Economics 204 Fall 2024
Department of Economics
Duke University

ECONOMETRICS AND DATA SCIENCE
Intuition, Theory and Applications

Instructor: Duncan Thomas
Office: Social Sciences 224
Email: dthomas@econ.duke.edu

Lectures: Monday and Wednesday 8.30-9.45am, Sanford 04

Office hours: Monday 5:00-6:30 pm in Soc Sci 327
(or by appointment in Soc Sci 224)

Class web site: http://ipl.econ.duke.edu/dthomas/ec204 (or use link from Canvas)

Teaching assistants

Graduate TAs	Angela Orozco	Angela.Orozco@duke.edu
	Tulio Salvio Sousa	TulioSalvio.Sousa@duke.edu
	Andres Santos Vargas	Andres.SantosVargas@duke.edu
Undergraduate TAs	Ian Bailey	Ian.C.Bailey@duke.edu
	Evan Greer	Evan.Greer@duke.edu
	Michael Thomas	Michael.S.Thomas@duke.edu
	Lewis Zhu	Lewis.Zhu@duke.edu

The class may only be taken for a letter grade and not on a satisfactory/unsatisfactory basis. Attendance at lectures is required. Attendance at sections is strongly recommended. Sections are led by outstanding students who have considerable expertise in applying econometric methods to data and teaching the material in this course. Sections and TA office hours times and locations are:

Sections and office hours	*Day/time*	*Place*
Section 01	Thursday 6:30-7:20 pm	Soc Sci 136
Section 02	Thursday 12:00-12:50 pm	Soc Sci 136
Office hours w/ Graduate TAs	Wednesday 5:45-6:45 pm	Soc Sci 124
	Friday 5:45-6:45 pm	Soc Sci 113
Office hours w/ Undergrad TAs	Saturday 10:00am-noon	Soc Sci 111

Please use these office hours. Note that the Economics Learning Center does not support this class.

Examinations

There will be two examinations: a midterm and a final. Both examinations are required.

Midterm: *Wednesday October 9, 8:30-9:45am* Sanford 04
The in-class mid-term will be open book. You may bring any materials you need to the exam including your laptop/tablet. You may not communicate with anyone or any AI-based tool inside or outside the room during the exam.

Final: *Monday December16, 2:00-5:00pm* Sanford 04
The final will be closed book. You may bring *one two-sided 8.5"*11" page of formulae and notes* that you think will be of use to you during the exam. That sheet must be handed in with your exam. The exam will not include questions that require the use of STATA or other statistical software. Please bring a calculator to the exam.

Course Objectives

Econ 204 is the second course in the two-course sequence in econometrics and data sciences. The overarching goal of this class is to provide you with a superior understading of the principles underlying *the theory and application* of modern econometric and data science methods so that you can effectively interpret empirical evidence to address questions in economics, finance, the social and health sciences. Mastery of the material will enable you to think critically when you evaluate the quality of evidence in support of a claim about how individuals behave, markets perform, firms make decisions, organizations operate, societies co-ordinate, etc. A recurrent theme in the class – and in the modern scientific literature – is the credibility of claims regarding identification of causal effects. To this end, this class is designed to facilitate the development of the skills needed to be an effective consumer and producer of empirical research and to apply these methods in practice. Emphasis is placed on *intuitive understanding* of underlying concepts with more rigorous arguments serving to strengthen the foundation of your knowledge. Central concepts are illustrated with *applications* using data.

Communication

It is imperative that you keep up with the class as it is very difficult catch up if you fall behind. Outstsanding support is provided by the PhD and undergraduate TAs. If you have substantive questions about the course material you want to ask me, please see me before class, after class or during my office hours. Please do not send me substantive questions about the material in this class by email. Emailed questions about the material are very hard for me to answer effectively. It has been my experience that I often need to know more about the problems you are encountering than you tell me in your email. In addition, I have no way to assess whether my e-mail response has cleared up the problem for you. I will, therefore, not attempt to answer substantive questions sent to me by email. If you cannot attend my office hours, meet before or after class, we can make an appointment at a time that is convenient for you.

Course Requirements

The class begins with a review of the linear regression model. Core statistical concepts covered in your statistics preparation will be applied to the regression model to provide a fuller understanding of the value of these tools to better understand economic phenomena and the world around us. One key goal of the class is for you to develop the skills to understand and evaluate xtensions to the linear regression model designed to address real world problems and methods designed to identify causal effects. A second key goal is to provide you with a foundation for rigorous data analysis and hypothesis testing that speaks to questions in the economics and related literatures. More broadly, the class is designed so you can develop the critical thinking skills that are necessary to successfully exploit the massive amount of data that surrounds us in this, the Age of Information.

In addition to understanding the theoretical concepts that underlie modern regression analysis, you should develop the practical skills necessary for good data analysis as well as learn how to interpret the results of your analyses. You will be required to do econometric analysis with real data from actual applications. You may use whatever computer hardware and software you like.

Instruction will be provided for using STATA which I encourage you to learn: it will serve you well in this class, in other classes, empirical research you do at Duke and beyond. We will provide instruction for STATA in the discussion sections for this class and I will use STATA in class. You may download a copy of STATA/SE for your own Mac, Windows or Linux computer. Instructions and licence information are on the class web site which has links to some of the excellent on-line resources that are publicly available.

Grading and Organization

Each week there will be two lectures. You are *required* to attend every lecture. Please put away all phones, laptops, etc. before class starts and please do not use them until the class has ended; they may not be used during class. You may take notes on a tablet but please do not use the tablet for anything else during lectures. Lectures will be supplemented by a weekly section led by an experienced TA. Sections will cover material not covered in this class that was covered in the pre-requisite class. You will be responsible for that material. Sections will also review lecture material, problem sets and provide computer instruction. Each week, one TA will be responsible for all sections and the same material will be covered in every section that week. You may attend the section that is more convenient for you. You are expected to understand and uphold the Honor Code, http://studentaffairs.duke.edu/conduct/about-us/duke-community-standard.

If you are eligible for testing accommodations, please let me know at the beginning of the semester and complete the SDAO forms. Please do not sign up to take exams in the Testing Center. I organize all exams and assure accommodations are met.

Problem sets

There will be a problem set every second week. Each is designed to help you understand important ideas in the theory and application of econometric methods and contribute to your developing a sophisticated understanding of how to interpret empirical evidence in a practical, real-world setting. To underscore the importance of taking the problem sets seriously, they will account for **30%** of the final course grade. Completed problem sets must be submitted in person at the beginning of the lecture on the due date. Late problem sets will not be accepted. If you cannot hand in your problem set at the beginning of the lecture, with my prior consent, you can hand it in earlier. Problem sets must be handed in on paper, preferably typewritten. Emailed problem sets will not be accepted.

I very strongly encourage you to work with other students in this class. The problem sets provide a useful focal point for those interactions. By collaborating with your peers, you will better understand the material. Explaining an idea, concept, method or result to a peer is one of the best ways to reinforce your own knowledge. However, to be sure that you do understand the work, *you must write up your own answersin your own words and submit your own work.* This applies to drawing on AI-based assistance. Please indicate on your problem set answers with whom you collaborated and what AI-tools you used, if any.

You may only refer to the answer keys provided for this year's problem sets which are posted on the class web page after the problem set is due. Answer keys from other years may not be used in any way or under any circumstances. A grade of zero will be assigned to all problem sets for the course for any student who violates this policy for any problem set.

For each problem set, the grades will be: 3 if you do an absolutely stellar job and your answers could serve as the model answer key; 2 if, roughly speaking, at least ¾ of your problem set is stellar; 1 if you did not acheieve that standard but your answers are mostly correct; and 0 otherwise. Problem sets handed in late will get a grade of 0. Answers to questions that call for intuition or interpretation are given more weight than calculations. A grade of 1 is a signal to you that you need help. If you feel you need help, please do not wait: please seek help from the TAs or me immediately. The final problem set grade will be based on the total score over all problem sets. I cannot overstate the value to you of making a good faith attempt to complete every problem set and submit it on time. Each problem set will be graded by one TA. If you have any questions, that TA will be the best person to review your answers after you have submitted your problem set. All questions about grades should be taken up with that TA.

Answer keys for each problem set will be posted on the class webpage soon after the submission deadline. Each problem set will be reviewed in section. The section will go over the answers to

questions that posed the greatest difficulty and provide insights into how to think about the problems to strengthen your understanding of the material covered in class.

Please carefully review the posted answer keys. There are three reasons this is a good use of your time. First, it is very difficult for us to identify every instance in which you do not understand something; it is your responsibility to make sure you understand all the material covered in each problem as laid out in the answer key. Second, the answer key is designed to provide a succinct, clear answer to each question; following the model used in the answer key will stand you in good stead in this class and beyond. Third, it is not realistic to provide individual-specific detailed written feedback tor each student on their answer.

Quizzes

Quizzes will be administered both in and out of class. The in-class quizzes will be administered at random times and are intended to assess your understanding of concepts covered in lectures up to that point and to provide feedback to you regarding your grasp of the material covered in the class. The quizzes will account for **15%** of the final grade. There are no make-up quizzes. If you are absent from class, no matter what the reason, your grade for that in-class quiz will be zero.

Weekly discussion section

Each weekly discussion section will be led by a TA. The TAs are experienced teachers who have expertise in econometrics. Sections will provide instruction in STATA, including classes on good programming and data management practices that will be helpful for other classes and beyond. Sections will review problem sets and extend ideas covered in the problem sets and reinforce material covered in this class or the stats pre-requisite. The sections in any week are designed to be identical; you are welcome to attend the section that fits your schedule best in any week.

Mid-term exam

There will be one mid-term which will account for **15%** of your course grade. The mid-term will be open book. You may bring your computer, tablet, books or materials you want to the exam. You may not communicate with anyone inside or outside the classroom during the exam and you may not use an AI-based tool to help answer questions. There will be no make-up midterm. If you are unable to take the midterm, you must provide me with a written explanation *before* the mid-term. If the written explanation provides a valid reason that is clearly beyond your control, and I judge that it is appropriate to do so, then I will substitute your final grade for your midterm grade. In any other instances, your midterm grade will be zero.

Final exam

The final exam will be closed book and will cover all the material in the course. You do not need to memorize formulae for this class. You may bring one 8½*11 page of formulae for reference during the final exam; you may use both sides of the page. No other reference material is allowed. You may not communicate with anyone inside or outside the classroom during the exam; this includes texting, emailing or any other form of electronic communication. The final exam score will contribute **40%** of the final grade for the course. If you miss the final exam for a reason that is outside of your control, with the approval of the Dean and if I judge that it is appropriate, you may be able to take the final exam the next time it is offered. In that case, I will substitute your grade in that exam for the final exam grade for this class after adjusting the grade so that the mean grade in both exams is the same. The date and time of the final exam is determined by the Registrar: it cannot be rescheduled.

Sharing class materials

All the material you need for this course will be available on the class web page. The material is copyrighted which means you may not re-distribute any material such as hand outs, slides, problem sets, exams or answer keys that we have created. Re-distributing includes posting material on a

website, server, shared drive, file host or similar service or providing material to someone who is not enrolled in the class this term. *Please* do not distribute any materials from this class. Since I change the class each year, I strongly discourage you from attempting to use material from prior classes; in my experience, that results in confusion and disappointment.

Reading

The recommended text for this class is

> Wooldridge, Jeffrey M. (2020) *Introductory Econometrics: A Modern Approach, 7th Edition,* Cengage (ISBN: 978-1337558860)

Access to this book is encouraged. If you purchase or rent the book, you may buy the paper or electronic version; you do not need access to the on-line resources and an earlier edition of the book will work well. You may be able to borrow the book from the library. If you find you do not like the presentation in Wooldridge's book, I encourage you to look at one or more of the following:

> Angrist J. and S. Pischke. (2014) *Mastering Metrics: The Path from Cause to Effect.* Princeton University Press.
>
> Goldberger, A. (1998). *Introductory Econometrics*, Harvard University Press.
>
> Gujarati, D. (2003), *Basic Econometrics*, Irwin-McGraw-Hill.
>
> Hill, C., W. Griffiths, and G. Lim (2011). *Principles of Econometrics*, Wiley.
>
> Stock, J and M. Watson (2011) *Introduction to Econometrics*, Pearson.
>
> Studenmund, A. H. (2013) *Using Econometrics: A Practical Guide*. Pearson.

None of these books covers statistical theory comprehensively. If during the course you feel you need a statistical reference, use the text from your statistics course or look at any of the following:

> Hogg, R. V. and E. Tanis and D. Zimmerman. (2015). *Probability and Statistical Inference*, MacMillan, 9th edition.
>
> DeGroot, M. H. and M. J. Schervish (2019): *Probability and Statistics*, Pearson, 4th edition.
>
> Rice, J. (2007): *Mathematical Statistics and Data Analysis*, Duxbury, 3rd edition.

I will provide handouts throughout the course to supplement the lectures and textbook. There is no shortage of alternative text books and there is an abundance of material on-line that will supplement this course. The class web-site has a section on "Additional material to support this class" which includes, for example, links to videos that cover multivariable calculus. If you feel you need help with the concepts covered, please review that material.

Course Outline and Required Reading

Readings from Wooldridge are intended to complement the lectures.

		Wooldridge
1.	*Introduction to econometrics*	Chapters 1 and 19
2.	*Simple linear regression model*	Chapter 2, 9.5, 9.6
3.	*Fundamentals of multiple regression*	Chapter 3, 9.2 9.3, 9.4
4.	*Theory of estimation and inference*	Appendices B and C
5.	*Linear regression model: Inference*	Chapter 4, 5
6.	*Linear regression model: Interpretation*	Chapter 6
7.	*Indicator variables*	Chapter 7
8.	*Non parametric regression*	

9.	*Non-spherical errors: Heteroskedasticity and correlated errors* *Resampling methods*	Chapter 8
10.	*Limited dependent variable models*	Chapter 17
11.	*Identification of causal effects: The problems* *Omitted variables and sample selectivity*	Chapter 9
12.	*Identification of causal effects: Some potential solutions*	
12a.	*Experimental approaches*	
12b.	*Non-experimental approaches*	
12c.	*Matching methods, regression discontinuity designs*	
12d.	*Instrumental variables and two stage least squares*	Chapters 15 and 16
12e.	*Panel data methods*	Chapters 13 and 14

If you have difficulty with the material and Wooldridge's presentation, you should consult one of the alternate readings. They are listed, with chapter references, below. Alternate readings are identified only to assist you and are not required.

Alternate readings

Sections 1, 2 and 3: Introduction to regression model

Angrist and Pischke	Chapters 1 and 2
Goldberger	Chapters 1 and 2
Gujarati	Chapters 1, 5 and 6
Hill, Griffiths and Lim	Chapters 1, 2 and 4
Stock and Watson	Chapter 1

Section 4 : Theory of estimation and inference

Goldberger	Chapters 2-4
Gujarati	Chapters 2, 3
Hill, Griffiths and Lim	Chapters 1P, 3
Stock and Watson	Chapters 2-3

Section 5, 6 and 7: Classical multiple regression model

Angrist and Pischke	Chapter 2
Goldberger	Chapters 6 -12
Gujarati	Chapters 7-9
Hill, Griffiths and Lim	Chapters 5-7
Stock and Watson	Chapters 4 -7

Sections 8, 9 and 10: Relaxing assumptions of the regression model

Goldberger	Chapters 13-17
Gujarati	Chapters 10-14
Hill, Griffiths and Lim	Chapter 8
Stock and Watson	Chapters 8, 9 and 11

Sections 11: Unobserved heterogeneity and instrumental variable methods

Angrist and Pischke	Chapter 3
Goldberger	Chapters 18 and 20
Gujarati	Chapter 15
Hill, Griffiths and Lim	Chapters 10-11
Stock and Watson	Chapter 12

Section 12: Panel data methods

Angrist and Pischke	Chapter 5
Gujarati	Chapters 10, 11 and 12
Hill, Griffiths and Lim	Chapter 15
Stock and Watson	Chapter 10