Gokhale Institute of Politics and Economics
*Syllabus for **Post Graduate Certificate Programme on Econometric Application and
Data Science (EADS).**:*

This is an applied quantitative course using cutting-edge methods, including econometrics, statistical software applications (*STATA, R, EViews, etc*). What distinguishes these modules is the adoption of the modern learning-by-doing approach to teaching econometrics, which emphasizes the application of econometrics to real world problems. The focus is on understanding the theoretical aspects that are critical in applied work, the ability to correctly interpret empirical results and to unmask the various technical nuances for precise understanding.

Underpinning this program is a strong emphasis on quantitative skills applicable to the private and public sectors as well as a focus on faculty and student research.

Course Structure: The content of the course will be presented by means of practical training sessions. The sessions will demonstrate the concepts and methods of econometrics with practical examples of different dimensions. The course will cover wide data sets and solution of some of the problem set will be discussed in the sessions.

Course Objectives: This course is ideal for advanced graduate students, early-career academic researchers, working professionals from analytic sector and researchers in the public, private or non-profit sector. There will be a strong emphasis on the concepts and application of probability theory, random variables, distributions, sampling theory, statistical inference, correlation and regression. Statistical inference techniques such as estimation and significance testing are important in the fitting and interpretation of econometric models. Correlation and regression analysis are essential tools for measuring relationships between variables and for prediction. This course should be of value to those interested in in-depth data analysis employing modeling and suitable econometric techniques.

Outcome: By the end of the module you should be competent in applying the mathematical and statistical tools used in econometrics, and should be able to:

1. handle data
2. apply econometric methods
3. understand the conditions under which particular estimators are appropriate
4. apply appropriate estimators to the type of numerical data
5. judging theoretical methods for different types of numerical data
6. understand the basic theory of the ordinary least squares, generalized least squares, time series and panel data models
7. interpret empirical results in applied economics literature
8. write and present technical material lucidly

Texts: As stand-alone resources will be provided, there will be no need to rely on particular texts. However, there are several good texts now available which can be used as supplementary materials, including:

- *WOOLDRIDGE, J M, Introductory Econometrics – A Modern Approach* (4th ed), South-Western, 2009
- *RUUD, Paul A., An Introduction to Classical Econometric Theory*, Oxford, 2000.
- *GREENE, William H., Econometric Analysis*, Prentice Hall, 2000.
- *BALTAGI, B.H., Econometric Analysis of Panel Data* (4th Edition), John Wiley & Sons, 2005
- *ULLAH A, GILES D.E. (eds.), Handbook of Applied Economic Statistics*, Marcel Dekker, New York, 1998
- *MUKHERJEE, Chandan., Econometrics and Data Analysis for Developing Countries: A Guide to Econometric Practice*, (with Howard White and Marc Wuyts), Routledge, UK, 1998

The course consists of ten modules; starting with **Exploratory Data Analysis (EDA)**. This module is divided in five sections. The first section is on **text analysis utilities** which will involve data preparation by cleaning special characters and spaces from texts, removal of user-defined words, stemming for related words, bag of words to observe list and number of unique words allowing quantitative analysis, discriminant analysis, clustering and use of dictionaries. The next section is **model specification** comprising data classification, understanding the role of data in modeling, exploratory data analysis vs. formal statistical inference, focus on different dimensions of data –time, spatial etc. The final section of this module is **modeling average**; this involves working with real world dataset (usefulness and shortcomings of modeling mean, relative efficiency of mean and median, confidence interval and its interpretation, etc.)

The second module on **outlier detection, shape of the distribution and data transformation** consists of two sections. The first section of this module focuses on **least square principle**, consists of concept of resistance, mean-based vs. ordered-based statistics, etc. This section will be followed by a section on **data distribution** entailing outlier detection, validity of normality assumption, data transformation using cross-sectional data.

The third module is on **linear regression.** The first section is on **concepts**, comprises of basic principles and discussion on theoretical background. The second and final section of this module on **regression with graphics** entails checking the model assumption, exploratory band regression, checking of high leverage and influential points, transformation of data towards linearity, double-log transformation vs. semi-log transformation, etc.

The fourth module on **interpretation of multiple regression coefficients** consists of two sections. The first section consists of **hierarchical sequence** of simple regression, introduction of added variable

plot, etc. followed by the second section on **different macroeconomic data set,** working on money illusion in the demand function, partial vs. multiple regressions, decomposing the sum of squares, coefficient of determination, partial regression vs. partial coefficients, checking the assumptions of regression model, etc.

The fifth module is **model selection/misspecification in multiple regression,** in which the first section comprises of **omitted variable** bias, testing zero restrictions/non zero linear restrictions, etc. followed by the second section on **parameter stability**, which would use macroeconomic database, dummy variables, etc.

The sixth module is on **heteroscedasticity** with two sections. The first section is on **detecting heteroscedasticity**, involving use of test statistics**,** and transformation towards homoscedasticity. The second section is **graphical detection of heteroscedasticity,** this would use survey data to find appropriate transformation, working with group data, etc.

The seventh module is **regression on categorical variable.** This module involves handling different data sets where dependent variable is categorical, contingency tables, test of association, logit and probit modeling, comparing probit with linear probability model. Emphasis is given on microeconometrics, using sample from big data appropriate for micro-econometric exercises.

The eighth module on **time series** has two sections. The first section is on stationarity and non-stationarity of data, different types of ordering of data, structural autocorrelation vs. temporal autocorrelation, autoregressive process, multivariate time series, etc. The is followed by second section on role of moving average in modelling exercise followed with auto-regressive moving average process, cointegration, ARCH-GARCH modelling, inferring causality and reverse causality, etc.

The ninth module is **panel data analysis**. The first section comprises of **understanding panel data**, complicacy of panel data analysis, short vs. long panel, balanced vs. unbalanced panel, clustering of residual, types of regressor, fixed vs. random effects models, role of instrumental variable to control endogeneity.

The tenth module is preparation of dissertation on topic of interest. The candidate will be asked to take one optional module (Module 10) and to prepare a rigorous econometric exercise and preparation of dissertation on different topics from demography, development issues, etc. (Module 11).

Description of the course structure: (6 months programme)

Module No.	Content
Module1:	**Exploratory Data Analysis (EDA):** Section 1: **Text analysis utilities:** Data preparation by cleaning special characters and spaces from texts, removal of user-defined words, stemming for related words, discriminant analysis, clustering and use of dictionaries. Section 2: **Model specification**: Data classification, role of data, exploratory data analysis vs. formal statistical inference, different dimensions of data. Section 3: **Modeling average**: working with real world dataset (usefulness and shortcomings of modeling mean, relative efficiency of mean and median, confidence interval and its interpretation, etc.)
Module 2:	**Outlier detection, shape of the distribution and data transformation**: Section1: Least square principle and concept of resistance, mean-based vs. ordered-based statistics, etc. Section 2: **Data Distribution** :Outlier detection, validity of normality assumption, data transformation using cross-sectional data.
Module 3:	**Linear regression:** Section 1: **Concepts**: Basic principles and discussion on theoretical background Section 2: **Regression with graphics** – checking the model assumption, exploratory band regression, checking of high leverage & influential points, transformation of data towards linearity, double-log transformation vs. semi-log transformation, etc.
Module 4:	**Interpretation of multiple regression coefficients:** (Proposed resource persons (tentative): Mr. Nayak) Section 1: Hierarchical sequence of simple regression, added variable plot, etc. Section 2: Working with **different macroeconomic data set**, money illusion in the demand function, partial vs. multiple regressions, decomposing the sum of squares, coefficient of determination, partial regression vs. partial coefficients, checking the assumptions of regression model, etc.
Module 5:	**Model selection/misspecification in multiple regression:** Section1: Omitted variable bias, testing zero restrictions/non zero linear restrictions, etc. Section2: Parameter stability using macroeconomic database, use of dummy variables, etc.
Module 6:	**Heteroscedasticity:** Section 1: **Detecting heteroscedasticity**, different test statistics**,** and transformation towards homoscedasticity. Section 2: **Detecting heteroscedasticity by graphical method** using survey data, finding appropriate transformation, working with group data, etc.
Module 7:	**Regression on categorical variable:** Section1: Handling different data sets where dependent variable is categorical Section 2: Preparation, distributional and contingency tables Section 3:test of association, logit and probit modeling, comparing probit with linear probability model.

	Emphasis is given on microeconometrics, using sample from big data appropriate for micro-econometric exercises.
Module 8:	**Time series analysis:** Section 1: Stationarity and Non-Stationarity of data, different types of ordering of data, structural autocorrelation vs. temporal autocorrelation, autoregressive process, multivariate time series, etc. Section 2: Role of moving average in modelling exercise, auto-regressive moving average process, cointegration, ARCH-GARCH modelling, inferring causality and reverse causality, etc.
Module 9:	**Panel data analysis:** (Proposed resource persons (tentative): Dr. Siddhanta) Section 1: **Panel data**, complicacy of panel data analysis, short vs. long panel, balanced vs. unbalanced panel, clustering of residual, types of regressor, fixed vs. random effects models, role of instrument variable to control endogeneity. Section 2: Organization of real world dataset for panel analysis, panel identifier, time identifier, regression analysis, interpretation of result, report writing.
Module 10	**Demographic Analysis and Population Projections (Optional 1)** Section 1: Population age-sex structure: a tool to understand demographic change and evaluate data quality, Concepts and measurements related to demographic change: fertility, mortality, and migration Section 2: Direct and indirect estimation techniques for fertility and mortality data from registration systems, household surveys, and censuses
	Spatial Data Analysis (Optional 2) Section1: Spatial Autoregressive model Section 2: Spataial Moving Average model Section 3: Spatial Autoregressive Moving Average model Section 4: Spatio-temporal model Section 5: Spatio-temporal panel
	Cluster Analysis for Market Research (Optional 3) Section 1: Market segmentation: identification homogenous groups of objects (or cases, observations) called clusters - hierarchical methods and partitioning methods Section 2: Conducting a Cluster Analysis: Selection of Clustering Variables, Types of clustering variables, factor-cluster segmentation Section 3: Selection of Clustering Procedure: Hierarchical Clustering Methods
Module 11:	**Project: Demography/Development Economics**

Finally, students will be asked for assessment tests which include submission of dissertation and formal examination. Grades for the course will be computed based on the following allocation:
1. Final exam (20%) [short objective types + practical exercises + result interpretation and report writing on the basis of analysis]
2. Dissertation (80%)