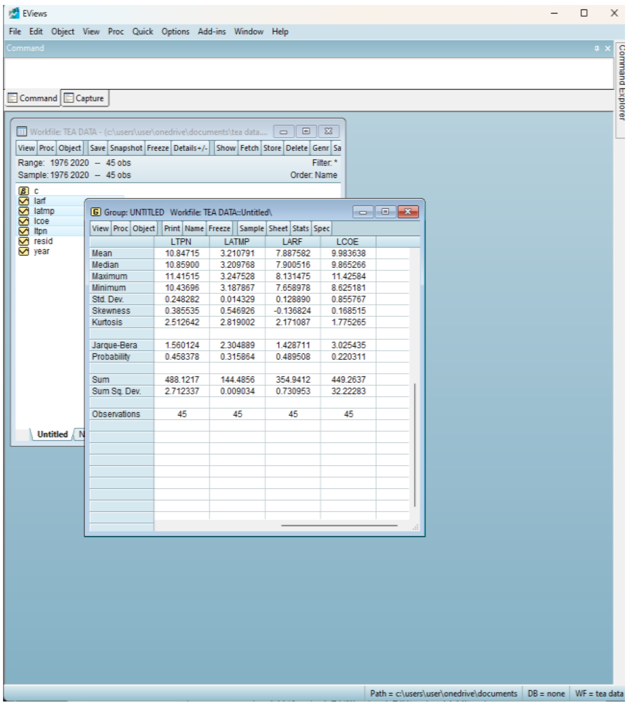


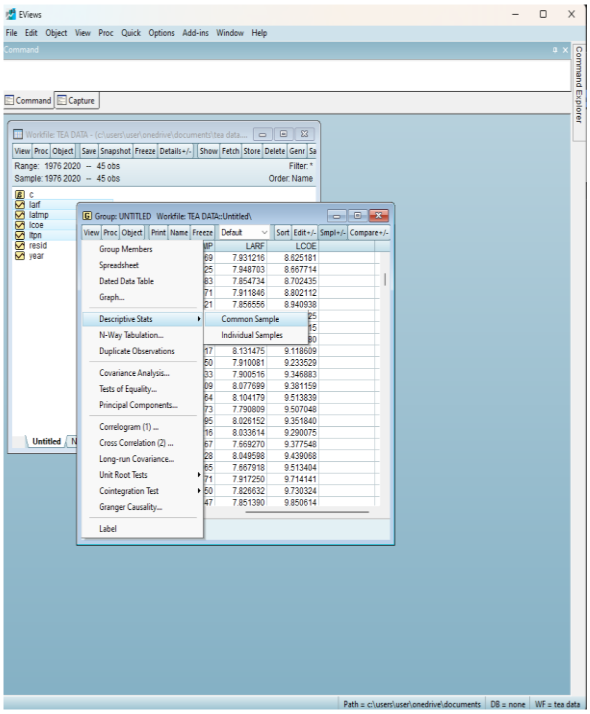
Fig. 2 Methodological approach of the Study

How to calculate descriptive statistics

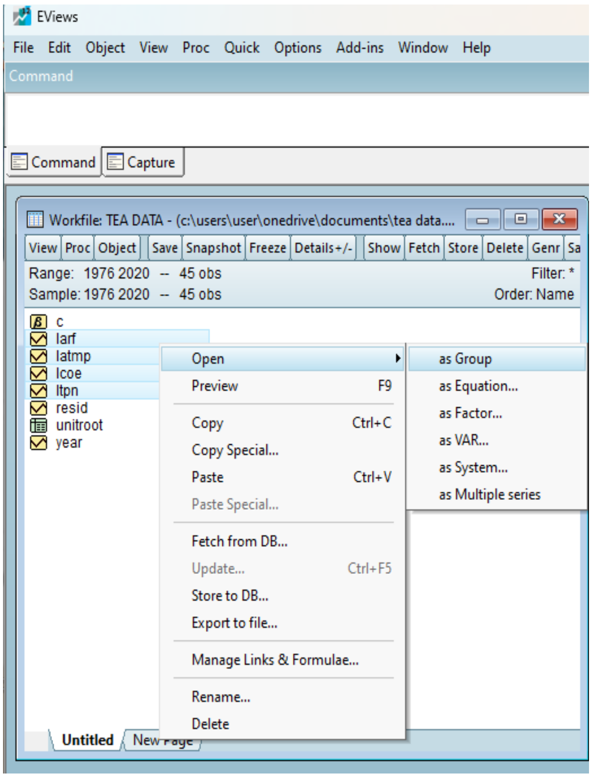


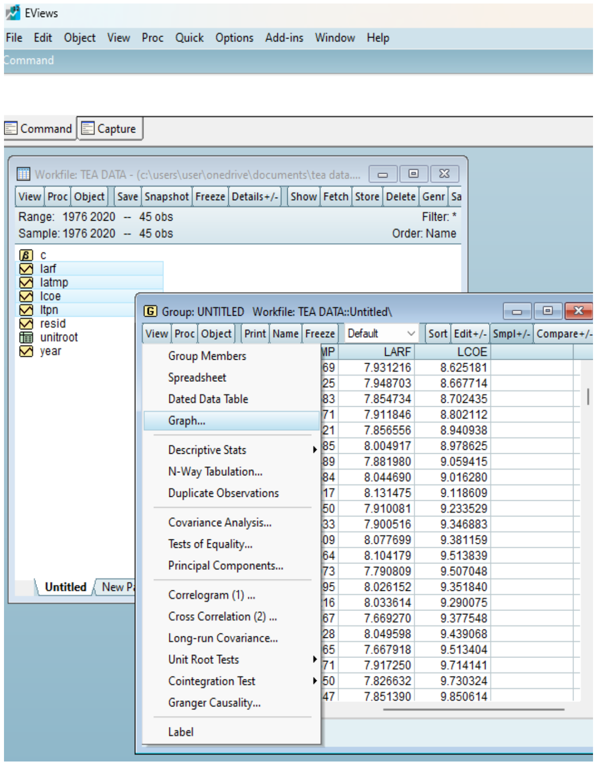


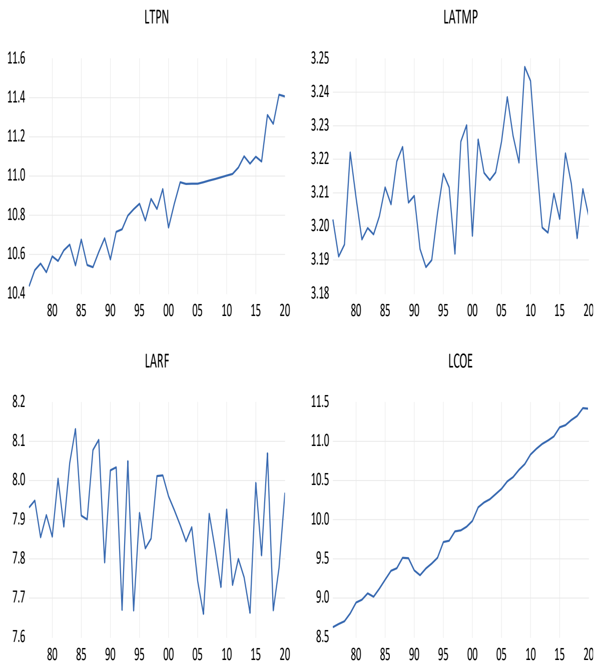


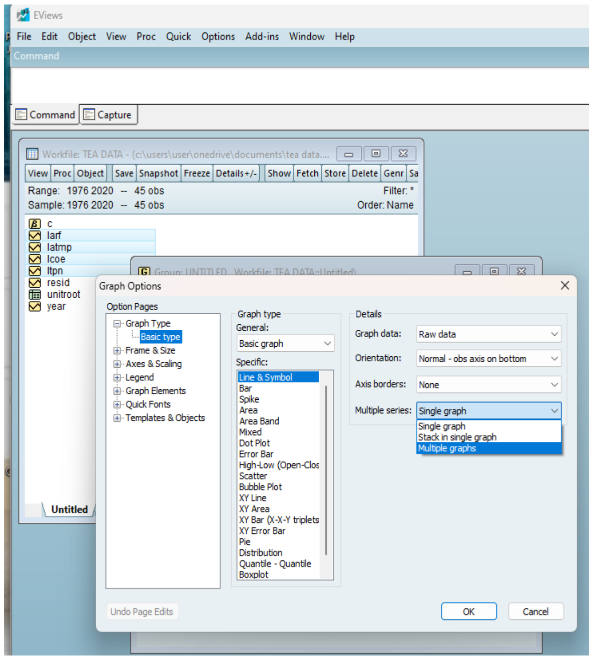


How to Make a Graph



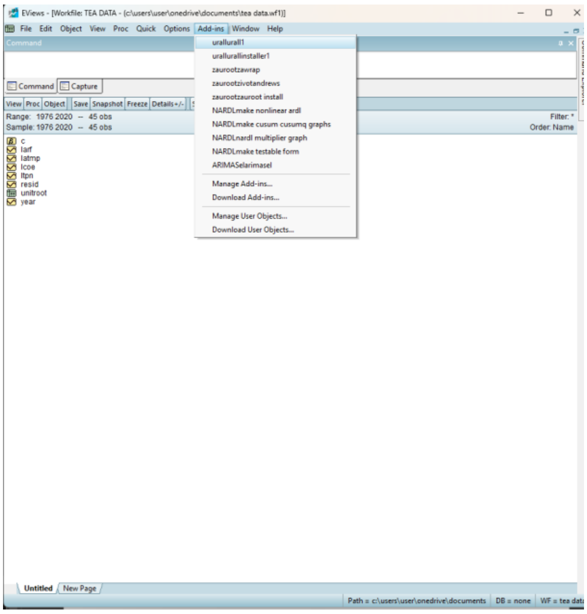


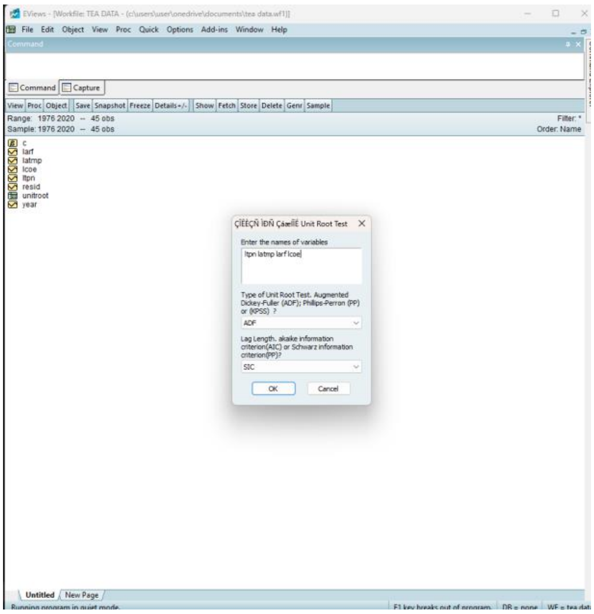


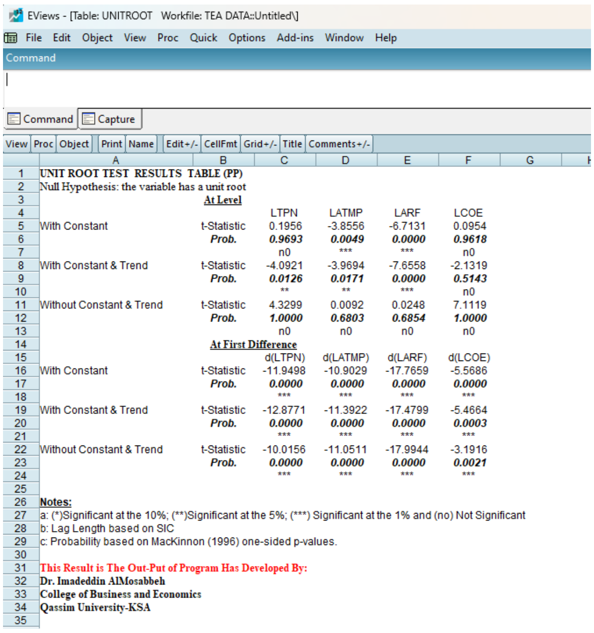


Correlation Matrix

Check for Unit Root Test



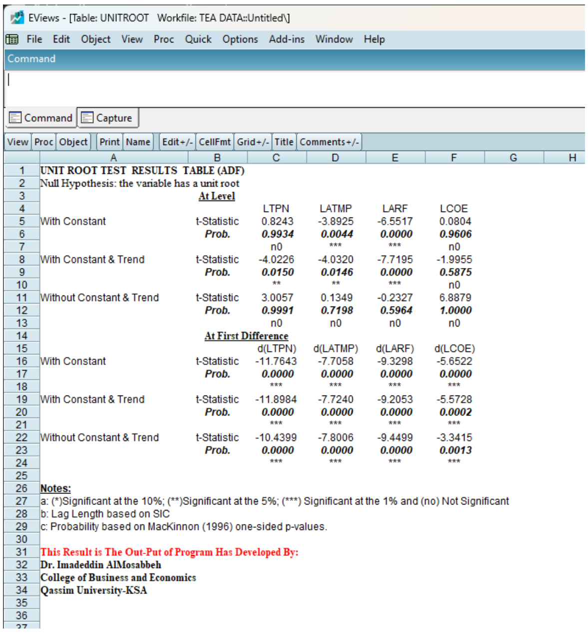




UNIT ROOT TEST RESULTS TABLE (PP)					
Null Hypothesis: the variable has a unit root					
	At Level				
		LTPN	LATMP	LARF	LCOE
With Constant	t-Statistic	0.1956	-3.8556	-6.7131	0.0954
	Prob.	*0.9693*	*0.0049*	*0.0000*	*0.9618*
		n0	***	***	n0
With Constant & Trend	t-Statistic	-4.0921	-3.9694	-7.6558	-2.1319
	Prob.	*0.0126*	*0.0171*	*0.0000*	*0.5143*
		**	**	***	n0
Without Constant & Trend	t-Statistic	4.3299	0.0092	0.0248	7.1119
	Prob.	*1.0000*	*0.6803*	*0.6854*	*1.0000*
		n0	n0	n0	n0
	At First Difference				
		d(LTPN)	d(LATMP)	d(LARF)	d(LCOE)
With Constant	t-Statistic	-11.9498	-10.9029	-17.7659	-5.5686
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0000*
		***	***	***	***
With Constant & Trend	t-Statistic	-12.8771	-11.3922	-17.4799	-5.4664
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0003*
		***	***	***	***
Without Constant & Trend	t-Statistic	-10.0156	-11.0511	-17.9944	-3.1916
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0021*
		***	***	***	***

Notes:
a: (*)Significant at the 10%; (**)Significant at the 5%; (***) Significant at the 1% and (no) Not Significant
b: Lag Length based on SIC
c: Probability based on MacKinnon (1996) one-sided p-values.

This Result is The Out-Put of Program Has Developed By:
Dr. Imadeddin AlMosabbeh
College of Business and Economics
Qassim University-KSA

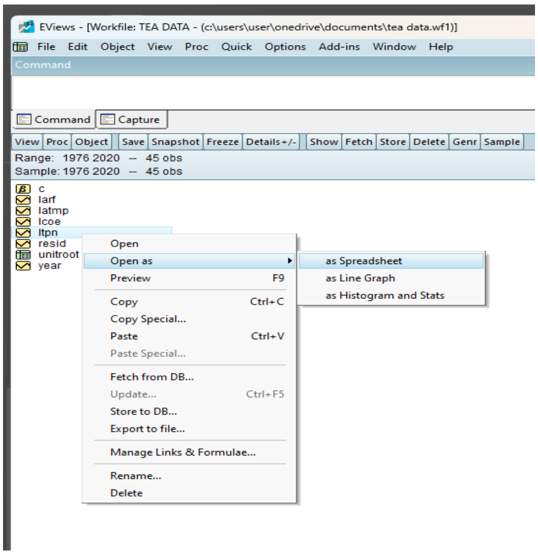


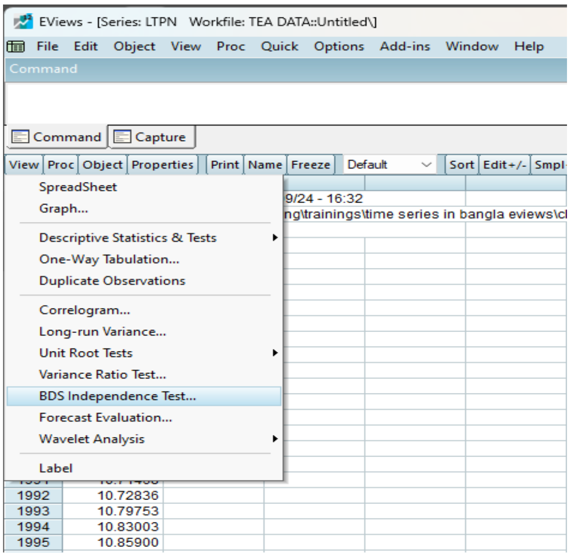
UNIT ROOT TEST RESULTS TABLE (ADF)					
Null Hypothesis: the variable has a unit root					
	At Level				
		LTPN	LATMP	LARF	LCOE
With Constant	t-Statistic	0.8243	-3.8925	-6.5517	0.0804
	Prob.	*0.9934*	*0.0044*	*0.0000*	*0.9606*
		n0	***	***	n0
With Constant & Trend	t-Statistic	-4.0226	-4.0320	-7.7195	-1.9955
	Prob.	*0.0150*	*0.0146*	*0.0000*	*0.5875*
		**	**	***	n0
Without Constant & Trend	t-Statistic	3.0057	0.1349	-0.2327	6.8879
	Prob.	*0.9991*	*0.7198*	*0.5964*	*1.0000*
		n0	n0	n0	n0
	At First Difference				
		d(LTPN)	d(LATMP)	d(LARF)	d(LCOE)
With Constant	t-Statistic	-11.7643	-7.7058	-9.3298	-5.6522
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0000*
		***	***	***	***
With Constant & Trend	t-Statistic	-11.8984	-7.7240	-9.2053	-5.5728
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0002*
		***	***	***	***
Without Constant & Trend	t-Statistic	-10.4399	-7.8006	-9.4499	-3.3415
	Prob.	*0.0000*	*0.0000*	*0.0000*	*0.0013*
		***	***	***	***

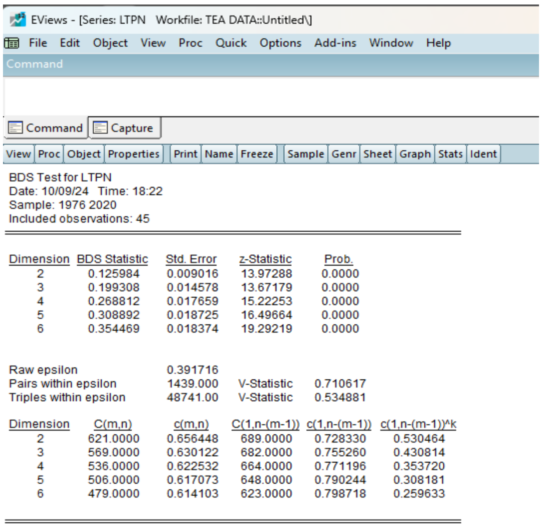
Notes:
a: (*)Significant at the 10%; (**)Significant at the 5%; (***) Significant at the 1% and (no) Not Significant
b: Lag Length based on SIC
c: Probability based on MacKinnon (1996) one-sided p-values.

This Result is The Out-Put of Program Has Developed By:
Dr. Imadeddin AlMosabbeh
College of Business and Economics
Qassim University-KSA

BDS test for nonlinearity







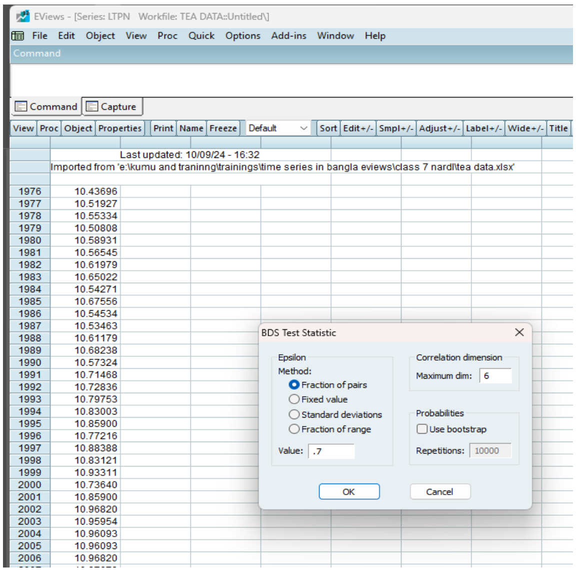
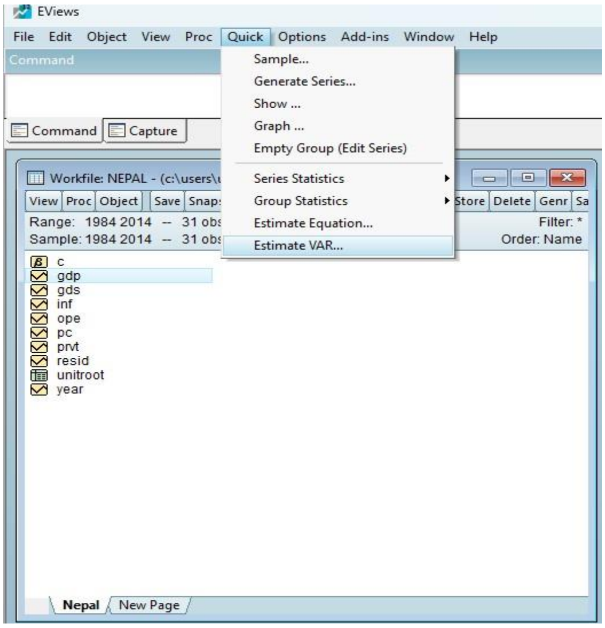


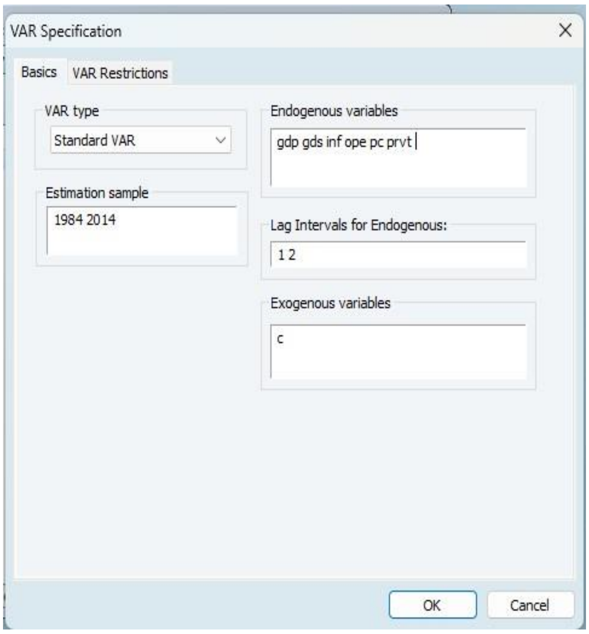
Table 6 BDS test for nonlinearity

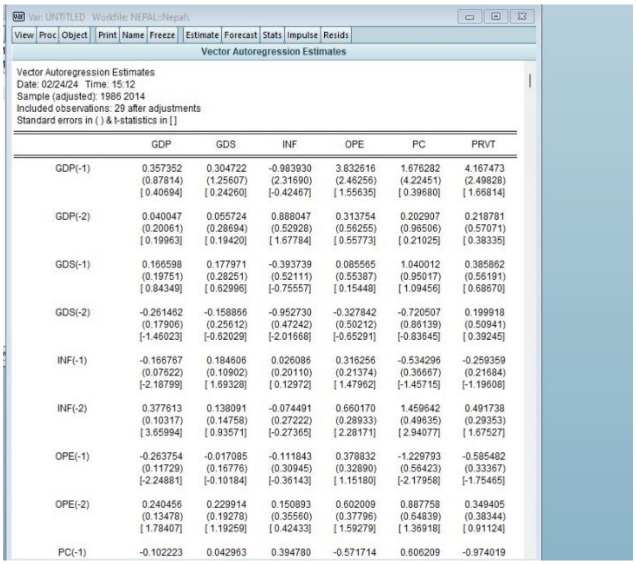
Variables	BDS statistics				
	D=2	D=3	D=4	D=5	D=6
LTPN	0.125***	0.199***	0.268***	0.308***	0.354***
LATMP	0.029***	0.045***	**0.038****	**0.039****	**0.031***
LARF	−0.001	0.002	0.002	−0.001	0.005
LCOE	0.189***	0.316***	0.401***	0.460***	0.504***

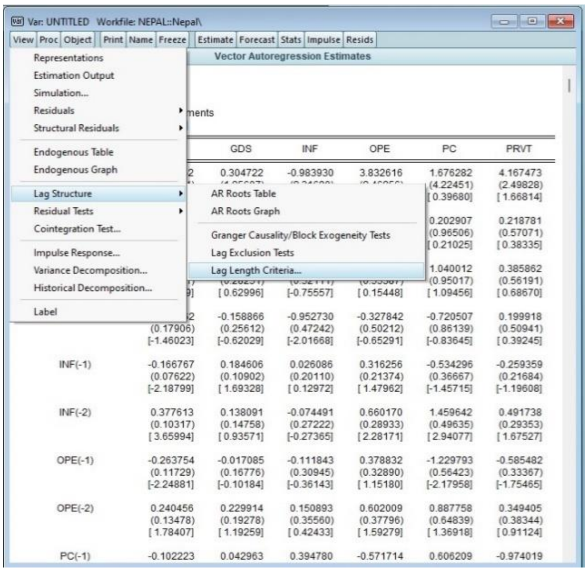
The differences in statistical significance could imply varying degrees of non-linearity or relationships between the variables being analyzed

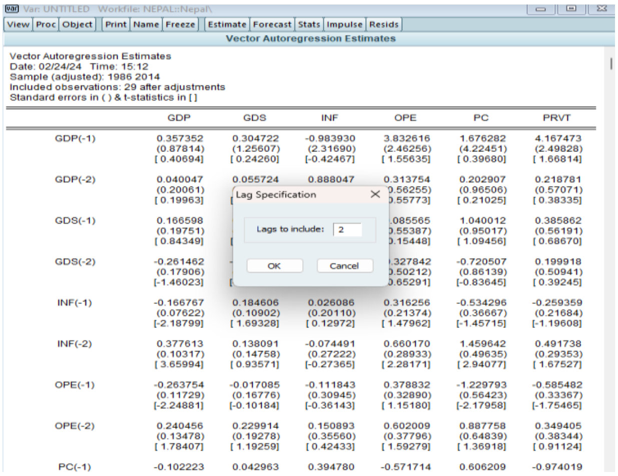
Select Optimum Lag Length





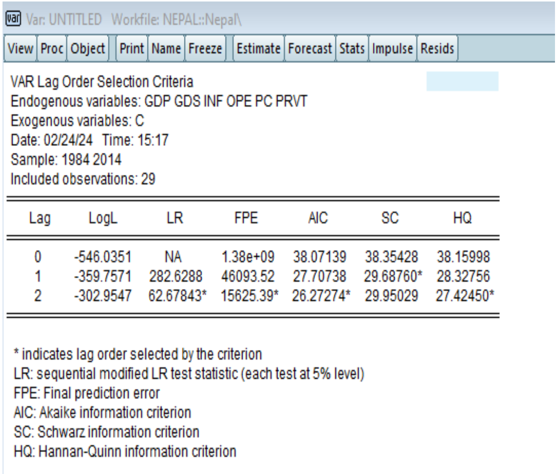






Vector Autoregression Estimates
Date: 02/24/24 Time: 15:12
Sample (adjusted): 1986 2014
Included observations: 29 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	GDP	GDS	INF	OPE	PC	PRVT
GDP(-1)	0.357352	0.304722	-0.983930	3.832616	1.676282	4.167473
	(0.87814)	(1.25607)	(2.31690)	(2.46256)	(4.22451)	(2.49828)
	[0.40694]	[0.24260]	[-0.42467]	[1.55635]	[0.39680]	[1.66814]
GDP(-2)	0.040047	0.055724	0.888047	0.313754	0.202907	0.218781
	(0.20061)	[illegible]	[illegible]	.56255)	(0.96506)	(0.57071)
	[0.19963]	[illegible]	[illegible]	.55773]	[0.21025]	[0.38335]
GDS(-1)	0.166598	[illegible]	[illegible]	.085565	1.040012	0.385862
	(0.19751)	[illegible]	[illegible]	.55387)	(0.95017)	(0.56191)
	[0.84349]	[illegible]	[illegible]	.15448]	[1.09456]	[0.68670]
GDS(-2)	-0.261462	[illegible]	[illegible]	.327842	-0.720507	0.199918
	(0.17906)	[illegible]	[illegible]	.50212)	(0.86139)	(0.50941)
	[-1.46023]	[illegible]	[illegible]	.65291]	[-0.83645]	[0.39245]
INF(-1)	-0.166767	0.184606	0.026086	0.316256	-0.534296	-0.259359
	(0.07622)	(0.10902)	(0.20110)	(0.21374)	(0.36667)	(0.21684)
	[-2.18799]	[1.69328]	[0.12972]	[1.47962]	[-1.45715]	[-1.19608]
INF(-2)	0.377613	0.138091	-0.074491	0.660170	1.459642	0.491738
	(0.10317)	(0.14758)	(0.27222)	(0.28933)	(0.49635)	(0.29353)
	[3.65994]	[0.93571]	[-0.27365]	[2.28171]	[2.94077]	[1.67527]
OPE(-1)	-0.263754	-0.017085	-0.111843	0.378832	-1.229793	-0.585482
	(0.11729)	(0.16776)	(0.30945)	(0.32890)	(0.56423)	(0.33367)
	[-2.24881]	[-0.10184]	[-0.36143]	[1.15180]	[-2.17958]	[-1.75465]
OPE(-2)	0.240456	0.229914	0.150893	0.602009	0.887758	0.349405
	(0.13478)	(0.19278)	(0.35560)	(0.37796)	(0.64839)	(0.38344)
	[1.78407]	[1.19259]	[0.42433]	[1.59279]	[1.36918]	[0.91124]
PC(-1)	-0.102223	0.042963	0.394780	-0.571714	0.606209	-0.974019

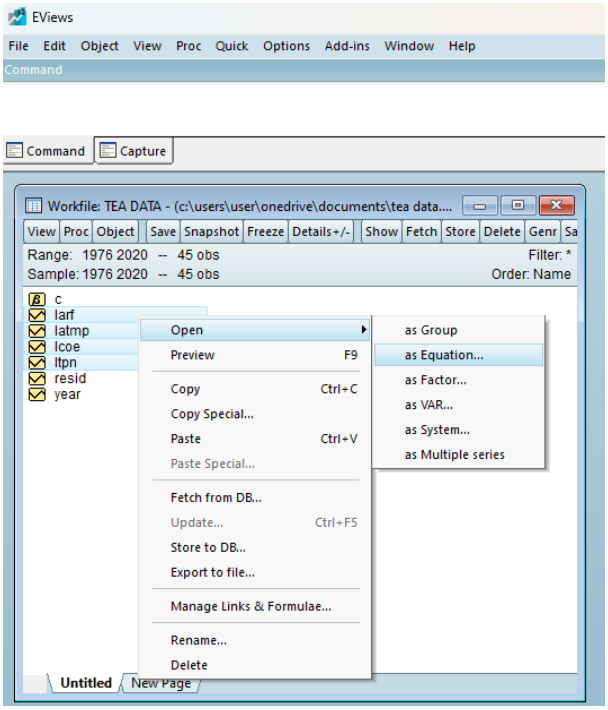


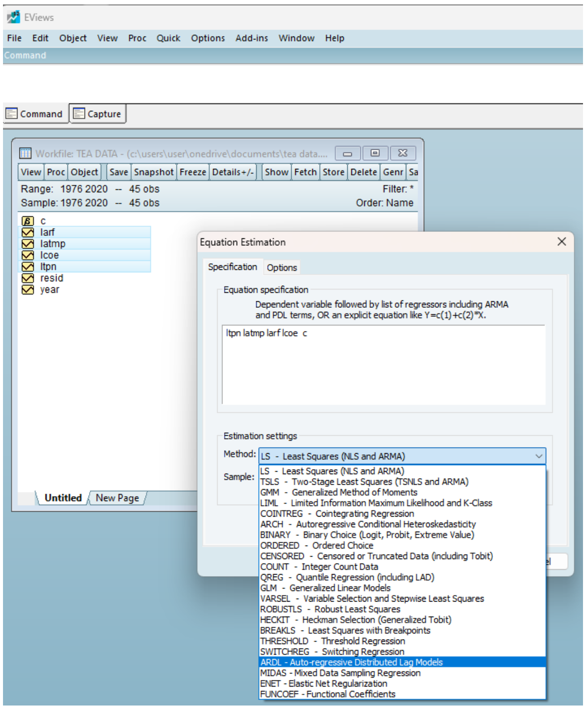
VAR Lag Order Selection Criteria
Endogenous variables: GDP GDS INF OPE PC PRVT
Exogenous variables: C
Date: 02/24/24 Time: 15:17
Sample: 1984 2014
Included observations: 29

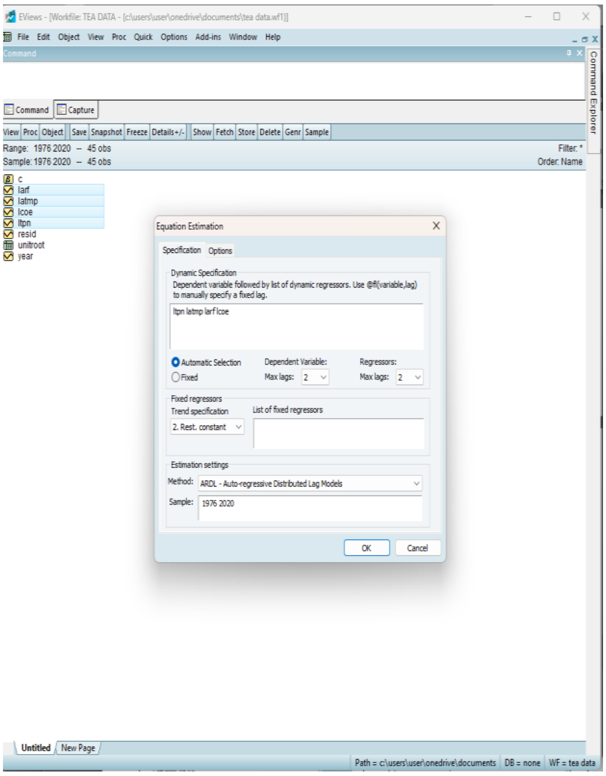
Lag	LogL	LR	FPE	AIC	SC	HQ
0	-546.0351	NA	1.38e+09	38.07139	38.35428	38.15998
1	-359.7571	282.6288	46093.52	27.70738	29.68760*	28.32756
2	-302.9547	62.67843*	15625.39*	26.27274*	29.95029	27.42450*

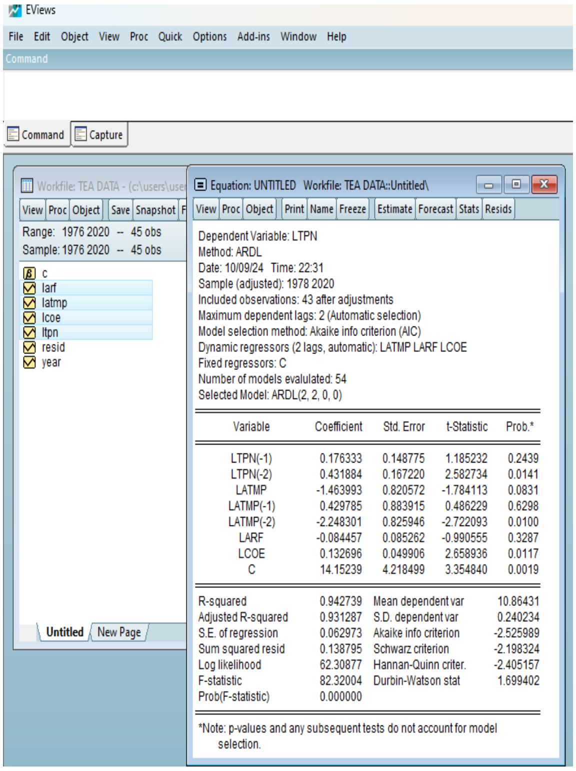
* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

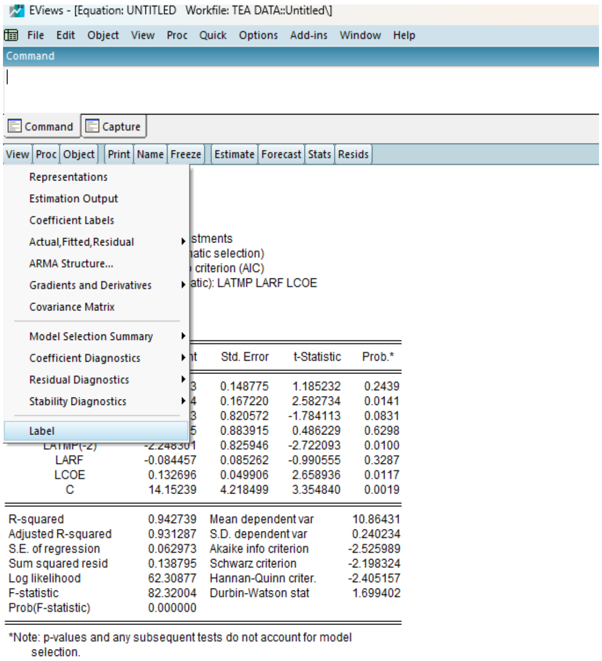
Specify ARDL model with optimum Lag

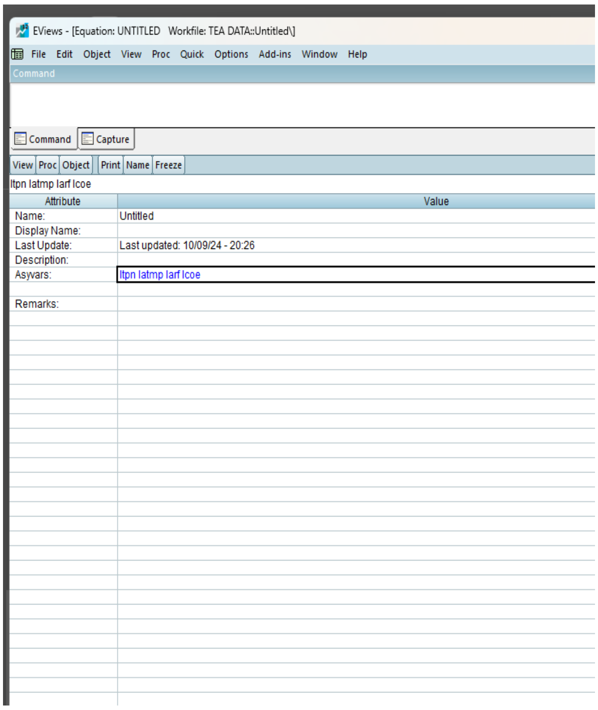


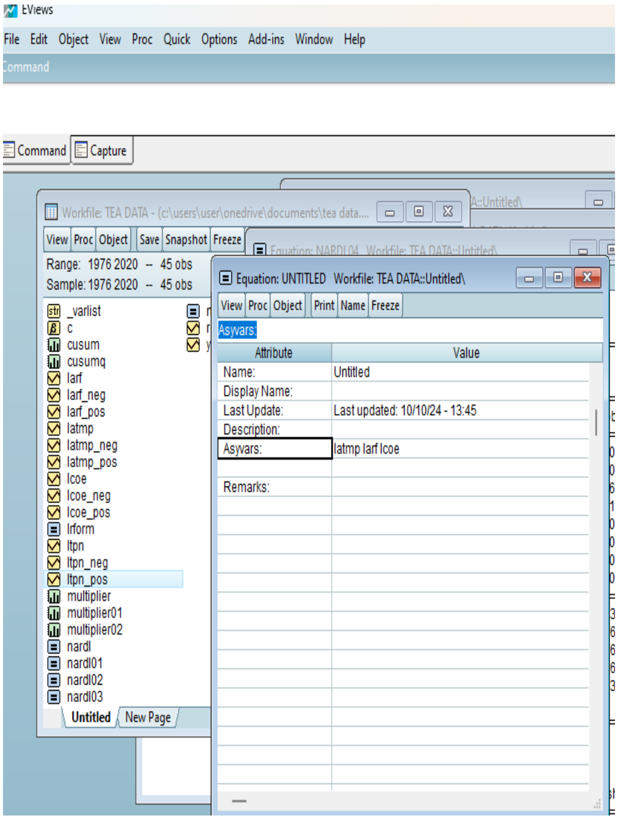


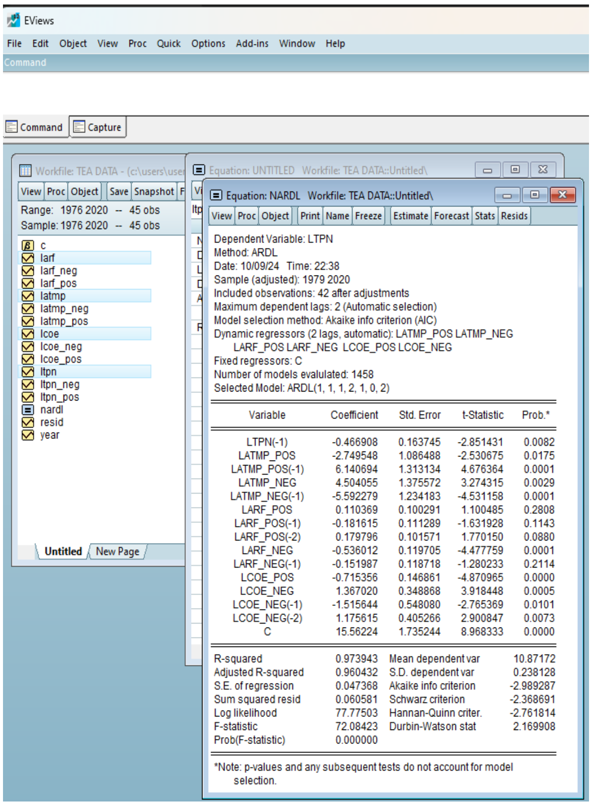

EViews - [Workfile: TEA DATA - (c:\users\user\onedrive\documents\tea data.wf1)]
File Edit Object View Proc Quick Options Add-ins Window Help
Command
Command Capture
View Proc Object Save Snapshot Freeze Details+/- Show Fetch Store Delete Genr Sample
Range: 1976 2020 -- 45 obs
Sample: 1976 2020 -- 45 obs
Filter: *
Order: Name
c
larf
latmp
lcoe
ltpn
resid
unitroot
year
Equation Estimation
Specification Options
Dynamic Specification
Dependent variable followed by list of dynamic regressors. Use @fl(variable,lag) to manually specify a fixed lag.
ltpn latmp larf lcoe
Automatic Selection
Fixed
Dependent Variable:
Max lags: 2
Regressors:
Max lags: 2
Fixed regressors
Trend specification
2. Rest. constant
List of fixed regressors
Estimation settings
Method: ARDL - Auto-regressive Distributed Lag Models
Sample: 1976 2020
OK
Cancel
Untitled New Page
Path = c:\users\user\onedrive\documents DB = none WF = tea data

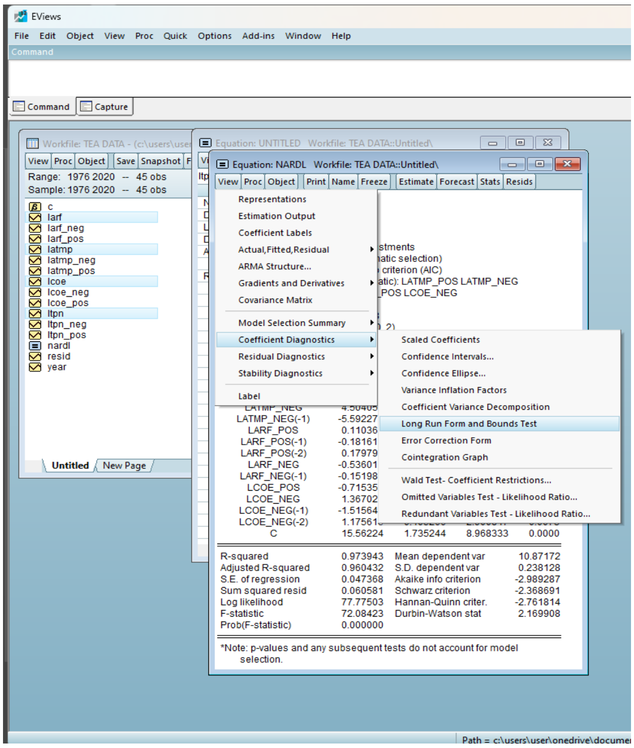

EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Command
Command Capture
Workfile: TEA DATA - (c:\users\use
View Proc Object Save Snapshot
Range: 1976 2020 -- 45 obs
Sample: 1976 2020 -- 45 obs
c
larf
latmp
lcoe
ltpn
resid
year
Untitled New Page
Equation: UNTITLED Workfile: TEA DATA::Untitled\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Dependent Variable: LTPN
Method: ARDL
Date: 10/09/24 Time: 22:31
Sample (adjusted): 1978 2020
Included observations: 43 after adjustments
Maximum dependent lags: 2 (Automatic selection)
Model selection method: Akaike info criterion (AIC)
Dynamic regressors (2 lags, automatic): LATMP LARF LCOE
Fixed regressors: C
Number of models evalulated: 54
Selected Model: ARDL(2, 2, 0, 0)
Variable Coefficient Std. Error t-Statistic Prob.*
LTPN(-1) 0.176333 0.148775 1.185232 0.2439
LTPN(-2) 0.431884 0.167220 2.582734 0.0141
LATMP -1.463993 0.820572 -1.784113 0.0831
LATMP(-1) 0.429785 0.883915 0.486229 0.6298
LATMP(-2) -2.248301 0.825946 -2.722093 0.0100
LARF -0.084457 0.085262 -0.990555 0.3287
LCOE 0.132696 0.049906 2.658936 0.0117
C 14.15239 4.218499 3.354840 0.0019
R-squared 0.942739 Mean dependent var 10.86431
Adjusted R-squared 0.931287 S.D. dependent var 0.240234
S.E. of regression 0.062973 Akaike info criterion -2.525989
Sum squared resid 0.138795 Schwarz criterion -2.198324
Log likelihood 62.30877 Hannan-Quinn criter. -2.405157
F-statistic 82.32004 Durbin-Watson stat 1.699402
Prob(F-statistic) 0.000000
*Note: p-values and any subsequent tests do not account for model selection.

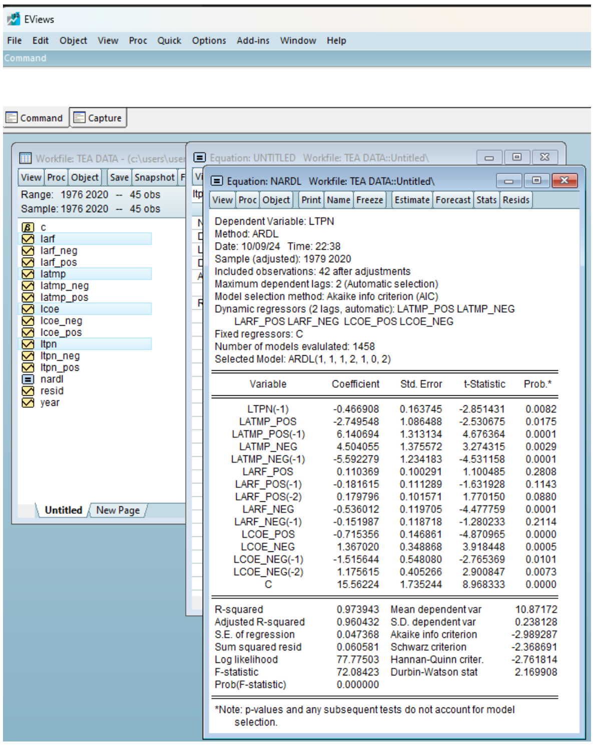

EViews - [Equation: UNTITLED Workfile: TEA DATA::Untitled\]
File Edit Object View Proc Quick Options Add-ins Window Help
Command
Command Capture
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Representations
Estimation Output
Coefficient Labels
Actual,Fitted,Residual
ARMA Structure...
Gradients and Derivatives
Covariance Matrix
Model Selection Summary
Coefficient Diagnostics
Residual Diagnostics
Stability Diagnostics
Label
stments
atic selection)
criterion (AIC)
atic): LATMP LARF LCOE
Std. Error t-Statistic Prob.*
0.148775 1.185232 0.2439
0.167220 2.582734 0.0141
0.820572 -1.784113 0.0831
0.883915 0.486229 0.6298
0.825946 -2.722093 0.0100
LARF -0.084457 0.085262 -0.990555 0.3287
LCOE 0.132696 0.049906 2.658936 0.0117
C 14.15239 4.218499 3.354840 0.0019
R-squared 0.942739 Mean dependent var 10.86431
Adjusted R-squared 0.931287 S.D. dependent var 0.240234
S.E. of regression 0.062973 Akaike info criterion -2.525989
Sum squared resid 0.138795 Schwarz criterion -2.198324
Log likelihood 62.30877 Hannan-Quinn criter. -2.405157
F-statistic 82.32004 Durbin-Watson stat 1.699402
Prob(F-statistic) 0.000000
*Note: p-values and any subsequent tests do not account for model selection.

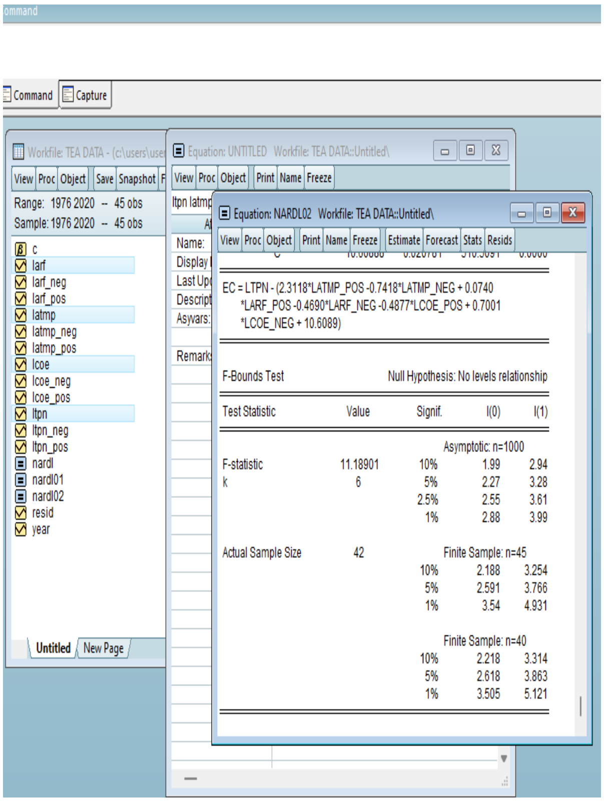

EViews - [Equation: UNTITLED Workfile: TEA DATA::Untitled\]
File Edit Object View Proc Quick Options Add-ins Window Help
Command
Command Capture
View Proc Object Print Name Freeze
ltpn latmp larf lcoe
Attribute Value
Name: Untitled
Display Name:
Last Update: Last updated: 10/09/24 - 20:26
Description:
Asyvars: ltpn latmp larf lcoe
Remarks:

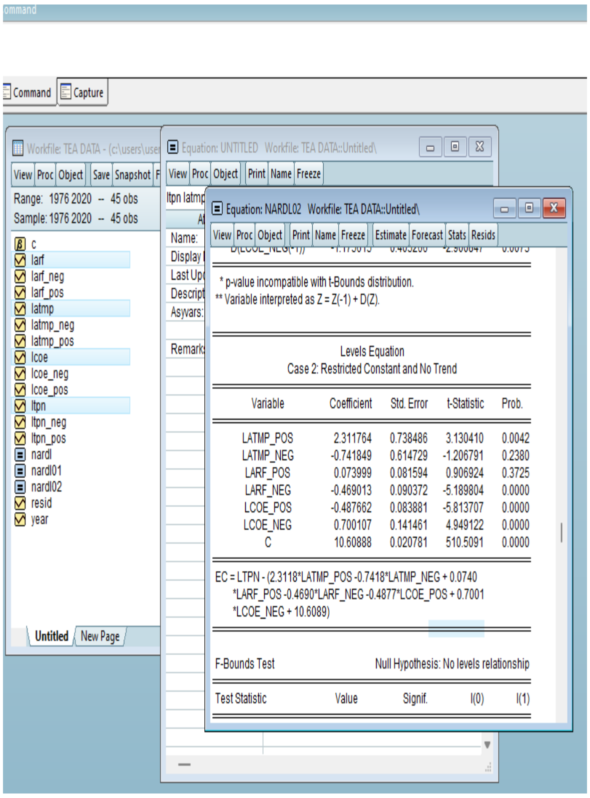

EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Command
Command Capture
Workfile: TEA DATA - (c:\users\user\onedrive\documents\tea data....
View Proc Object Save Snapshot Freeze
Range: 1976 2020 -- 45 obs
Sample: 1976 2020 -- 45 obs
_varlist
c
cusum
cusumq
larf
larf_neg
larf_pos
latmp
latmp_neg
latmp_pos
lcoe
lcoe_neg
lcoe_pos
lrform
ltpn
ltpn_neg
ltpn_pos
multiplier
multiplier01
multiplier02
nardl
nardl01
nardl02
nardl03
Untitled New Page
Equation: UNTITLED Workfile: TEA DATA::Untitled\
View Proc Object Print Name Freeze
Asyvars:
Attribute
Value
Name:
Untitled
Display Name:
Last Update:
Last updated: 10/10/24 - 13:45
Description:
Asyvars:
latmp larf lcoe
Remarks:

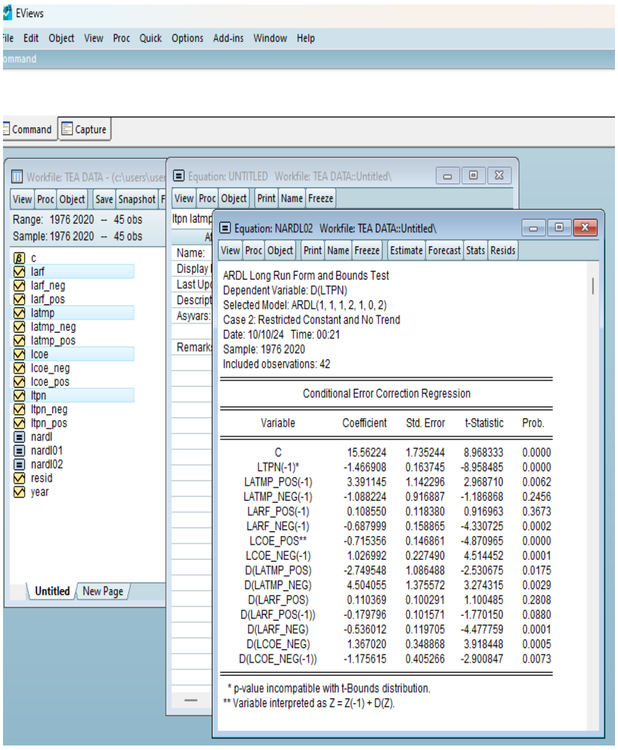

EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Command
Command Capture
Workfile: TEA DATA - (c:\users\use
Range: 1976 2020 -- 45 obs
Sample: 1976 2020 -- 45 obs
Equation: NARDL Workfile: TEA DATA::Untitled\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Dependent Variable: LTPN
Method: ARDL
Date: 10/09/24 Time: 22:38
Sample (adjusted): 1979 2020
Included observations: 42 after adjustments
Maximum dependent lags: 2 (Automatic selection)
Model selection method: Akaike info criterion (AIC)
Dynamic regressors (2 lags, automatic): LATMP_POS LATMP_NEG
LARF_POS LARF_NEG LCOE_POS LCOE_NEG
Fixed regressors: C
Number of models evaluated: 1458
Selected Model: ARDL(1, 1, 1, 2, 1, 0, 2)
Variable Coefficient Std. Error t-Statistic Prob.*
LTPN(-1) -0.466908 0.163745 -2.851431 0.0082
LATMP_POS -2.749548 1.086488 -2.530675 0.0175
LATMP_POS(-1) 6.140694 1.313134 4.676364 0.0001
LATMP_NEG 4.504055 1.375572 3.274315 0.0029
LATMP_NEG(-1) -5.592279 1.234183 -4.531158 0.0001
LARF_POS 0.110369 0.100291 1.100485 0.2808
LARF_POS(-1) -0.181615 0.111289 -1.631928 0.1143
LARF_POS(-2) 0.179796 0.101571 1.770150 0.0880
LARF_NEG -0.536012 0.119705 -4.477759 0.0001
LARF_NEG(-1) -0.151987 0.118718 -1.280233 0.2114
LCOE_POS -0.715356 0.146861 -4.870965 0.0000
LCOE_NEG 1.367020 0.348868 3.918448 0.0005
LCOE_NEG(-1) -1.515644 0.548080 -2.765369 0.0101
LCOE_NEG(-2) 1.175615 0.405266 2.900847 0.0073
C 15.56224 1.735244 8.968333 0.0000
R-squared 0.973943 Mean dependent var 10.87172
Adjusted R-squared 0.960432 S.D. dependent var 0.238128
S.E. of regression 0.047368 Akaike info criterion -2.989287
Sum squared resid 0.060581 Schwarz criterion -2.368691
Log likelihood 77.77503 Hannan-Quinn criter. -2.761814
F-statistic 72.08423 Durbin-Watson stat 2.169908
Prob(F-statistic) 0.000000
*Note: p-values and any subsequent tests do not account for model selection.

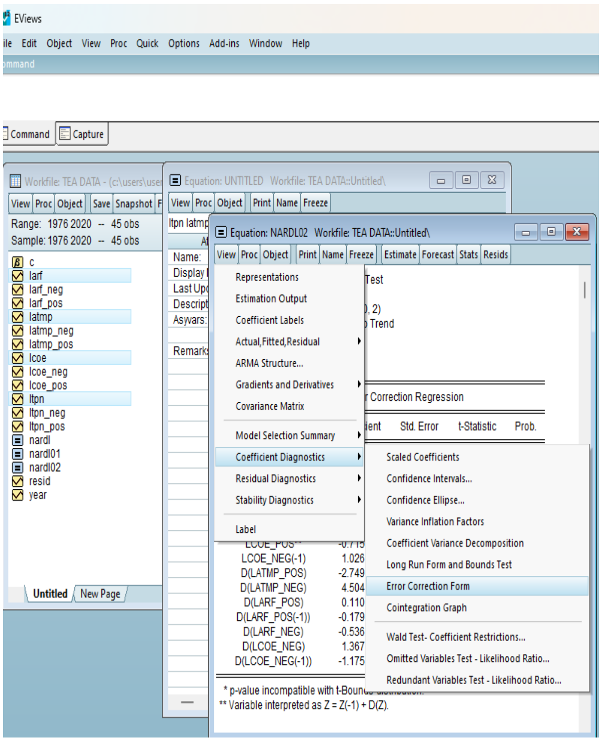

EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Command
Command Capture
Equation: NARDL Workfile: TEA DATA::Untitled\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Representations
Estimation Output
Coefficient Labels
Actual,Fitted,Residual
ARMA Structure...
Gradients and Derivatives
Covariance Matrix
Model Selection Summary
Coefficient Diagnostics
Residual Diagnostics
Stability Diagnostics
Label
Scaled Coefficients
Confidence Intervals...
Confidence Ellipse...
Variance Inflation Factors
Coefficient Variance Decomposition
Long Run Form and Bounds Test
Error Correction Form
Cointegration Graph
Wald Test- Coefficient Restrictions...
Omitted Variables Test - Likelihood Ratio...
Redundant Variables Test - Likelihood Ratio...
R-squared 0.973943 Mean dependent var 10.87172
Adjusted R-squared 0.960432 S.D. dependent var 0.238128
S.E. of regression 0.047368 Akaike info criterion -2.989287
Sum squared resid 0.060581 Schwarz criterion -2.368691
Log likelihood 77.77503 Hannan-Quinn criter. -2.761814
F-statistic 72.08423 Durbin-Watson stat 2.169908
Prob(F-statistic) 0.000000
*Note: p-values and any subsequent tests do not account for model selection.
Path = c:\users\user\onedrive\docume

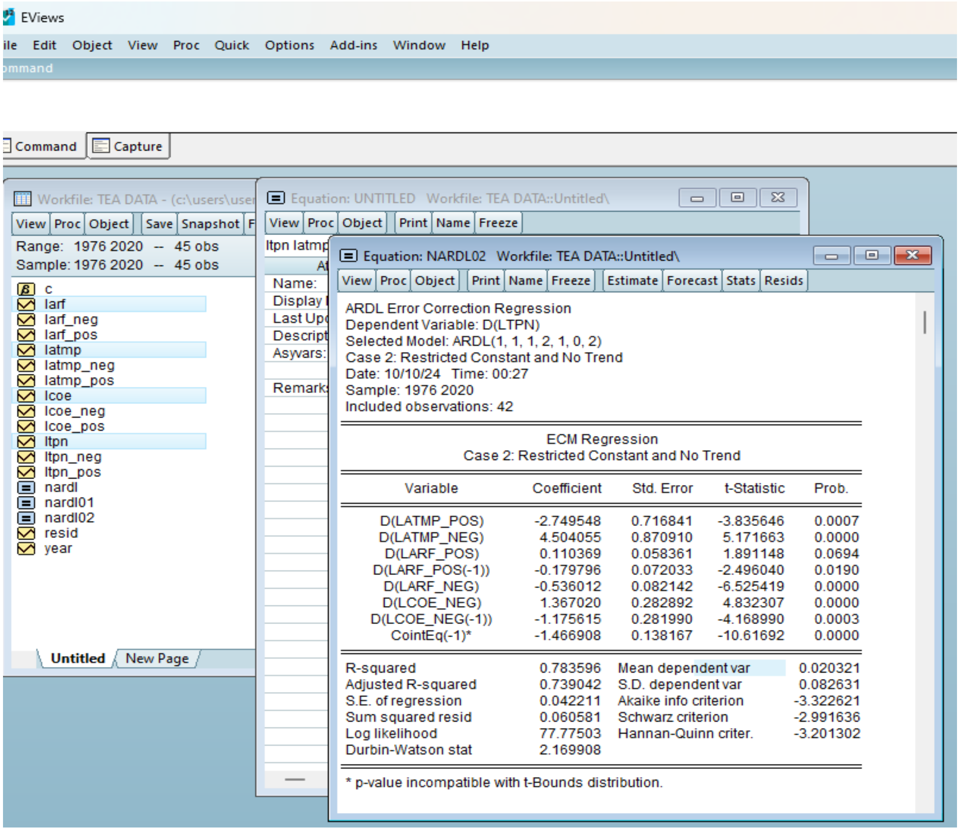

EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Command
Command Capture
Equation: NARDL Workfile: TEA DATA::Untitled\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Dependent Variable: LTPN
Method: ARDL
Date: 10/09/24 Time: 22:38
Sample (adjusted): 1979 2020
Included observations: 42 after adjustments
Maximum dependent lags: 2 (Automatic selection)
Model selection method: Akaike info criterion (AIC)
Dynamic regressors (2 lags, automatic): LATMP_POS LATMP_NEG
LARF_POS LARF_NEG LCOE_POS LCOE_NEG
Fixed regressors: C
Number of models evaluated: 1458
Selected Model: ARDL(1, 1, 1, 2, 1, 0, 2)
Variable Coefficient Std. Error t-Statistic Prob.*
LTPN(-1) -0.466908 0.163745 -2.851431 0.0082
LATMP_POS -2.749548 1.086488 -2.530675 0.0175
LATMP_POS(-1) 6.140694 1.313134 4.676364 0.0001
LATMP_NEG 4.504055 1.375572 3.274315 0.0029
LATMP_NEG(-1) -5.592279 1.234183 -4.531158 0.0001
LARF_POS 0.110369 0.100291 1.100485 0.2808
LARF_POS(-1) -0.181615 0.111289 -1.631928 0.1143
LARF_POS(-2) 0.179796 0.101571 1.770150 0.0880
LARF_NEG -0.536012 0.119705 -4.477759 0.0001
LARF_NEG(-1) -0.151987 0.118718 -1.280233 0.2114
LCOE_POS -0.715356 0.146861 -4.870965 0.0000
LCOE_NEG 1.367020 0.348868 3.918448 0.0005
LCOE_NEG(-1) -1.515644 0.548080 -2.765369 0.0101
LCOE_NEG(-2) 1.175615 0.405266 2.900847 0.0073
C 15.56224 1.735244 8.968333 0.0000
R-squared 0.973943 Mean dependent var 10.87172
Adjusted R-squared 0.960432 S.D. dependent var 0.238128
S.E. of regression 0.047368 Akaike info criterion -2.989287
Sum squared resid 0.060581 Schwarz criterion -2.368691
Log likelihood 77.77503 Hannan-Quinn criter. -2.761814
F-statistic 72.08423 Durbin-Watson stat 2.169908
Prob(F-statistic) 0.000000
*Note: p-values and any subsequent tests do not account for model selection.


Equation: NARDL02 Workfile: TEA DATA::Untitled\
EC = LTPN - (2.3118*LATMP_POS -0.7418*LATMP_NEG + 0.0740
*LARF_POS -0.4690*LARF_NEG -0.4877*LCOE_POS + 0.7001
*LCOE_NEG + 10.6089)
F-Bounds Test
Null Hypothesis: No levels relationship
Test Statistic Value Signif. I(0) I(1)
Asymptotic: n=1000
F-statistic 11.18901 10% 1.99 2.94
k 6 5% 2.27 3.28
2.5% 2.55 3.61
1% 2.88 3.99
Actual Sample Size 42
Finite Sample: n=45
10% 2.188 3.254
5% 2.591 3.766
1% 3.54 4.931
Finite Sample: n=40
10% 2.218 3.314
5% 2.618 3.863
1% 3.505 5.121


Equation: NARDL02 Workfile: TEA DATA::Untitled\
* p-value incompatible with t-Bounds distribution.
** Variable interpreted as Z = Z(-1) + D(Z).
Levels Equation
Case 2: Restricted Constant and No Trend
Variable Coefficient Std. Error t-Statistic Prob.
LATMP_POS 2.311764 0.738486 3.130410 0.0042
LATMP_NEG -0.741849 0.614729 -1.206791 0.2380
LARF_POS 0.073999 0.081594 0.906924 0.3725
LARF_NEG -0.469013 0.090372 -5.189804 0.0000
LCOE_POS -0.487662 0.083881 -5.813707 0.0000
LCOE_NEG 0.700107 0.141461 4.949122 0.0000
C 10.60888 0.020781 510.5091 0.0000
EC = LTPN - (2.3118*LATMP_POS -0.7418*LATMP_NEG + 0.0740
*LARF_POS -0.4690*LARF_NEG -0.4877*LCOE_POS + 0.7001
*LCOE_NEG + 10.6089)
F-Bounds Test
Null Hypothesis: No levels relationship
Test Statistic Value Signif. I(0) I(1)


Equation: NARDL02 Workfile: TEA DATA::Untitled\
ARDL Long Run Form and Bounds Test
Dependent Variable: D(LTPN)
Selected Model: ARDL(1, 1, 1, 2, 1, 0, 2)
Case 2: Restricted Constant and No Trend
Date: 10/10/24 Time: 00:21
Sample: 1976 2020
Included observations: 42
Conditional Error Correction Regression
Variable Coefficient Std. Error t-Statistic Prob.
C 15.56224 1.735244 8.968333 0.0000
LTPN(-1)* -1.466908 0.163745 -8.958485 0.0000
LATMP_POS(-1) 3.391145 1.142296 2.968710 0.0062
LATMP_NEG(-1) -1.088224 0.916887 -1.186868 0.2456
LARF_POS(-1) 0.108550 0.118380 0.916963 0.3673
LARF_NEG(-1) -0.687999 0.158865 -4.330725 0.0002
LCOE_POS** -0.715356 0.146861 -4.870965 0.0000
LCOE_NEG(-1) 1.026992 0.227490 4.514452 0.0001
D(LATMP_POS) -2.749548 1.086488 -2.530675 0.0175
D(LATMP_NEG) 4.504055 1.375572 3.274315 0.0029
D(LARF_POS) 0.110369 0.100291 1.100485 0.2808
D(LARF_POS(-1)) -0.179796 0.101571 -1.770150 0.0880
D(LARF_NEG) -0.536012 0.119705 -4.477759 0.0001
D(LCOE_NEG) 1.367020 0.348868 3.918448 0.0005
D(LCOE_NEG(-1)) -1.175615 0.405266 -2.900847 0.0073
* p-value incompatible with t-Bounds distribution.
** Variable interpreted as Z = Z(-1) + D(Z).


Equation: NARDL02 Workfile: TEA DATA::Untitled\
Representations
Estimation Output
Coefficient Labels
Actual,Fitted,Residual
ARMA Structure...
Gradients and Derivatives
Covariance Matrix
Model Selection Summary
Coefficient Diagnostics
Residual Diagnostics
Stability Diagnostics
Label
Scaled Coefficients
Confidence Intervals...
Confidence Ellipse...
Variance Inflation Factors
Coefficient Variance Decomposition
Long Run Form and Bounds Test
Error Correction Form
Cointegration Graph
Wald Test- Coefficient Restrictions...
Omitted Variables Test - Likelihood Ratio...
Redundant Variables Test - Likelihood Ratio...



ARDL Error Correction Regression
Dependent Variable: D(LTPN)
Selected Model: ARDL(1, 1, 1, 2, 1, 0, 2)
Case 2: Restricted Constant and No Trend
Date: 10/10/24 Time: 00:27
Sample: 1976 2020
Included observations: 42

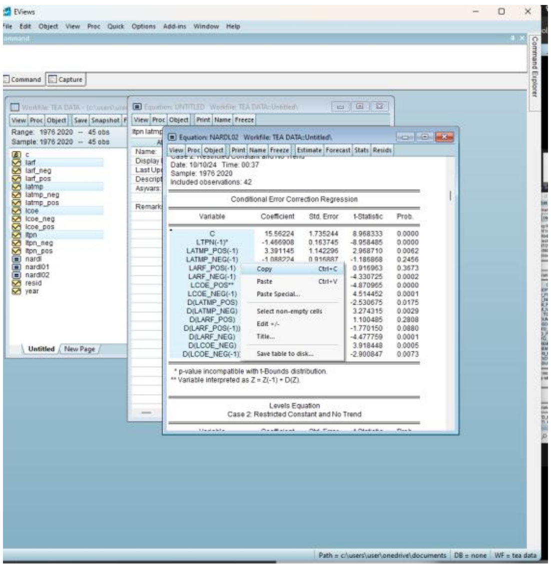
ECM Regression
Case 2: Restricted Constant and No Trend

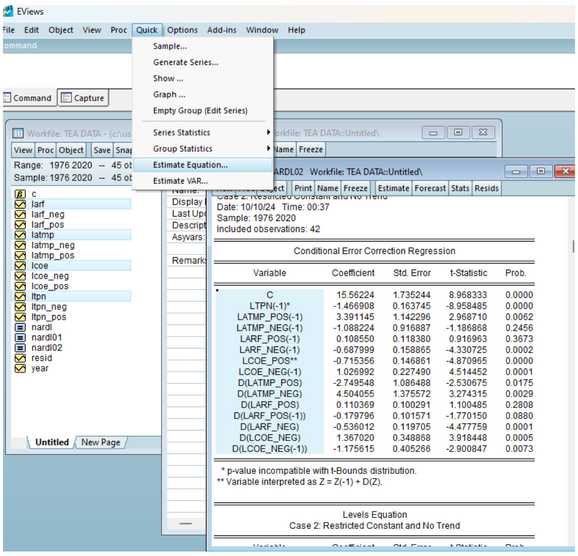
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LATMP_POS)	-2.749548	0.716841	-3.835646	0.0007
D(LATMP_NEG)	4.504055	0.870910	5.171663	0.0000
D(LARF_POS)	0.110369	0.058361	1.891148	0.0694
D(LARF_POS(-1))	-0.179796	0.072033	-2.496040	0.0190
D(LARF_NEG)	-0.536012	0.082142	-6.525419	0.0000
D(LCOE_NEG)	1.367020	0.282892	4.832307	0.0000
D(LCOE_NEG(-1))	-1.175615	0.281990	-4.168990	0.0003
CointEq(-1)*	-1.466908	0.138167	-10.61692	0.0000

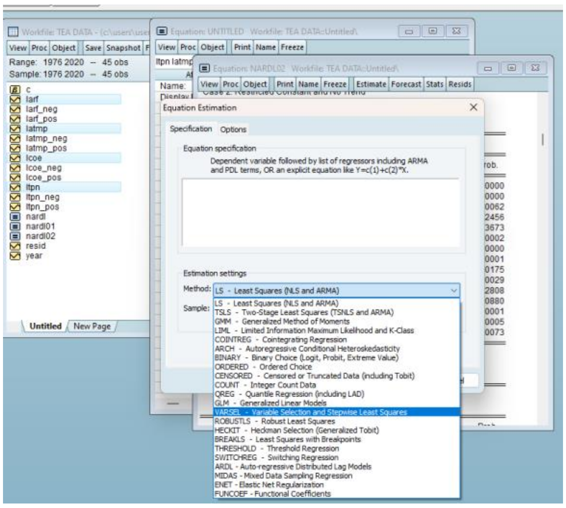
R-squared	0.783596	Mean dependent var	0.020321
Adjusted R-squared	0.739042	S.D. dependent var	0.082631
S.E. of regression	0.042211	Akaike info criterion	-3.322621
Sum squared resid	0.060581	Schwarz criterion	-2.991636
Log likelihood	77.77503	Hannan-Quinn criter.	-3.201302
Durbin-Watson stat	2.169908		

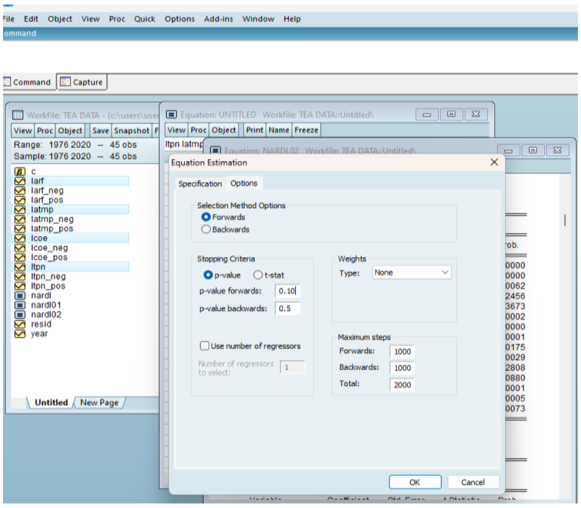
* p-value incompatible with t-Bounds distribution.

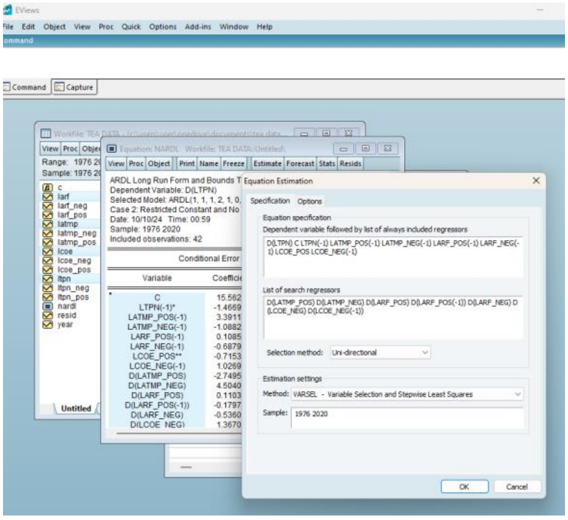
Testing the presence of asymmetries using Wald test

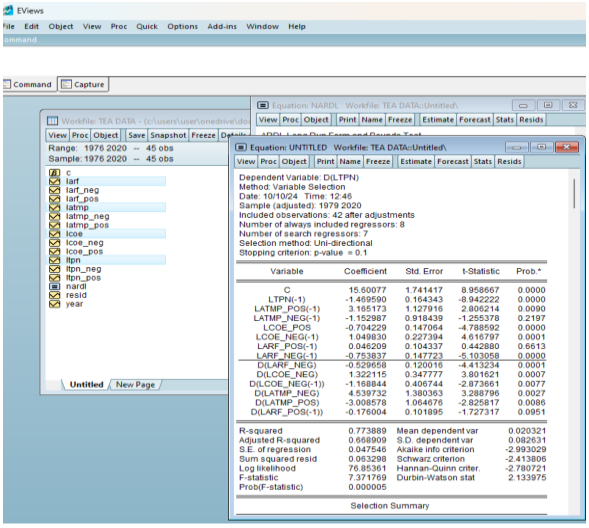


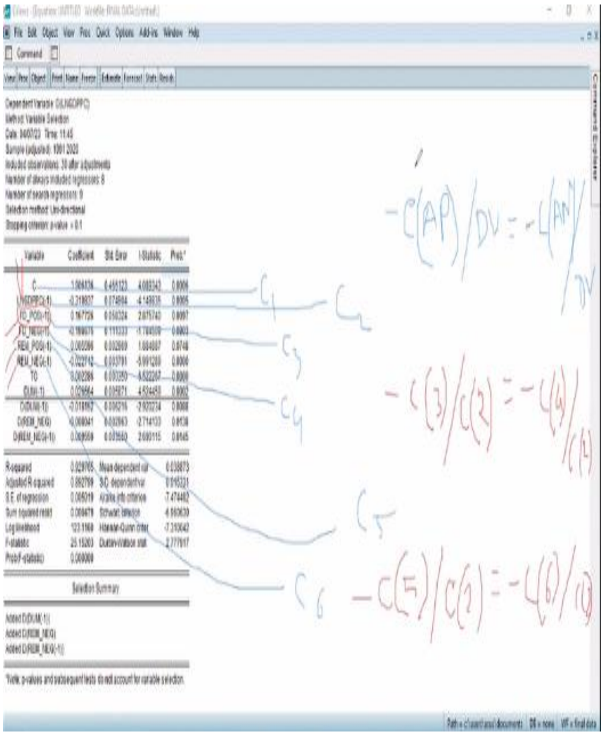


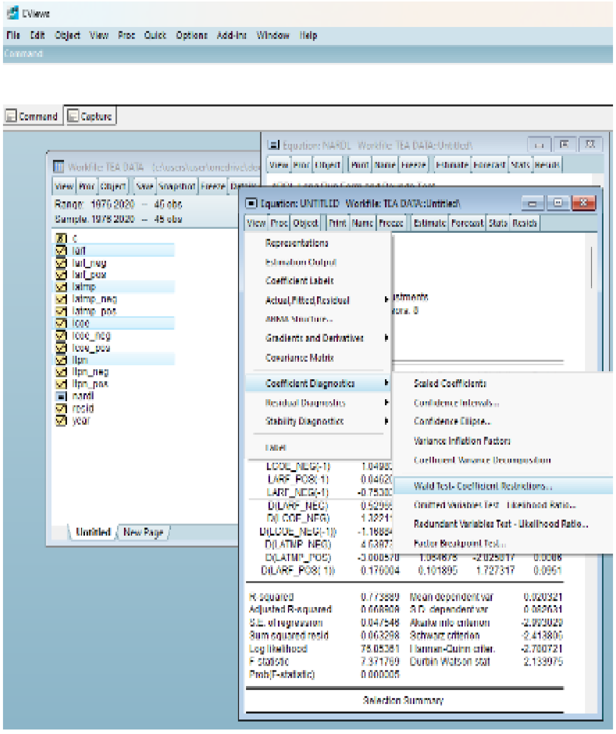


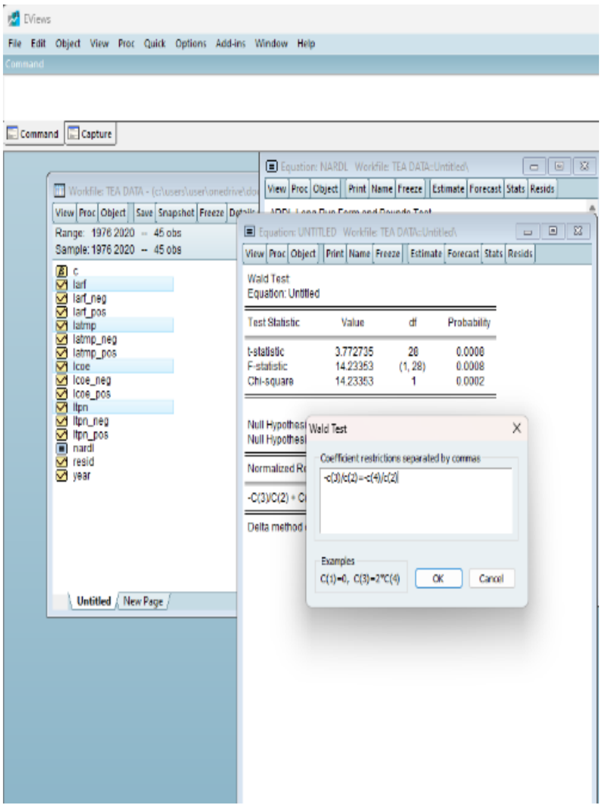


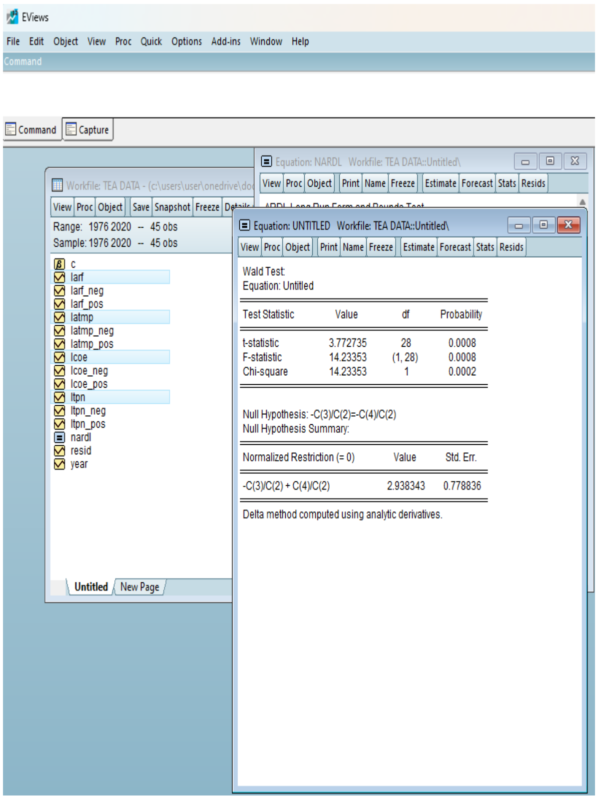



Equation: UNTITLED Workfile: TEA DATA::Untitled\
Dependent Variable: D(LTPN)
Method: Variable Selection
Date: 10/10/24 Time: 12:46
Sample (adjusted): 1979 2020
Included observations: 42 after adjustments
Number of always included regressors: 8
Number of search regressors: 7
Selection method: Uni-directional
Stopping criterion: p-value = 0.1
Variable Coefficient Std. Error t-Statistic Prob.*
C 15.60077 1.741417 8.958667 0.0000
LTPN(-1) -1.469590 0.164343 -8.942222 0.0000
LATMP_POS(-1) 3.165173 1.127916 2.806214 0.0090
LATMP_NEG(-1) -1.152987 0.918439 -1.255378 0.2197
LCOE_POS -0.704229 0.147064 -4.788592 0.0000
LCOE_NEG(-1) 1.049830 0.227394 4.616797 0.0001
LARF_POS(-1) 0.046209 0.104337 0.442880 0.6613
LARF_NEG(-1) -0.753837 0.147723 -5.103058 0.0000
D(LARF_NEG) -0.529658 0.120016 -4.413234 0.0001
D(LCOE_NEG) 1.322115 0.347777 3.801621 0.0007
D(LCOE_NEG(-1)) -1.168844 0.406744 -2.873661 0.0077
D(LATMP_NEG) 4.539732 1.380363 3.288796 0.0027
D(LATMP_POS) -3.008578 1.064676 -2.825817 0.0086
D(LARF_POS(-1)) -0.176004 0.101895 -1.727317 0.0951
R-squared 0.773889 Mean dependent var 0.020321
Adjusted R-squared 0.668909 S.D. dependent var 0.082631
S.E. of regression 0.047546 Akaike info criterion -2.993029
Sum squared resid 0.063298 Schwarz criterion -2.413806
Log likelihood 76.85361 Hannan-Quinn criter. -2.780721
F-statistic 7.371769 Durbin-Watson stat 2.133975
Prob(F-statistic) 0.000005
Selection Summary



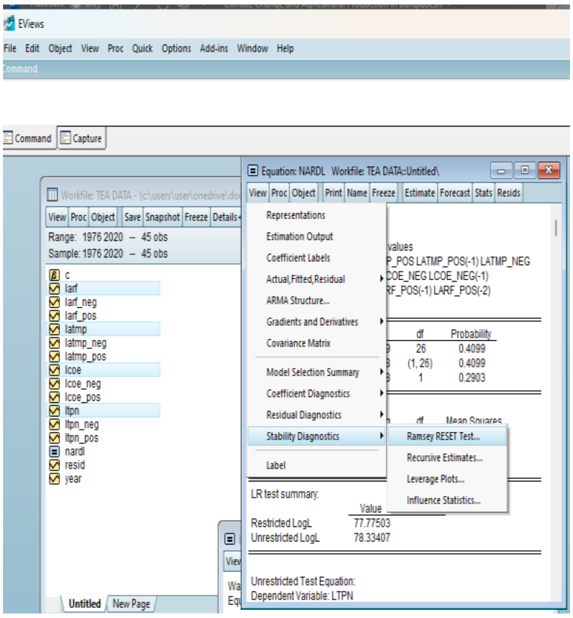

EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Command Capture
Equation: UNTITLED Workfile: TEA DATA::Untitled\
Coefficient Diagnostics
Scaled Coefficients
Confidence Intervals...
Confidence Ellipse...
Variance Inflation Factors
Wald Test- Coefficient Restrictions...
Omitted Variables Test - Likelihood Ratio...
Redundant Variables Test - Likelihood Ratio...
Selection Summary

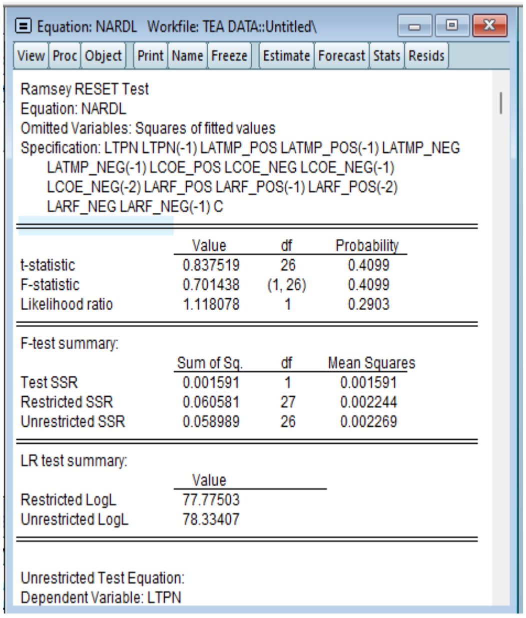

EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Command Capture
Equation: UNTITLED Workfile: TEA DATA::Untitled\
Wald Test
Equation: Untitled
Test Statistic Value df Probability
t-statistic 3.772735 28 0.0008
F-statistic 14.23353 (1, 28) 0.0008
Chi-square 14.23353 1 0.0002
Wald Test
Coefficient restrictions separated by commas
-c(3)/c(2)=-c(4)/c(2)
Examples
C(1)=0, C(3)=2*C(4)
OK
Cancel


EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Command Capture
Equation: UNTITLED Workfile: TEA DATA::Untitled\
Wald Test:
Equation: Untitled
Test Statistic Value df Probability
t-statistic 3.772735 28 0.0008
F-statistic 14.23353 (1, 28) 0.0008
Chi-square 14.23353 1 0.0002
Null Hypothesis: -C(3)/C(2)=-C(4)/C(2)
Null Hypothesis Summary:
Normalized Restriction (= 0) Value Std. Err.
-C(3)/C(2) + C(4)/C(2) 2.938343 0.778836
Delta method computed using analytic derivatives.

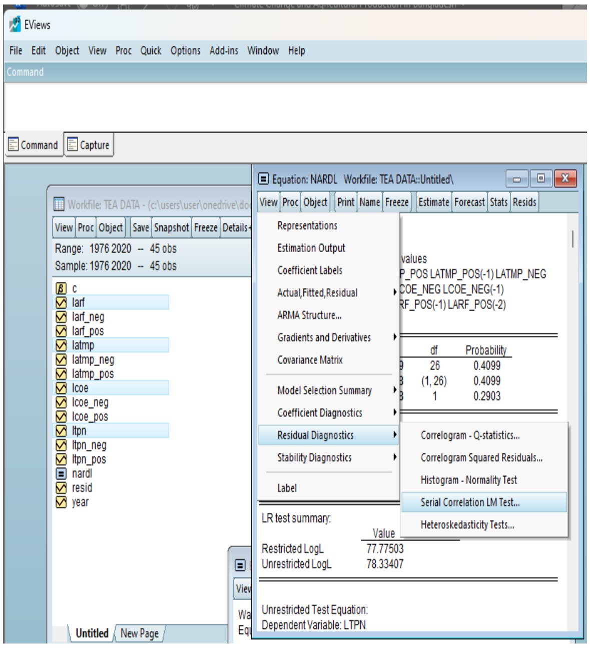
Diagnostic tests

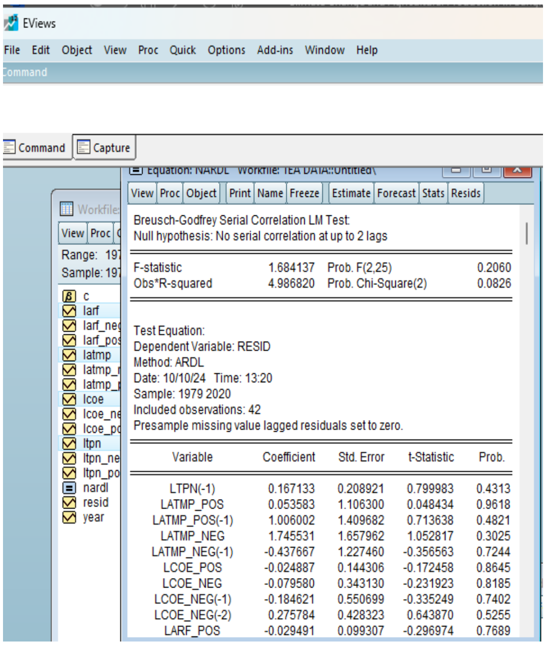
Ramsey RESET Test



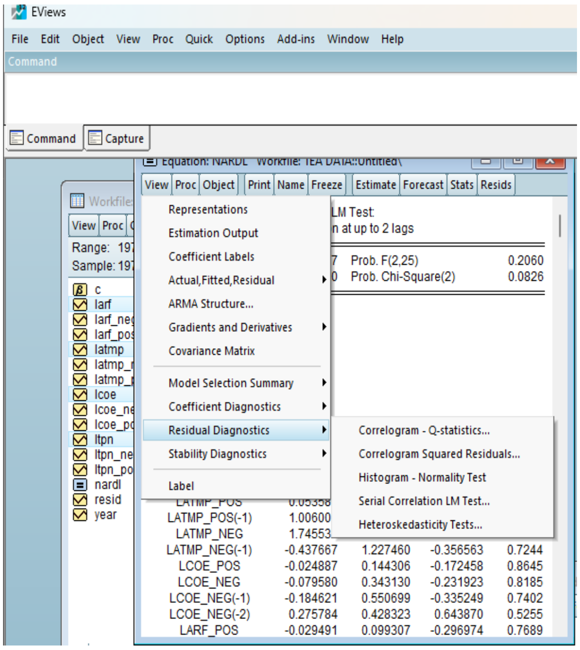


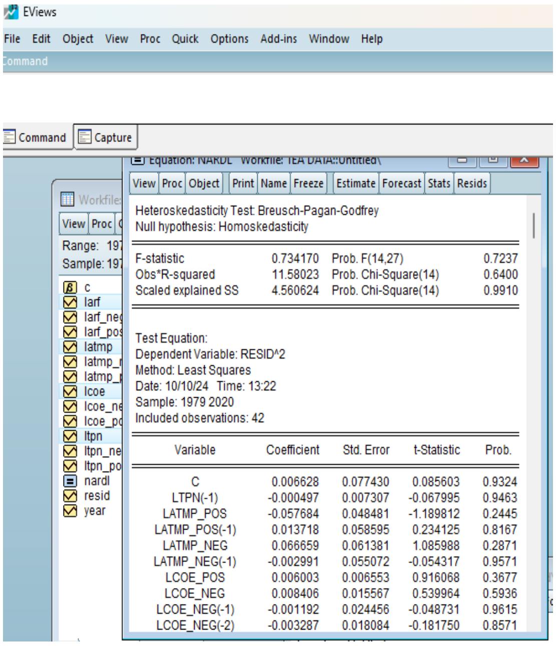
LM for serial correlation



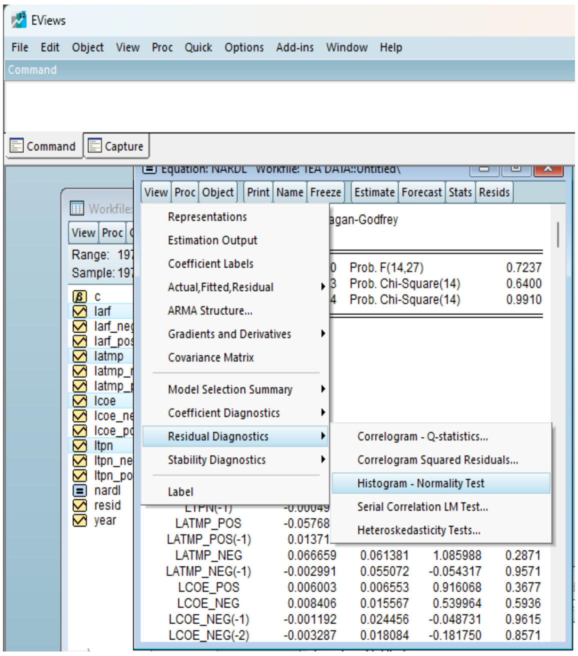


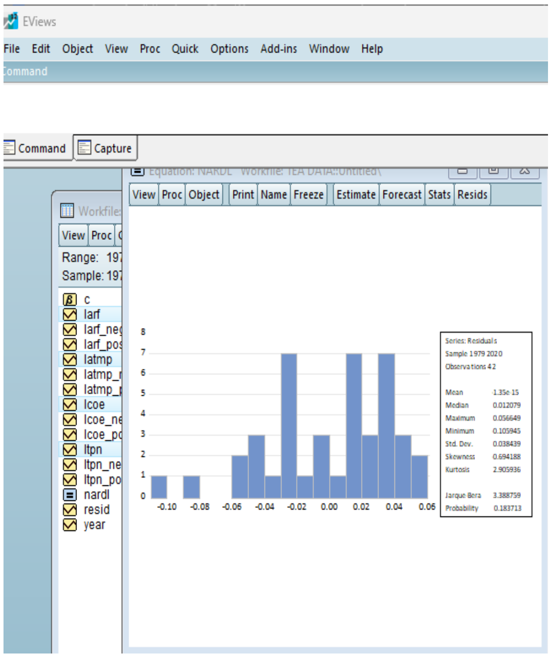
BPG for heteroscedasticity



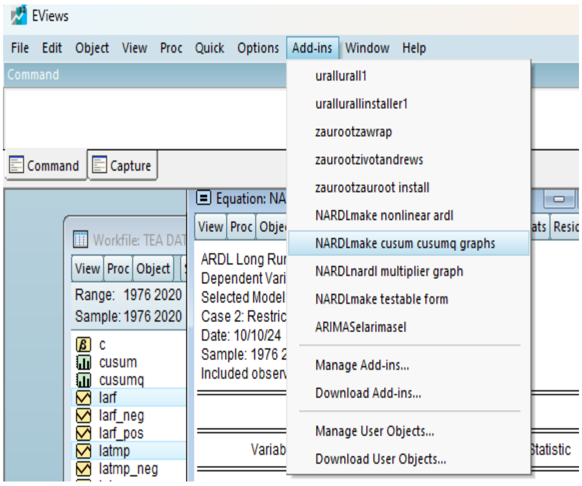


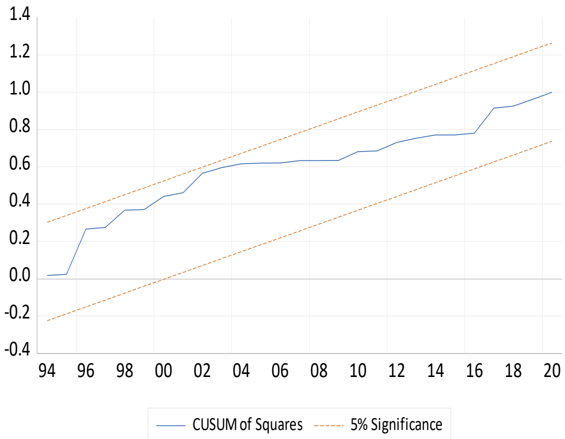
Jarque-Bera test of normality

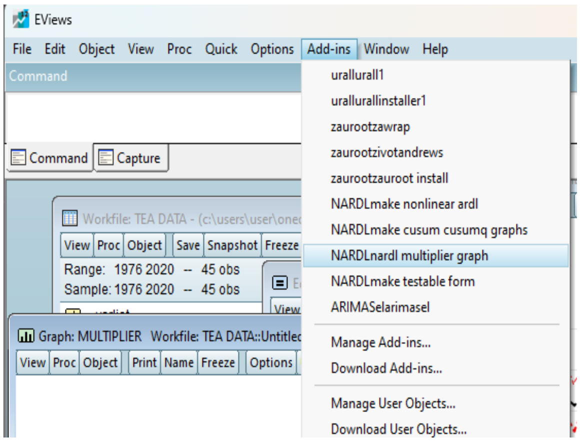


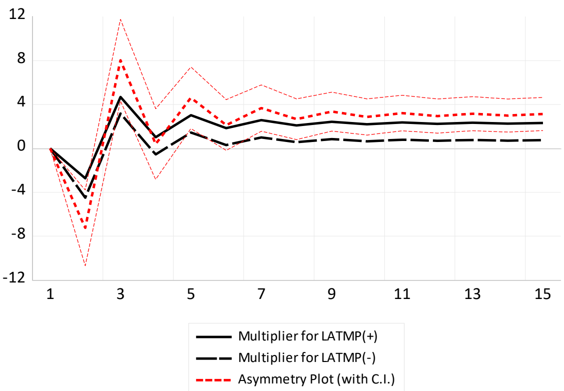


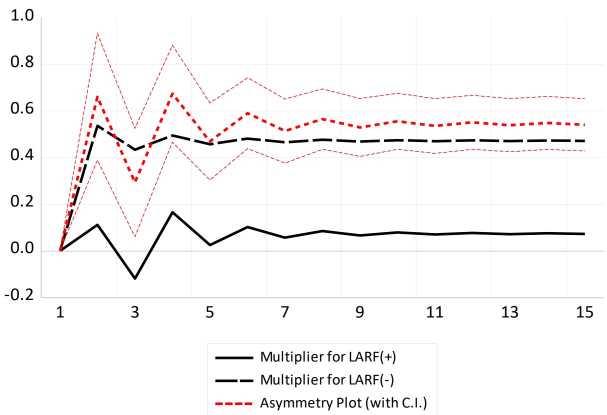
CUSUM and CUSUM square

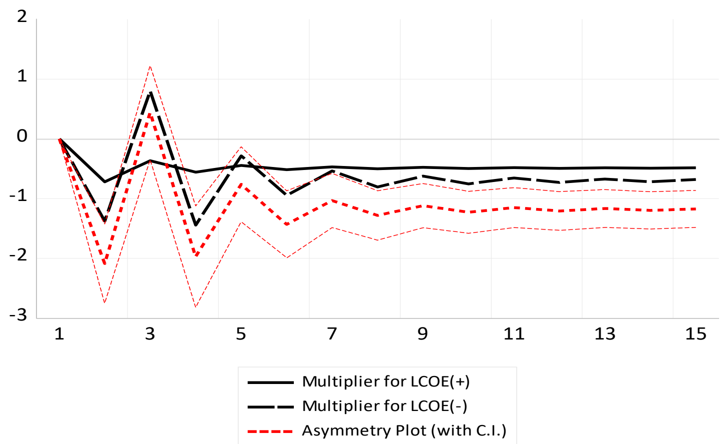




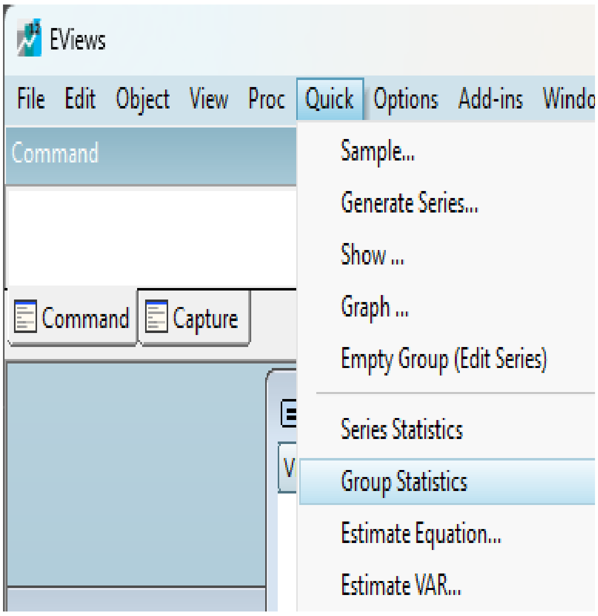


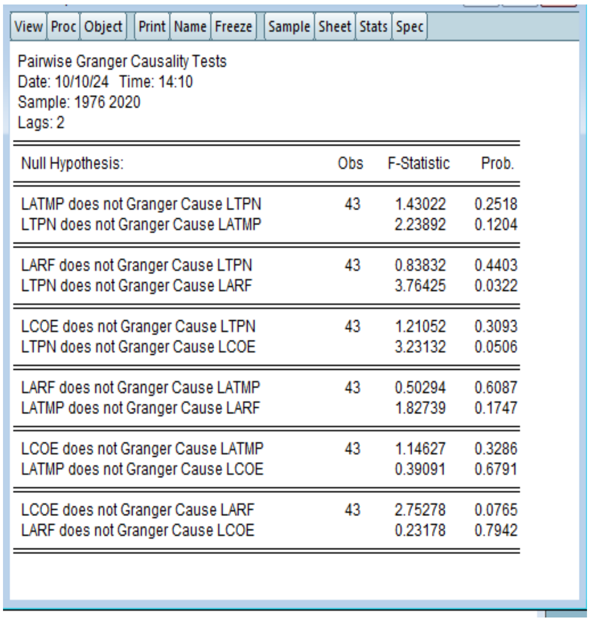






Pairwise Granger causality test




View Proc Object Print Name Freeze Sample Sheet Stats Spec

Pairwise Granger Causality Tests
Date: 10/10/24 Time: 14:10
Sample: 1976 2020
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
LATMP does not Granger Cause LTPN	43	1.43022	0.2518
LTPN does not Granger Cause LATMP		2.23892	0.1204
LARF does not Granger Cause LTPN	43	0.83832	0.4403
LTPN does not Granger Cause LARF		3.76425	0.0322
LCOE does not Granger Cause LTPN	43	1.21052	0.3093
LTPN does not Granger Cause LCOE		3.23132	0.0506
LARF does not Granger Cause LATMP	43	0.50294	0.6087
LATMP does not Granger Cause LARF		1.82739	0.1747
LCOE does not Granger Cause LATMP	43	1.14627	0.3286
LATMP does not Granger Cause LCOE		0.39091	0.6791
LCOE does not Granger Cause LARF	43	2.75278	0.0765
LARF does not Granger Cause LCOE		0.23178	0.7942

……………………..End……………….

Estimation of Supply and Demand Elasticities of the Agricultural Farm

Dr. Md. Abdus Salam
Deputy Program Director
APCU-BARC, PARTNER
Bangladesh Agricultural Research Council, Farmgate, Dhaka
Email address: asalam_36@yahoo.com
Mobile no.: 01716309251

Introduction

A basic assumption of most economic analysis of the firm behavior is that a firm act so as to maximize its profits, the difference of the revenue and the cost. This leads to the fundamental condition (Production Law).

A profit function approach is used to examine impacts of prices and fixed factors on farmers' resource allocation decisions. This is because profit function has a duality relationship with the underlying production function. An advantage of a profit function model is that it is specified as a function of prices and fixed factors which are exogenous in nature and, therefore, are free from possible endogeneity problem associated with a production function model (Rahman et al., 2012). The basic assumption is that farm management decisions can be described as static profit maximization problem.

Specifically, the farm household is assumed to maximize 'restricted' profits from growing specific crops, defined as the gross value of output fewer variable costs, subject to a given technology and given fixed factor endowments (Rahman and Parkinson,2007).

Choose the level of output such that marginal revenue = marginal cost. A firm must also face the decisions on how much of a specific input to use/hire. The second fundamental condition of profit maximization is the condition of equal long-run profits.

If the firm has a single output and two input, one input is quasi-fixed inputs (factors) since it is assumed as short run profit definition. Another one is variable input.

Quasi-fixed inputs (factors) are inputs that are held fixed or constant at the observed level for some period but eventually adjust to an optimum over a longer time period. These inputs are held fixed because of difficulties in making ready adjustments over comparatively short time periods (FAO). The firm wants to take new labours, it can face a higher wage, then it is a quasi-fixed cost but it seems to me that it should come under variable cost as the input cost changes with the quantity of output.

The profit function:

$\pi = py - w_i x_i - rz \;\; i = 1, 2$

where p is the price of the output and wi is the price of the ith input. If the firm faces output price p and input prices w_i, we can calculate the maximum profit that can be obtained by the firm by solving following optimization problem:

$\pi = py - w_i x_i - rz$ s.t. $y = f(x_i, z)$

where $y = f(x_i)$ is the *production function* of the firm.

This restricted maximization can be transformed into an unrestricted optimization by replacing y by the production function:

$$\pi = pf(x_i) - w_i x_i - rz$$

Then the first-order conditions for this special case are (*interior solutions only*):

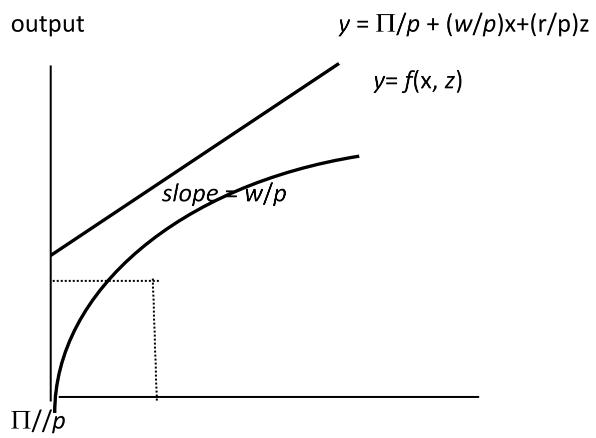
$\pi = pf(x_i) - w_i x_i - rz \quad i = 1, 2$

$p\frac{\partial\pi}{\partial x_i} = w_i$, i=1,2 or $p\nabla f(x_i) = w_i$

$p\frac{\partial\pi}{\partial z} = r$, i=1,2 or $p\nabla f\left(x_{i,}z\right) = r$

That is, the value of the marginal product of each factor must be equal to the factor's price. (Do we see that this is a special case of MR = MC?)

The diagram below illustrates the above FOC for single input case.



- The second order condition (sufficient) is as usual: the Hessian matrix of f is **negative semidefinite** at the optimal point.

Let A be an n × n symmetric matrix. Then: A is negative semidefinite if and only if all the kth order principal minors of A are ≤ 0 if k is odd and ≥ 0 if k is even.

<u>Properties of Profit Functions</u>: The above defined profit function π(**p**) is

1. non-decreasing in output prices, non-increasing in input prices;
2. homogeneous of degree 1 in prices (output and input);
3. convex in **p**;
4. continuous in **p**.

- Properties of the profit function have several uses. In particular, these properties offer some observable implications of profit-maximizing behavior:
    - Whenever some property is not true, we can claim that the firm is *not* a profit-maximizer.

Net Supply Functions and Hotelling's Lemma

Net Supply Functions - Input Demand & Output Supply Functions

The solution of the profit maximization problem: $\pi = \max_{x_i} pf(x_i) - w_i x_i$ is denoted by $y = y(p, w_i)$ which is commonly called net supply function of the firm. Clearly,

$$\pi(p, w_i) = py(p, w, r) - w_i x_i(p, w_i, r)$$

The following equation can be developed using profit maximization solution

$$y = y(p, w_i, r)$$

$$x_i = x_i(p, w_i, r)$$

$$z = z(p, w_i, r)$$

In particular,

- if x_i is an input, then the function $x_i(p, w_i)$ is called the input demand function, also known as factor demand function.
- similarly, if y is an output, then the corresponding function $y(p, w_i)$ is called the output supply function, or simply supply function.

FAO. Managing Fishing Capacity: Selected Papers on Underlying Concepts and Issues

Translog profit function Anderson, et al. [1996] point out three functional forms that seem to dominate in empirical production economics literature. Those forms are the translog, the normalized quadratic and the generalized Leontief functions. They concede that economic theory is not sufficient to determine the suitable functional form, although it does aid in identifying relevant variables and homogeneity restrictions. The preferred functional form is both data and method specific, thus making testing of alternative forms imperative to the selection process [Anderson et al., 1996; Ornelas, et al., 1993]. Correct specification of a functional form is important in so far as it impacts on predicted responses of modelled policy interventions [Anderson et al., 1996].

Suppose the farmers manages their production of output y by input x_i ($i = 1, \ldots ..2$) and quasi-fixed input z under the price p for output, price $w_i$for inputs and r for the price of quasi-fixed input. His/her profit π is described as the translog profit form.

The translog specification is a second-degree flexible function in prices and fixed inputs, with variable elasticities of substitution and is considered as a second order approximation of any functional form. Algebraically, the translog profit function is specified as follows [Christensen, et al., 1973; Capalbo et al., 1988]:

$$\begin{aligned}log\pi = \ & \alpha_0 + \ \alpha_p logp + \sum_i \alpha_{x_i} logw_i + \alpha_z logz + \frac{1}{2}\sum_i \beta_{pp}(logp)^2 \\ & + \frac{1}{2}\sum_i \beta_{pw_i} logp * logw_i + \frac{1}{2}\beta_{pz} logp * logz + \frac{1}{2}\sum_i \beta_{w_i w_j}\,(logw_i)^2 \\ & + \ \frac{1}{2}\sum_i \beta_{w_i p} logw_i * logp + \frac{1}{2}\sum_i \beta_{w_i z} logw_i * logz + \ \frac{1}{2}\beta_{zz}(lowz)^2 \\ & + \frac{1}{2}\beta_{zp} logz * logp + \frac{1}{2}\sum_i \beta_{zw_i} logz * log\, w_i\end{aligned}$$

In mathematics, the symmetry of second derivatives (also called the equality of mixed partials) refers to the possibility of interchanging the order of taking partial derivatives of a function.

According to Young theorem, symbolically the symmetry may be expressed as:

$\frac{\partial}{\partial w_i}\left(\frac{\partial \pi}{\partial p}\right) = \frac{\partial}{\partial p}\left(\frac{\partial \pi}{\partial w_i}\right)$ or $\frac{\partial^2 \pi}{\partial p \partial w_i} = \frac{\partial^2 \pi}{\partial w_i \partial p}$

This implies the main equation as

$$\begin{aligned}log\pi = \ & \alpha_0 + \ \alpha_p logp + \sum_i \alpha_{x_i} logw_i + \alpha_z logz + \frac{1}{2}\sum_i \beta_{pp}(logp)^2 \\ & + \frac{1}{2}\sum_i \beta_{pw_i} logp * logw_i + \frac{1}{2}\beta_{pz} logp * logz + \frac{1}{2}\sum_i \beta_{w_i w_j}\,(logw_i)^2 \\ & + \frac{1}{2}\sum_i \beta_{w_i z} logw_i * logz + \ \frac{1}{2}\beta_{zz}(lowz)^2\end{aligned}$$

Taylor expansion can be described as

$$\begin{aligned}f(x) = \ & f(x^*) + \sum_{j=1}^{n} \frac{\partial f(x^*)}{\partial x^*}\left(x_j - x_j^0\right) + \frac{1}{2!}\sum_{j_1=1}^{n} \sum\nolimits_{j_2=0}^{n} \frac{\partial^2 f(x^*)}{\partial x_{j_1} \partial x_{j_2}}\left(x_{j_1} - x_{j_1}^*\right)\left(x_{j_2} - x_{j_2}^*\right) \\ & + R_2\end{aligned}$$

Where R_2 is reminder

Hotelling's Lemma

If you know the profit function, then according to the following well-known lemma, Hotelling's Lemma, it is easy to find the net supply function: just differentiate the profit function.

Hotelling's Lemma. Let $y(p, w_i, r)$ be the firm's net supply function for output y. Then,

$\pi = \max_{p,w_i} pf(x_i, z) - w_i x_i - rz$

$$\frac{\partial \pi}{\partial p} = \ f(x_i, z) = y$$

$$\frac{\partial \pi(p, w_i, z)}{\partial w_i} = -x_i$$

$$\frac{\partial \pi(p, w_i, z)}{\partial w_i} = -z$$

assuming that the derivative exists and that $p > 0$.

Share equation

$$\frac{\partial log\pi}{\partial logp} = \frac{\partial \pi}{\partial p} * \frac{p}{\pi} = \frac{py}{\pi} = S_p$$

$$\frac{\partial log\pi}{\partial logw_i} = \frac{\partial \pi}{\partial w_i} * \frac{w_i}{\pi} = -\frac{w_i x_i}{\pi} = S_{w_i}$$

$$S_p = \frac{\partial log\pi}{\partial logp} = \alpha_p + \beta_{pp} logp + \sum_i \beta_{pw_i} logw_i + \beta_{pz} logz$$

$$S_{w_i} = \frac{\partial log\pi}{\partial logw_i} = \alpha_{w_i} + \sum_i \beta_{w_i w_j} logw_i + \sum_i \beta_{w_i p} logp + \sum_i \beta_{w_i z} logz$$

Since the input and output shares form a singular system of equations (by definition $S_y - \sum_i S_{w_i}$ = 1), one of the share equations, the output share, is dropped and the profit function and variable input share equations are estimated jointly using SURE procedure. The one of the share equations was dropped from the system estimations to avoid singularity of the covariance matrix when the determinant of a matrix is zero. The joint estimation of the profit function together with factor demand equations ensures consistent parameter estimates (Sidhu and Baanante, 1981).

"Profit shares" are never between zero and one but they sum up to one, as do "real" shares

$$S_y + \sum_i S_{x_i} = \frac{py}{\pi} - \frac{w_i x_i}{\pi} = \frac{py - \sum_i w_i x_i}{\pi} = \frac{\pi}{\pi} = 1$$

From the Homogeneity of degree one in price, these constraint

$$\alpha_p + \sum_i \alpha_{w_i} + \alpha_r = 1$$

$$\beta_{pp} + \sum_i \beta_{pw_i} = 0$$

$$\beta_{w_i w_j} + \sum_i \beta_{w_i p} = 0$$

$$\beta_{w_i p} - \beta_{pw_i} = 0$$

Derived output supply and input demand elasticities

Based on the derived output supply function (4.37) and the derived input demand functions (4.38), we can derive the output supply elasticities and the (unconditional) input demand elasticities:

$$\frac{\partial S_p}{\partial logp} = \frac{\partial S_y}{\partial p}.p$$

$$\frac{\partial S_p}{\partial p} = \frac{\partial(\frac{py}{\pi})}{\partial p} = \frac{y}{\pi} + \frac{p.\,{}^{\partial y}/_{\partial p}}{\pi} - \frac{py}{\pi^2}\frac{\partial \pi}{\partial p}$$

Therefore,

$$\frac{\partial S_p}{\partial logp} = \left(\frac{y}{\pi} + \frac{p.\,{}^{\partial y}/_{\partial p}}{\pi} - \frac{py}{\pi^2}\frac{\partial \pi}{\partial p}\right).p = \frac{py}{\pi} + \frac{py}{\pi}\,.\frac{{}^{\partial y}/_{y}}{{}^{\partial p}/_{p}} - (\frac{py}{\pi})^2$$

Hence in the translog form

$$\frac{\partial S_p}{\partial logp} = \beta_{pp}$$

That is

$$\beta_{pp} = \frac{py}{\pi} + \frac{py}{\pi}\,.\frac{{}^{\partial y}/_{y}}{{}^{\partial p}/_{p}} - (\frac{py}{\pi})^2$$

$$\beta_{pp} = S_p + S_p\,.\varepsilon_{yp} - {S_p}^2$$

$$\frac{\beta_{pp}}{S_p} = 1 + \varepsilon_{yp} - S_p$$

$$\varepsilon_{yp} = S_p + \frac{\beta_{pp}}{S_p} - 1$$

As well

$$\frac{\partial S_{w_i}}{\partial p} = \frac{\partial(-\frac{w_i x_i}{\pi})}{\partial p} = -\left[\frac{w_i\,{}^{\partial x_i}/_{\partial p}}{\pi} - \frac{w_i x_i}{\pi^2}\frac{\partial \pi}{\partial p}\right]$$

$$= -\left[\frac{w_i x_i}{\pi}\frac{\partial x_i}{\partial p} - \frac{w_i x_i}{\pi^2}\frac{\partial \pi}{\partial p}\right]$$

$$\frac{\partial S_{w_i}}{\partial logp} = \frac{\partial S_{w_i}}{\partial p}.p$$

$$\frac{\partial S_{w_i}}{\partial logp} = \left[\frac{w_i x_i}{\pi}\frac{\partial x_i}{\partial p} - \frac{w_i x_i}{\pi^2}\frac{\partial \pi}{\partial p}\right].p$$

$$\frac{\partial S_{w_i}}{\partial logp} = \left[\frac{w_i x_i}{\pi} \frac{\partial x_i / x_i}{\partial p / p} - \frac{w_i x_i}{\pi} \frac{py}{\pi}\right]$$

Hence in the translog form

$$\frac{\partial S_{w_i}}{\partial logp} = \beta_{w_i p}$$

That is

$$\beta_{w_i p} = \left[\frac{w_i x_i}{\pi} \frac{\partial x_i / x_i}{\partial p / p} - \frac{w_i x_i}{\pi} \frac{py}{\pi}\right]$$

$$\beta_{w_i p} = \left[\frac{w_i x_i}{\pi} \frac{\partial x_i / x_i}{\partial p / p} - \frac{w_i x_i}{\pi} \frac{py}{\pi}\right]$$

$$\beta_{w_i p} = S_{w_i} . \varepsilon_{w_i p} - S_{w_i} S_p$$

$$\varepsilon_{w_i p} = \frac{\beta_{w_i p}}{S_{w_i}} + S_p$$

Therefore

$$\varepsilon_{x_i w_i} = S_{w_i} + \frac{\beta_{w_i w_i}}{S_{w_i}} - \delta_{ij}$$

Where δ_{ij} is the Kronecker delta as

$$\delta_{ij} = \begin{cases} 1 \; if \;\; i = j \\ 0 \;\; if \; i \neq j \end{cases}$$

$$\varepsilon_{x_i w_i} = \begin{pmatrix} S_{y_1} & \cdots & S_{y_1} \\ \vdots & \ddots & \vdots \\ S_{x_n} & \cdots & S_{x_n} \end{pmatrix} + \begin{pmatrix} 1/S_{y_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/S_{x_n} \end{pmatrix} - \delta_{ij}$$

$$\delta = I$$

Testing the properties of the profit function

Under the assumptions of profit maximizing behavior with a continuous and a twice-differentiable profit function, the parameters of the estimated equations must satisfy symmetry, convexity, monotonicity and homogeneity conditions.

Monotonicity:

This property requires that the profit function strictly increases in output prices and strictly decreases in input prices [Chambers, 1988]. This property is tested through evaluation of the first derivatives of the profit function with respect to input and output prices.

In the translog case, this implies evaluation of the profit shares. For inputs, the first derivatives of the profit function with respect to the input price should be non-positive.

The first derivatives of the profit function with respect to the output prices should be non-negative. Since the functions approximate the true profit function and the first derivatives are expressions in the levels of the variables, the evaluation is done at the point of approximation. In the normalized quadratic case, this implies setting the values of the variables to zero and for the translog function the values are set to one.

For normalized quadratic case

$$\alpha_p + \sum_i \alpha_{w_i} + \alpha_r = 0$$

For translog case

$$\alpha_p + \sum_i \alpha_{w_i} + \alpha_r = 1$$

For the system of input demand and output supply functions to be compatible with profit maximization, monotonicity and convexity of the underlying profit function, as well as homogeneity and symmetry must hold. The unrestricted system could be estimated and then the theoretical constraints could be formally tested, both locally and globally [Capalbo et al., 1988]. The latter effectively provides a test for profit maximizing behaviour [Lopez, 1980].

Convexity

The necessary condition for convexity is that the Hessian matrix of second order derivatives of the profit function with respect to all prices be positive semi-definite. This implies that all the principal minors must have non-negative determinants {Capalbo et al., 1988]. This follows from the fact that $\frac{\partial^2\pi}{\partial p \partial p} > 0$ and $\frac{\partial^2\pi}{\partial w_i \partial w_i} > 0$) making the profit function convex in input and output prices (i.e. output supply is upward sloping and input demand is downward sloping). Algebraically, the Hessian matrix is represented as follows:

From the equation

$$\begin{bmatrix} \frac{\partial^2\pi}{\partial p_1 \partial p_1} & \cdots & \frac{\partial^2\pi}{\partial p_i \partial p_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\pi}{\partial p_1 \partial w_i} & \cdots & \frac{\partial^2\pi}{\partial w_i \partial w_n} \end{bmatrix}$$

Homogeneity

The Wald-test was used to test for the homogeneity restrictions. The translog profit function is homogeneous in prices. but not in fixeded factors. For the translog profit function to be homogeneous, the symmetry condition ($\beta_{pw_i} = \beta_{w_ip}$, $\beta_{pz} = \beta_{zp}$ and $\beta_{w_iz} = \beta_{zw_i}$) the additivity restriction ($\alpha_p + \sum_i \alpha_{w_i} + \alpha_r = 1$) well as the condition that the sum of the coefficients of the squared and interaction terms are zero ($\beta_{pp} = \sum_i \beta_{pw_i} = \beta_{w_iw_j} = \sum_i \beta_{w_ip} = \beta_{w_ip} = \beta_{pw_i} = 0$) must hold. However, homogeneity in prices can also be imposed by non-normalizing the translog profit function.

Symmetry

Symmetry is imposed due to the restricted sample size. Without the symmetry condition, there are not sufficient degrees of freedom in order to estimate all the parameters of the specified equations. The methodology described is used to estimate different combinations of single equation and system specification. The symmetry property was tested by imposing cross-equation restrictions of equality on the corresponding parameters between the profit function and the four factor demand equations.

Symmetry of the parameters

$\beta_{pw_i} = \beta_{w_ip}$

$\beta_{pz} = \beta_{zp}$

$\beta_{w_iz} = \beta_{zw_i}$

Estimation method of a system of equations

The econometric estimation of a system of equations can be done with various techniques1 [Johnston, 1984; Zellner, 1987; Pindyck and Rubinfield, 1991; Johnston and DiNardo, 1997; Greene, 1997]. For this study, Seemingly Unrelated Regression (SUR) and Ordinary Least Squares (OLS) estimation techniques are considered and compared to determine the appropriate estimation technique, given the data set. SUR estimation (also known as the multivariate regression or Zellner' s method) [Zellner, 1962] accounts for both heteroskedasticity and contemporaneous cross-equation error correlation. This technique is appropriate when all the right-hand side variables are assumed exogenous, and when some common factors, which are not explicitly modelled, influence the disturbances across equations [Zellner, 1962; Johnston and DiNardo, 1997]. Using the Iterative-SUR makes the system indifferent to the choice of the dropped share equation. In addition, the cross-equation symmetry restrictions and possible contemporaneous correlation between the errors of the various share equations, justify the choice of this method [Higgins, 1986; Pindyck and Rubinfeldt, 1991; Kotsoyannis, 1981; Johnston and DiNardo, 1997].

While the systems approach allows for cross-equation restrictions and takes account of cross-equation error correlation, it does come at a cost. Misspecification of an equation within the system may contaminate estimates of the other parameters. When employing single equation estimation, only the parameters of the mis-specified equation are affected. Thus, OLS provides an intuitive test for the correct specification of the different equations in the system approach. If the system estimation yields unsatisfactory results, the single equation OLS results may indicate which equation(s) causes the problems.

Each specified equation for both methods contains an additive error term that captures the unexplained difference between the profit maximizing levels of input and output versus the realized levels [Higgins, 1986]. The error term will inexorably capture the effect of all variables that are not explicitly specified as well as some quality differences in inputs and outputs. No quality distinctions are reported and could thus not be incorporated. In addition, the cross-sectional nature of the data leads to the use of White's Heteroskedasticity Consistent Variance Co-variance Estimator [White, 1980] to account for possible heteroskedasticity of unknown form.

Refernce:

Chambers, R.G. 1988. Applied Production Analysis. A Dual Approach. Cambridge University Press, Cambridge.

Capalbo, Susan, M. and John M. Antle (Ed.). (1988). Agricultural Productivity: measurement and explanation. Resources for the Future, Washington, D.C., USA.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau. (1971). "Conjugate duality and the transcendental logarithmic production function". Econometrica, Vol. 39:4, pp. 255 – 256.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau. (1973). "Transcendental logarithmic production frontiers". Review of economics and Statistics, Vol. 55, pp. 28 – 45.

Higgins, James. (1986). "Input demand and output supply on Irish farms – A microeconomic approach". European Review of Agricultural Economics, Vol. 13, pp. 477 – 493.

Rahman, S., Parkinson, R.J. 2007. Soil fertility and productivity relationships in rice production system, Bangladesh. Agricultural Systems. 92: 318-333.

An Introduction to Beta Regression: Key Concept and Application

Dr. Md. Shofiqul Islam
Principal Scientific Officer (PSO)
Agricultural Economics and Rural Sociology (AERS) Division
Bangladesh Agricultural Research Council, Farmgate, Dhaka
Email address: shafiqbau07@gmail.com
Mobile no.: 01704-778929

Introduction to Beta Regression

Beta regression is a type of regression used for modeling response variables that are continuous and restricted to the interval (0, 1). This is particularly useful for proportion data or rates that cannot take values outside this range, such as:

- Proportion of a population exhibiting a characteristic
- Rates of success or failure
- Probabilities

The Beta distribution is flexible and can model various types of distributions based on the values of its shape parameters.

Key Characteristics of Beta Regression

- **Range**: The dependent variable must lie between 0 and 1, exclusive.
- **Distribution**: Assumes that the response variable follows a Beta distribution.
- **Link Function**: Commonly uses the logit link, but other link functions (e.g., probit) can also be applied.

When to Use Beta Regression

Use beta regression when:

- The dependent variable is continuous and bounded (0, 1).
- The data is not normally distributed.
- There are issues with heteroscedasticity in OLS regression.

STATA software commands for Beta Regression

Setting Up Your Data

Ensure your data is in the correct format. The dependent variable should be in the interval (0, 1).

```
clear

id y x1 x2
1 .1 10 20
2 .5 15 30
3 .9 20 25
```

Loading Required Package

If you haven't installed the *betareg* package yet, do so:

ssc install *betareg*

Running Beta Regression

Use the betareg command to fit a beta regression model. Here, y is the dependent variable and x1 and x2 are independent variables.

betareg y x1 x2

Interpreting Results After running the command, STATA will output the coefficients, standard errors, z-values, and p-values. Interpret the coefficients as the change in the log-odds of the response variable per unit increase in the predictor.

Checking Model Fit Check the goodness-of-fit for your model:

estat ic

This command will display information criteria like AIC and BIC, which help in model comparison.

Predictions: You can obtain predictions from your model:

predict yhat, mu

Here, yhat will contain the predicted values based on your model.

Diagnostic Plots: To visualize the fit and check for assumptions, you can create diagnostic plots. For example, you can plot residuals:

scatter residuals yhat

Where residuals are calculated as:

gen residuals = y - yhat

Example Application: Let's consider an example of modeling the proportion of students passing an exam based on hours studied and previous grades.

Load your dataset

use "exam_data.dta", clear

Run Beta Regression

betareg pass_rate hours_studied previous_grade

Interpret the coefficients Analyze the output to understand how hours studied and previous grades affect the pass rate.

www.barc.gov.bd