

BASIC & ADVANCED ECONOMETRIC ANALYSIS USING NSSO UNIT LEVEL DATA  
[Theoretical and Applied]

(TUS 24 & CAMS 79<sup>th</sup> ROUND)

Four-Month Long Certificate Course [90- 100 Hours]

1st July to 30<sup>th</sup> October 2026

7:00 pm – 8:30 pm (Monday to Thursday)

Online (Webex)

Department of Humanities, Arts and Social Sciences,  
National Institute of Technology Calicut  
Kozhikode (Kerala)

## About the Course

NIT Calicut believes that research skills should be accessible to every student, at any level, from anywhere. Statistical and econometric analyses are essential for solving real-world challenges, including socio-economic, geographical, psychological, and environmental issues. These courses are to equip students, scholars, and professionals with both theoretical knowledge and hands-on applied skills.

### Econometric Analysis-

Balancing theory and application is essential in econometric analysis—not just integrating skills from allied fields like mathematics and statistics, but also committing to consistent practice of the applied aspects. A key challenge is that these subjects are often absent from curricula or taught in isolation, in a segmented rather than narrative fashion. Moreover, effective software use demands a solid grasp of underlying methods. Today, many prioritise skills in statistical packages amid market demands, yet over the past 20 years, nearly all such packages have incorporated built-in functions for statistical and econometric tools, simplifying precise calculations. Still, results require logical explanations, which hinge on understanding the theory and its synergy with the research question or problem.

To address these learning challenges, the theory of each method will be taught first, followed by software-based applications. This approach helps scholars grasp both the rationale and limitations simultaneously. The course content deliberately balances theory, application, and software skills, highlighting the role of theory during practical sessions. With every demonstration, results will be interpreted to underscore the need for theoretical depth before drawing applications or conclusions.

For beginners, extensive graphical methods and real-world examples will be incorporated to make concepts more intuitive and easier to imagine. The teaching methodology fosters independence by avoiding ready-made answers or resources, instead encouraging participants to invest equivalent effort in mastering every aspect.

## NSSO Unit Level Data Analysis-

The National Sample Survey Office (NSSO) of India's Ministry of Statistics and Programme Implementation conducts regular socioeconomic surveys across the country and publishes the results. The reports contain aggregate (macro) level information that is relevant to research and policy. The unit-level data from various rounds of the NSS surveys, on the other hand, is a rich source of information widely used by researchers, corporate executives, and policymakers. NSSO conducts surveys on variables related to agriculture, employment, consumption, health, education, time use, housing conditions, disabilities, migration, unorganised manufacturing, debt, and many other important segments.

Comprehensive Annual Modular Survey (CAMS): CAMS is introduced to address the emerging need for information on high-frequency socio-economic indicators that are not available from other sources, such as administrative data. CAMS will collect information required to generate SDG indicators and sub-indicators for Global Indices.

The Time Use Survey (TUS) provides a framework for measuring the time dispositions of the population across different activities. One distinguishing feature of the Time Use Survey from other household surveys is that it can capture time disposition on different aspects of human activities, be it paid, unpaid or other activities, with such details which are otherwise not possible in other surveys. In recent years, time use surveys have gained much impetus among policymakers and other data users for their usefulness in measuring various aspects of gender statistics.

Because there is no detailed manual for obtaining unit-level data from the NSSO and training in data extraction, merging, and variable construction is limited, researchers often find it difficult to use the NSSO's unit-level data in their research. As a result, only a small number of scholars are using these large national-level data sets in their research. This is where this FDP will come in handy, as it will enable researchers with the applied skills to use NSSO's unit-level data in research. The objective of the course is to provide a step-by-step process (methodology) of extracting, merging and analysing the unit-level data of NSSO surveys by using STATA software.

## Objectives:

- To develop an understanding of national-level surveys on various types of economic and social variables of national importance.
- To raise awareness among participants on the utilisation of the national level data for their respective field of research
- To increase the use of these unutilized data in research to find solutions for economic, social, geographical and environmental problems in future.
- To help researchers apply basic and advanced econometric analysis tools to unit-level data across various disciplines.

## Pedagogy:

A mix of Pedagogical tools will be used. Discussion and hands-on training with real data using statistical software. Participants will be guided with the relevant resources to facilitate the

hands-on training\*.

\*STATA participants have to arrange for themselves.

## Learning Outcomes:

- Participants will learn Basic STATA commands: gen, rename, label, egen, recode, tag, destring (with ignore and force), count, list, describe, substring, split, sort, browse.
- Participants will be introduced to official sources to download NSSO raw data and required files from ICSSR and MOSPI.
- Participants will understand the role of required documents and raw data files. Like, Read me file, Schedule, Instructions to field staff, Layout file and Reports.
- Learn how to extract the data of different levels.
- Participants will learn merging of files with different natures/combinations of files.
- Participants will understand the use of weight and cross-verification of results with the report.
- Participants will learn econometric methodology, OLS (method, assumptions and properties).
- Participants will learn multiple simple linear regression models, dummy variable models and OLS assumption diagnostics (Multicollinearity, Heteroscedasticity, Autocorrelation, and Model specification errors testing).
- Participants will learn the application of MLE (Maximum likelihood estimation on various applicable models in cross-section and panel data).
- Participants will learn, Qualitative Response Regression Models (Logit/Probit/Tobit/M-Logit/O-Logit), Count Data Models, Simultaneous-Equation Models, Dynamic Econometric Models, Panel Data Regression Models, Non-Linear Regression Models (Iteration method), Quantile Regression, Principal Component Analysis

### Target Participants:

Master students, Research Scholars, Faculty Members and Working Professionals in the area of economics, health, population studies, planning and related areas in India and abroad (If there is a provision of accepting the fee).

**A certificate of completion will be provided for 70% attendance.**

### Number of Participants:

60 (Maximum)

### Mode of Conduction:

Online (Webex)

### Registration Fee:

Research Scholars -Rs. 6500 (Rs. 5000 for NIT Calicut research scholars)

Faculty Member- Rs. 8500

## Course Details-

### Segment One-

#### Unit Level Data Analysis and Basic Econometric Analysis for decision making with STATA/R

Approximate Training Hours: 42-45 hours spanning over 7-8 weeks (6 hours per week).

Prerequisites: Basic understanding of probability, Basic Calculus, and descriptive statistics such as location and dispersion.

|                        | NSSO UNIT LEVEL DATA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | BASIC ECONOMETRIC ANALYSIS                                                                                                                                                                                                                                                                                                                                                                             |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Week 1-2<br>(12 hrs)   | <p>Extraction-</p> <p>How to apply the CONCATENATE command in the layout file.</p> <p>To make dct files in STATA using details from the layout file.</p> <p>To apply the command in the Do-file Editor to extract level-wise data in STATA to get all the files in the dta format.</p> <p>About cross-verifying the data files with the details provided in the readme file, to test the accuracy of the extracted files.</p> <p>Append function for old rounds.</p>                                                                                                                                                                                                                               | <p>Introduction to the STATA software and essential STATA commands, such as gen, rename, label, egen, recode, label, tag, destring (with ignore and force), count, list, describe, substring, split, sort, browse.</p> <p>Learning about types of data (cross-sectional and time series), the nature of the variable (nominal/ordinal/interval/ratio scale), and Econometric Methodology.</p>          |
| Week 3-4-5<br>(18 hrs) | <p>Merging and Matching</p> <ul style="list-style-type: none"> <li>To identify household and individual levels using the schedule layout, questionnaire and instructions to the field staff.</li> <li>How to generate individual IDs and household IDs using common primary keys.</li> <li>About merging one file with another for the following combinations 1:1, 1:m, m:1, and m:m.</li> <li>Household with household, household with individual, individual with individual (with and without duplicates)</li> <li>How to deal with duplicates.</li> <li>How to generate and apply weights.</li> <li>About cross-verifying the extracted data with the report, to validate the data.</li> </ul> | <ul style="list-style-type: none"> <li>OLS method with Bivariate Regression,</li> <li>Assumptions, Statistical and Mathematical properties,</li> <li>Evaluation of results, including hypothesis testing.</li> <li>Estimation of multiple regression, Interpretation of multiple regression equation, Multiple coefficients of determination, Adjusted R-squared and evaluation of results.</li> </ul> |

|                     |                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                          |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Week 6-7-8 (18 hrs) | Construction of variables <ul style="list-style-type: none"> <li>After extracting and merging the data files level-wise, one needs to calculate the variables as per the research objectives.</li> <li>To illustrate these cases in detail, we teach how to calculate important indicators of NSSO rounds (TUS 24 &amp; CAMS 79<sup>th</sup> ROUND).</li> </ul> | <ul style="list-style-type: none"> <li>OLS method-ANOVA models, Mixture of quantitative and qualitative regressors (ANCOVA), Interaction Effects, Use in seasonal Analysis and Piece-wise Linear Regression</li> <li>Multicollinearity, Heteroscedasticity,</li> <li>Autocorrelation and Model Specification (Nature, Causes, Consequences, Detection and Remedial Procedure)</li> </ul> |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## Segment two-

### Theoretical and Applied Advanced Econometric Analysis with STATA

Approximate Training Hours: 55-60 hours spanning over the period of 10 weeks (6 hours per week).

Prerequisites: Basic calculus, matrices, Probability distributions and Basic Econometrics.

Detailed course outline of the topics covered under Theoretical and Applied Advanced Econometric Analysis for Decision Making with STATA (*using OLS, MLE, and GMM methods*):

|                   | Topic                                         | Details of the topic                                                                                                                                                                                                                                                            |
|-------------------|-----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Week 1 (6 hrs)    | Methods                                       | OLS, MLE, GMM and the Iteration method                                                                                                                                                                                                                                          |
| Week 2-3 (12 hrs) | <b>Qualitative Response Regression Models</b> | Qualitative Response Regression Models: Nature of qualitative response models, Linear Probability Model (LPM), Maximum Likelihood Estimation method, The Logit Model (Odds Ratio and Margins), The Probit Model, Multinomial Logit model (Relative Risk Ratio) and Tobit model. |
| Week 3-4 (12 hrs) | <b>Count Data Models</b>                      | Poisson Regression, Negative Binomial Regression, Zero-Inflated Count Models (Zero-inflated Poisson, Zero-inflated Negative Binomial), Zero-Truncated Count Models (Zero-truncated Poisson, Zero-truncated Negative Binomial), Hurdle Models and Random-effects Count Models.   |

|                      |                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|----------------------|-------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Week 5<br>(6 hrs)    | <b>Simultaneous Equation Models</b>                                           | Simultaneous Equation Models-<br>Learning about the nature of the simultaneous equation model, the identification problem: Under Identification, Just or Exact Identification, Overidentification, Rules for Identification (Order and Rank Condition), A Test of Simultaneity (Hausman Specification Test), Tests for Exogeneity. Approaches to estimation: Recursive Models and OLS, the method of Indirect Least Squares, the method of Two-Stage Least Squares (Overidentified equation) and 3SLS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Week 5-6<br>(12 hrs) | <b>Dynamic Econometric Models</b>                                             | Dynamic Econometric Models-<br>The reasons for lags, Estimation of distributed lag models: Ad Hoc Estimation, The Koyck Approach: The Adaptive Expectations Model, Stock Adjustment or Partial Adjustment Model.<br>Estimation of Autoregressive Models, The method of Instrumental Variables, The Almon Approach to Distributed Lag Models.<br>Causality Test (Granger Causality).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Week 7<br>(6 hrs)    | <b>Panel Data Regression Models</b>                                           | Panel Data Regression Models-<br><br>Why Panel Data is used, Estimation of Panel Data Regression Models: Pooled OLS, First difference, Fixed effect (within estimator), Fixed effect (LSDV), Random effect, Random coefficient model, Dynamic Panel data models (DL, ARDL).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Week 8<br>(6 hrs)    | <b>DID Models &amp; Quantile Regression</b>                                   | DID is a quasi-experimental design that uses longitudinal data from treatment and control groups to obtain an appropriate counterfactual for estimating a causal effect. DID is typically used to estimate the effect of a specific intervention or treatment (such as a passage of law, enactment of policy, or large-scale program implementation) by comparing the changes in outcomes over time between a population that is enrolled in a program (the intervention group) and a population that is not (the control group).<br><b>Methods- Regression Adjustment((teffects), didregress, xtdidregress</b><br><br>Quantile Regression.<br>In Quantile regression models, the default is median regression, where the objective is to estimate the median of the dependent variable conditional on the independent variables. This method is similar to ordinary regression, where the objective is to estimate the conditional mean of the dependent variable. Simply put, median regression finds a line through the data that minimises the sum of the absolute residuals rather than the sum of the squares of the residuals, as in ordinary regression.<br><b>Stata functions- qreg, qreg with quantile and bsqreg(bootstrap)</b> |
| Week 9<br>(6 hrs)    | <b>Non-Linear Regression Models and Extensions of the Two-Variable Linear</b> | Non-Linear Regression Models and Extensions of the Two-Variable Linear Regression Model-<br>Scaling and units of measurement, Regression on standardised variables, functional forms of regression models, Log-Linear Model, Semi-Log models, Reciprocal Models.<br>Estimating a nonlinear regression model-                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

|                    |                                                                        |                                                                                                                                                                                                            |
|--------------------|------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                    | <b>Regression Model</b>                                                | The trial-and-Error Method, Approaches to estimating nonlinear regression models: Direct search or trial and error of derivative-free method, Direct Optimisation, and the Iterative Linearization Method. |
| Week 10<br>(6 hrs) | <b>Principal Component Analysis &amp; Linear Discriminant Analysis</b> | <b>Principal Component Analysis</b> with axioms for weightage and scale-free method.                                                                                                                       |

### Data sets for practice

1. STATA press- Example Data sets.
2. NSSO data sets for practice from extracted rounds.
3. IHDS data set 2004-05 and 2011-12
4. NSE data sets (in progress)
5. World Bank data, (country wise and indicator wise).
6. Data sets by Wooldridge, and exercise tables from Gujarati.

### References:

1. Business Statistics, 3rd Edition Noreen R. Sharpe, Georgetown University Richard D. De Veaux, Williams College Paul F. Velleman, Cornell University.
2. Basic Statistics for Business & Economics by Douglas Lind & William Marchal- 5<sup>th</sup> Edition
3. Statistical Analysis with Business and Economic Applications by Ya-Lun Chou, 1969 Edition.
4. A Handbook of Statistical Analyses using Stata, by Sophia & Brian, Third Edition
5. An Introduction to Stata Programming, By Christopher F. Baum, First Edition.
6. Introductory Econometrics: A Modern Approach by Wooldridge (with data sets) -2 Edition
7. Basic Econometrics by Damodar Gujarati – 4<sup>th</sup> Edition
8. All derivations from A. Koutsoyannis "Theory of Econometrics" -2<sup>nd</sup> Edition
9. Econometric Analysis by William Greene -5<sup>th</sup> Edition
10. Introduction to Econometrics, by G.S. Maddala-3<sup>rd</sup> Edition
11. Econometric Analysis of Cross Section and Panel Data, Jeffrey M. Wooldridge

**Course Coordinator-**

Dr. Mohamad Aslam

Assistant Professor  
Department of Humanities, Arts and Social Sciences  
NIT Calicut

**Resource Person-**

Dr. Mohamad Aslam, Assistant Professor, Department of Humanities, Arts and Social Sciences, National Institute of Technology, Calicut.

He is currently working as an Assistant Professor in the Department of Humanities, Arts and Social Sciences at NITC. Prior to that, he worked as a Postdoctoral Fellow at IISER, Mohali. His Ph.D is from Panjab University. He has completed his M.Phil. from Jawaharlal Nehru University. His research interests lie primarily in agrarian studies, the labour market, economic transformation among tribals, education, and inequality. Currently, he is working on economic changes in the Himalayan States using NSSO unit-level data. And on the challenges of tribals in occupational transformation in the North Western Himalayas. He has more than half a decade of teaching experience in theoretical and applied statistics and econometrics. He has already conducted several such workshops and FDPs with Institutes of National Importance, including national seminars and many other courses to train scholars and working professionals in applied skills.