

# Mahatma Gandhi College Thiruvananthapuram

## **Department of Economics**

# Certificate Course on

# DATA ANALYSIS USING GRETL

**Faculty** 

Sruthi S.

**Assistant Professor** 

Dept. of Economics

MG College, Tvpm

Course Duration 30 hours

Starting from 15.07.2021

Interested students may contact the department

From

The HOD

Department of Economics

Mahatma Gandhi College, Tvpm.

31st March 2021

Thiruvananthapuram

To

The Principal

Mahatma Gandhi College, Tvpm.

[Request for the approval of Add on course for the academic year 2021-2022]

Sir/Ma'am,

The department Id planning to offer an add on course during the academic **year 2**021-2022 under the title "EC 32: Certificate course on Data analytics using GRETL". Soft Sruthi S, Assistant Professor, Department of Economics, Mahatma Gandhi college will be the course coordinator as well as the instructor. We have planned for a 30-hr course from July 15<sup>th</sup> and is expected to complete before mid of October 2021. The targeted group is the Economics Graduate students. The details of the course are attached here with.

Thanking You

Yours' sincerely

HoD

Primar

Head of the Department PG Department of Economics Mahatma Gandhi College Thiruvananthapuram 1

2	n	1	1	<b>つ</b>	$\sim$	1	1
Z	U	Z	J	L-2	U	Z	Z

**EC 32: Data Analysis Using GRETL** 

Academic Year 2021-22 Mahatma Gandhi College Thiruvananthapuram

### **Course Synopsis**

This is a course designed for final year undergraduate B A Economics or any M.A Economics student who has done a basic course in Econometrics in their undergraduate programme. The course is designed to extend students' knowledge of basic econometric concepts and techniques learnt in Econometrics. Students will learn multiple linear regression techniques with dummy variables, more functional forms, concepts of measurement errors, multicollinearity and heteroskedasticity. These skills can be utilized in analysing data across multiple disciplines such as economics, political science, finance, business etc.

GRETL is a powerful free statistical software that allows us numerous ways to analyse, manipulate and present data. This course is designed to provide an introduction to this software, which can perform both basic as well as advanced statistical analyses. It is **not** intended to explain you the statistical and/or econometric methods. Students will be introduced to some basic features of GRETL (e.g., learning how to do data analysis using econometric tools) leading to efficient data management skills. We will include several topics on statistical analyses, performing regression analyses, as well as delving into the territory of graphics.

### **Prerequisites**

- Basic Econometrics, Basic Mathematics, Simple understanding of Hypothesis testing and Distributions.
- Laptop or desktop is recommended. [Those who want to do the course but don't have laptop or desktop can contact the instructor and get lab assistance from the college.]

### **Course Outcome**

Participants who successfully complete the course are expected to understand:

- 1. How to estimate OLS (simple linear and multiple linear regressions)
- 2.CLRM assumptions and how violations of these assumptions can affect statistical inferences;
- 3. How to interpret OLS statistics in different functional forms
- 4. Multicolliniarity and Heteroskedasticity (various test and remedies)
- 5. Instrumental variable approach to regression analysis;
- 6. Basics of the GRETL or Microsoft Excel used by economists to analyse economic data.

### **Class Timing**

Morning Slot: 8:00 to 9:30 Evening Slot: 3:30 to 5:30

The class timings are fixed dates will be intimated later after the inaugural class.

### **Instructors and Hours (30hrs)**

Instructor : Sruthi S

Preferred Contact : <a href="mailto:sruthi@mgcollegetvm.org">sruthi@mgcollegetvm.org</a>

### **Readings**

- Required Textbook: Wooldridge, J. M. (2009). Introductory Econometrics: A modern approach, 6th edition, Cengage Learning, India. [JW]
- Stock, J. H. & Watson, M. W. (2019). Introduction to Econometrics, 4th edition, Pearson [SW].
- Damodar N Gujarati and Dawn C Porter (2009): Basic Econometrics, Fifth Edition, McGraw Hill International Edition. [GJ]
- Damodar N Gujarati (2011): Econometrics by Example, First Edition, Palgrave, MacMillan.[GJX]
- AH Studenmund: Using Econometrics: A Practical Guide, Fifth Edition, Pearson Education [SM]

### **Assessment**

Evaluation will be 50% (internal lab exam in GRETL), 40% (Internal Written Exam: MCQ) and 10% for attendance and class participation.

### Assessments:

You will have several assessments throughout the course. The classes will include lecture, lab sessions and the assessment will also be done continuously based on your performance in understanding the software and also your efficiency in understanding various econometric problems using GRETL.

**Exams**: There will be no assignment for the course. There will be a final exam (40% of your grade). The final exam will be held according to university schedule.

**Academic Integrity**: Academic Honesty, Cheating, and Plagiarism as per university policy.

**Attendance Policy**: As per University policy for any other regular course degree course. **Course Contents** 

### **Syllabus**

### 30 Hrs

### Module I: Simple Linear & Multiple Regression Model

10 Hrs

Introduction to Econometrics -The concept of PRF & SRF -Significance of stochastic error term-Method of ordinary least squares- Assumptions underlying the method of least squares-Properties of estimators- Gauss Markov Theorem-Coefficient of determination, r2 -Hypothesis testing- t and F tests-P value- Practical versus statistical significance-PredictionMultiple coefficients of determination R2 and adjusted R2-Hypothesis testing-Testing the overall significance of the regression model-F test

### Module IL: Econometric Problems and their testing

10 Hrs

Multicollinearity- Nature, consequences, detection and remedial measures-Autocorrelation-Nature, consequences, detection, and remedial measures- Heteroskedasticity-Nature, consequences, detection and remedial measures.

### Module TO Dummy Variable Regression Model

10 Hrs

Dummy variable-ANOVA models-ANCOVA models-Dummy variable trap-Dummy variables and seasonal analysis-Structural analysis-Piecewise linear regression.

Important: This syllabus is intended to give the student guidance in what may be covered during the semester and will be followed as closely as possible. However, the instructor reserves the right to modify, supplement and make changes as the course needs arise.

• Multiple Regression Analysis: Further Issues

JW: Chapter 6

Multiple Regression Analysis with Qualitative Information

JW: Chapter 7 and SW: Chapter 11

Heteroskedasticity

Prepared by

SRUTHI S **Assistant Professor Department of Economics** Mahatma Gandhi College, TVPM

Head of the Department **PG Department of Economics** Mahatma Gandhi College Thiruvananthapuram

	202	1-22			
Data analysis using GRETL					
	Faculty:	SRUTHI S			
	Start Date:	15th July 2021			
	End Date:	28th September 2021			
	Hours:	32 hours			
	No of Days	27 days			
	Total no of reg students:	30			
		50			
	STUDENTS LIST				
sl no	Register number	Name			
1	55020118001	AISWARYA A S			
2	55020118002	ALFIS			
3	55020118003	ARUNIMA. P			
4	55020118004	GAYATHRI S			
5	55020118005	GOURI V NAIR			
6	55020118006	KRISHNA. K. P			
7	55020118007	LEENA GEORGE			
8	55020118008	PARVATHY A S			
9	55020118010	R.S.MALAVIKA			
10	55020118011	RESHMA MOHAN			
11	55020118012	REVATHY U S			
12	55020118014	SNIGDHA S			
13	55020118015	SURYA PRAKASH			
. 14	55020118016	VIGNESH M S			
15	55020118017	VISHNU VARDHAN M S			
16	55021118002	ADITHYA S PILLAI			
17	55021118003	AISWARYA J R			
18	55021118004	AKHILA G K			
19	55021118005	ANANA M			
20	55021118006	ARJUN J B			
21	55021118007	KARTHIKA S			
22	55021118008	KRISHNA MURALEEDHARAN			
23	55021118009	KRISHNA PRIYA V V			
24	55021118010	LAVANYA.S. A			
25	55021118011	LIJIMOL.L. A			
26	55021118012	M M SREELAKSHMI			
27	55021118013	S MEGHA NAIR			
28	55021118015	SOUFIYA BADAR			
29	55021118016	SWATHY S KUMAR			
30	55021118017	VARSHA B M			

Faculty in charge

SRUTHI S
Assistant Professor
Department of Economics
Mahatma Gandhi College, TVPM

PRIVAL G
Head of the Department
PG Department of Economics
Mahatma Gandhi College
Thiruvananthapuram

### Department of Economics Mahatma Gandhi College, Tvpm

Data analysis using GRETL (EC 32)

September 15th 2021

NAME: S MEGHA HAIR SECTION: 55021118013	V.Good
Instructions: Calculators are allowed. Total Marks: 20 points.	
1. [1 point] In the equation $y = \theta_0 + \theta_1 x + u$ , $\theta_0$ is the	
(a) dependent variable	
(b) independent variable	
(c) slope parameter (d) intercept parameter	
2. [1 point] If an independent variable in a multiple linear regre other independent variables, the model suffers from the proble	
(a) perfect collinearity	
(b) homoskedasticity	
(c) heteroskedasticity	
(d) omitted variable bias	
3. [1 point] The assumption that there are no exact linear relation multiple linear regression model fails if, where n is the	nships among the independent variables in a sample size and k is the number of parameters.
(a) n > 2	
(b) $n = k + 1$	
(c) $n > k$	
(d) n < k + 1	
4. [1 point] The Gauss-Markov theorem will not hold if	
(a) the error term has the same variance given any values of the	ne explanatory variables
(b) the error term has an expected value of non-zero given any	values of the independent variables
(c) the independent variables have no exact linear relationships	

- (d) the regression model relies on the method of random sampling for collection of data
- 5. [1 point] A model suffers from the problem of endogeneity if
  - (a) OLS estimators are not predicted precisely.
  - (b) The unobservables do not have constant variance.
  - (c)-Zero conditional mean assumption does not hold.
  - (d) When x, is uncorrelated with u
- 6. [1 point] True or False: The estimate ô is interesting because it is an estimate of the standard deviation of the unobservables affecting y. In other words, it estimates the standard deviation in y after the effect of x has been taken out and is called the standard error (s.e.).
- 7. [1 point] True or False: The term "linear" in a multiple linear regression model means that the equation is linear in the independent variables.
- 8. [1 point] True or False: The regression model,  $ceosalory = \hat{\alpha}_0 + \hat{\alpha}_1 ceoten + \hat{\alpha}_2 ceoten^2 + \hat{\alpha}_3 gender violates the assumption MLR.3.$
- 9. It point] True or False: Overspecifying a model that satisfies MLR. 1. through MLR. 4. has undesirable effects on the unbiasedness and efficiency of OLS estimators.
- 10. 11 point] True or False: MLR. 3. rules out certain relationships between explanatory variables and MLR. 4. rules out relationships between unobservables and regressors.
- 11. [2 points] True or False: Larger the sampling variance in x in an SLR model, larger the sampling variance for OLS estimators.
- 12. [2 points] CEO salary and return on equity regression model looks like the following

$$salary = 963.191 + 18.501roe$$
  
 $n = 209, R^2 = 0.0132$ 

The percentage of salary variations for CEO's salary left unexplained by roe is

- (a) 13.2%
- (b) 1.32%
- (0) 98.68%
- (d) 18.501%
- 13. Using data on 5000 college students, the following equation was estimated by OLS

$$colGPA = 1.467 - 0.0128 hsperc + 0.00192 sat$$
  
 $n = 5000 R^2 = 0.23.4$ 

where colGPA is measured on a 4-point scale, hsperc is the percentile in the high school graduating class (defined so that, for example, hsperc = 5 means the top 5% of the class), and sat is the combined math and verbal scores on the student achievement test.

(a) [2 points] The predicted college GPA when hsperc = 20 and sat = 1,050 is:

1. 2.676

Ji 3.227

7\_ iii 2.978

iv. 3.576

(b) [2 points] Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but Student A's SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students?

1.0.2688

ii. 1.0934

iil. 0.3012

iv. 0.2072

(c) 12 points! Holding hsperc fixed, what approximate difference in SAT scores leads to a predicted colgpo difference of .50, or one-half of a grade point?

i. 338

ii. 200

iii. 220

iv 260

Evaluated by

SRUTHI S
Assistant Professor
Department of Economics

Mahatma Gandhi College, TVPM

PRIVATE G

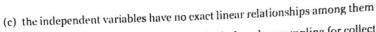
Head of the Department
PG Department of Economics
Mahatma Gandhi College
Thiruvananthapuram

### Department of Economics Mahatma Gandhi College, Tvpm

## Data analysis using GRETL (EC 32)

September 15" 2021

SECTION:	Question paper with answerkey)
Instructions: Total Marks	Calculators are allowed. : 20 points.
	he equation $y = \theta_0 + \theta_0 x + \phi_0 \theta_0$ is the
(a) dependent	variable
(b) independer	nt variable
(c) slope paras	neter
(d) intercept	parameter
2. [1 point] If an other independen	independent variable in a multiple linear regression model is an exact linear combination of transless, the model suffers from the problem of
(a) perfect co	llinearity
(b) homoskeda	sticity
(c) heteroskeda	asticty
(d) omitted var	iable bias
3. [1 point] The as multiple linear re	sumption that there are no exact linear relationships among the independent variables in a gression model fails if, where n is the sample size and k is the number of parameters.
(a) $n > 2$	
(b) $n = k + 1$	
(c) $n > k$	
(d) n < k + 1 As	is: D
. [1 point] The Ga	uss-Markov theorem will not hold if
(a) the error terr	m has the same variance given any values of the explanatory variables
(b) the error to variables	erm has an expected value of non-zero given any values of the independent



- (d) the regression model relies on the method of random sampling for collection of data
- [1 point] A model suffers from the problem of endogeneity if
  - (a) OLS estimators are not predicted precisely.
  - (b) The unobservables do not have constant variance.
  - (c) Zero conditional mean assumption does not hold.
  - (d) When x, is uncorrelated with u
  - **6. [1 point]** True or False: The estimate  $\hat{\sigma}$  is interesting because it is an estimate of the standard deviation of the unobservables affecting y. In other words, it estimates the standard deviation in y after the effect of x has been taken out and is called the standard error (s.e.). **True**
  - 7. [1 point] True or False: The term "linear" in a multiple linear regression model means that the equation is linear in the independent variables. False
  - **8.** [1 point] True or False: The regression model, ceosalary =  $\hat{\alpha}_0 + \hat{\alpha}_1 ceoten + \hat{\alpha}_2 ceoten^2 + \hat{\alpha}_3 gender violates the assumption MLR.3. False$
  - 9. [1 point] True or False: Overspecifying a model that satisfies MLR. 1. through MLR. 4. has undesirable effects on the unbiasedness and efficiency of OLS estimators. False
- 10.[1 point] True or False: MLR. 3. rules out certain relationships between explanatory variables and MLR 4. rules out relationships between unobservables and regressors. True
- 11. [2 points] True or False: Larger the sampling variance in x in an SLR model, larger the sampling variance for OLS estimators. False
- 12. [2 points] CEO salary and return on equity regression model looks like the following

$$salary = 963.191 + 18.501roe$$
  
 $n = 209, R^2 = 0.0132$ 

The percentage of salary variations for CEO's salary left unexplained by roe is

- (a) 13.2%
- (b) 1.32%
- (c) 98.68%
- (d) 18,501% Ans: C
- 13. Using data on 5000 college students, the following equation was estimated by OLS

$$colGPA = 1.467 - 0.0128 h speic + 0.00192 sat$$
  
 $n = 5000 R^2 = 0.23.4$ 

where colGPA is measured on a 4-point scale, hsperc is the percentile in the high school graduating class (defined so that, for example, hsperc = 5 means the top 5% of the class), and sot is the combined math and verbal scores on the student achievement test.

- (a) [2 points] The predicted college GPA when hsperc = 20 and sat = 1,050 is:
  - i. 2.676
  - ii. 3.227
  - iii. 2.978
  - iv. 3.576 Ans: ii
- (b) [2 points] Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but Student A's SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students?
  - i. 0.2688
  - ii. 1.0934
  - iii. 0.3012
  - iv. 0.2072 Ans: i
- (c) [2 points] Holding hsperc fixed, what approximate difference in SAT scores leads to a predicted colgpa difference of .50, or one-half of a grade point?
  - i. 338
  - ii. 200
  - iii. 220
  - iv. 260 Ans: iv

haites, A i ·115 1 1, 1

Department of Economics

**Assistant Professor** Mahatma Gandhi College, TVPM

Head of the Department PG Department of Economics Mahatma Gandhi College

Thiruvananthapuram

# DEPARTMENT OF ECONOMICS

MAHATMA GANDHI COLLEGE, THIRUVANANTHAPURAM

KERALA

# CERTIFICATE OF PARTICIPATION

In Mr. / Ms. S. Megha. Nain. (550.2.111.80.13). of 1st MA Economics, Mahatma Gandhi for completing the add-on course on "Data Analysis using GRETL" organized by the department of Economics, Mahatma Gandhi

College, Thiruvananthapuram in the academic year 2021-2022

Principal

Mahatma Gandhi College Thiruvananthapuram Principal

0

**)** 

PG Department of Económics Mahatma Gandhi College

	2021-22			
	Data anal	ysis using GRETL		
Fa	culty:	SRUTHI S	1	
Sta	irt Date:	15th July 2021		
En	d Date:	28th September 2021		
Но	urs:	32 hours		
No	of Days	27 days		
	al no of reg students			
M	ARK-SHEET			
sl no	Register number	Name	Mark	
1	55020118001	AISWARYA A S	18	
2	55020118002	ALFIS	18	
3	55020118003	ARUNIMA. P	17	
4	55020118004	GAYATHRI S	20	
5	55020118005	GOURI V NAIR	20	
5	55020118006	KRISHNA, K. P	17	
7	55020118007	LEENA GEORGE	18	
3	55020118008	PARVATHY A S		
9	55020118010	R.S.MALAVIKA	16	
.0	55020118011	RESHMA MOHAN	16	
1	55020118012	REVATHY U S	20	
	55020118014	SNIGDHA S	20	
	55020118015	SURYA PRAKASH	17	
1	55020118016	VIGNESH M S	19	
5	55020118017	VISHNU VARDHAN M S	18	
5	55021118002	ADITHYA S PILLAI	17	
	55021118003	AISWARYA J R	17	
	55021118004	AKHILA G K	17	
	55021118005	ANJANA M	17	
	55021118006	ARJUN J B	16	
	55021118007	KARTHIKA S	18	
	55021118008	KRISHNA MURALEEDHARAN	17	
3	55021118009	KRISHNA PRIYA V V	17	
	55021118010	LAVANYA.S. A	18	
	55021118011	LIJIMOL.L. A	19	
5	55021118012	M M SREELAKSHMI	17	
		S MEGHA NAIR	20	
	55021118015	SOUFIYA BADAR	17	
	55021118016	SWATHY S KUMAR	20	
		VARSHA B M THAN MS	17	

FACULTY:

Q5-45

SRUTH S
Assistant Professor
Department of Economics
Mahatma Gandhi College, TVPM

PRIYA L G

Head of the Department
PG Department of Economics
Mahatma Gandhi College
Thiruvananthapuram

### EC 32: Data Analysis Using GRETL

This is a course designed for final year undergraduate B A Economics of any M.A Economics student who has done a basic course in Econometrics in their undergraduate programme. Only PG students registered for the course. The course is designed to extend students' knowledge of basic econometric concepts and techniques learnt in Econometrics.

GRETL is a powerful free statistical software that allows us numerous ways to analyse, manipulate and present data. This course is designed to provide an introduction to this software, which can perform both basic as well as advanced statistical analyses. It is **not** intended to explain you the statistical and/or econometric methods. Students will be introduced to some basic features of GRETL (e.g., learning how to do data analysis using econometric tools) feading to efficient data management skills.

Due to the lack of availability of laptops students attended the online classes and used the nearby Akshaya centres and took help from their neighbourhood to access GRETL. The COVID-19 restrictions did affect the delivery of the course but within these limitations the course was of benefit to students especially during their project work. Final exam was conducted offline at the campus and was an MCQ type exam. At that time lab exam was also conducted and evaluated in the campus itself. Most of the students did very well as very basic econometric analysis were only asked for. The course intends to equip students to apply statistical and econometric analysis in their project works.

The students suggested for more lab hours and the department rectified that in the upcoming courses. Further the department has decided to bring in more U G students as well for the course next year.

SRUTHIS

Assistant Professor Department of Economics Mahatma Gandhi College, TVPM

PRIMALE

Head of the Department
G Department of Economics
Mahathi Gandhi Congo

Thirus manthapuram

20 (1 hm & 1 h los days)

30 200 of Brownia A.S End date Susa Kry Souping Chuminus P Vasitina B M Vistmu Stallydate: Revally U.S M. H skelakshini Languaya. S. A dujana. M A I'swamp. J. R Adithuya Enigoha Kauxbra K.P Gows V Nois R.S. Malaurka myrohor S. Megha Nain Kerishwa Parya. VV Karshara Munaludha Reshous Mehais Acium. J Kaskika. S lignesh. Surva Psakarh anvally AS certa provide Bruill ·L·A Vandhan ATTENDANCE REGISTER FOR THE MONTH OF 15" July 2021 SX S Kunnan Badan Pillar (in Inschangement 1 2 100 SHOT OU 1502118002 08 < 40 05 = < 6 5 5 7 5 3 5 7 5 < < < 5 < 7 5 5 7 অ < < < < < 80 30 < < 5 < < < < 5 < 5 < < < < < < ? 5 < 5 83 < 8 7 < < 5 30 < 5 9 < < 7 < < 0 ( 1 8 7 7 5 7 7 5 < 5 5 7 5 5 5 7 < 30 5 < ς < 5 5 5 < 5 5 ? 5 7 9 8 < 7 1 2 > 5 1 < 0 7 7 7 7 5 7 7 7 7 7 7 7 0 7 29 2 2 7 7 ? w 2 5 8 1 5 5 1 8 7 30 5 1 < 7 ? 1 7 7 7 7 7 7 4 0 ? 1 0 30 1 1 7 5 1 ? < 9 1 ? 1 7 1 7 7 5 1 7 9 1 1 9 9 1 7 < 7 Assistant Professor
Department of Economics
Mahatma Gandhi College, TVPM 0 30 7 7 7 < 2 < 7 9 5 ( 5 -4 9 1 9 29 29 1 1 ( 7 7 5 2 9 < \_ 20 7 0 -29 5 6 2 G Depar 8 - 0 ( 2 5 9 13 9 29 30 1 2 5 5 9 F andhi College 5 9 7 5 1 E 2 5 5 25 92% 100 % 786 96% 100% 1.96 100% 100% 100% 100% 1.96 168 7.001 100% 85% 100% 126 100% 188 20 76% 96% 96% 81% 100% 1001/ 92% 100% 96% 1,001 25 30/00 20 Marie 200 18 18 1 17 10 8 -0 16 P. O. 1 18 July 81 7 5 17 800 DEENY -7 Duyes Dupos Days absent Leave