

Core Empirical Research Methods

It's nearly impossible to overstate the value that economists ascribe to cleverness. Like most obsessions, this one is not altogether healthy.

– [David Autor](#)

My general philosophy in life is never to rely on being clever; instead I want to rely on being thorough and having a justifiable workflow.

– [Richard McElreath](#)

A theoretical statistician knows all about measure theory but has never seen a measurement whereas the actual use of measure theory by the applied statistician is a set of measure zero.

– [Stephen Senn](#)

Overview

This is the course website for *Core Empirical Research Methods* (core ERM), a 1st-year MPhil course in the Economics Department at the University of Oxford. Core ERM will help you develop the basic skills you'll need to carry out applied economic research. It will cover a mix of applied econometrics, programming/computing, and research skills. The prerequisites are basic familiarity with programming in *some language*, not necessarily R, and an introductory course in econometrics at the masters level. If you are interested in auditing this course see [Auditing Core ERM](#) below.

Because Core ERM is about doing economics, it will not be a traditional lecture course. Students should bring their laptops to lectures so that they can follow along with live demos and work on examples in small groups. While there will still be some lecture-style material, the overall format will be closer to a "lab" in the natural sciences. As such **attendance is mandatory if you are taking this course for credit**. GTAs (Graduate Teaching Assistants) will attend each lecture to help give you individualized help if you get stuck while working through in-class exercises. See [Required Software](#) for details on how to configure your machine for core ERM.

Personnel

Lecturer: [Francis J. DiTraglia](#)

Teaching Assistants

- [Inbar Amit](#)
- [Ana Verdnik](#)

Times and Locations

All class meetings will take place in the [Manor Road Building \(MRB\)](#)

Lectures

Weeks 1-8 of Trinity Term, MRB Lecture Theatre. **Lecture attendance is required if you are taking this course for credit.**

- Wednesdays 11-12:30pm
- Thursdays 11:30-1pm
- Fridays 11:30-1pm

Drop-in Surgeries

Weeks 2-9 of Trinity Term in MRB Seminar Room D. (For the North Americans among us, these are *GTA office hours*.)

- Mondays 2-5pm
- Tuesdays 2-5pm

Required Software

In this course we will use the [R programming language](#) via a front-end called [RStudio](#). Both are freely available on all major platforms. To install them [follow these instructions](#). To smooth out the inevitable start-of-term kinks, during week 1 we will work with RStudio via [Posit Cloud](#). Please [sign up for a free account here](#). This will allow you to get right to work at the start of term even if you encounter problems installing R. Eventually you will need to get R and RStudio working on your own machine, however. The week 2 [drop-in surgery](#) is an excellent place to get help with installation issues.

Assessment

This course is pass/fail and will be assessed entirely on the basis of a single coursework assignment: there is no exam. Before we go any further: yes it is possible to fail core ERM. See [Re-sits](#) for more details. Your final assignment will be released in Week 8 and will be due at ***noon on Wednesday, June 19th, 2024*** (TT Week 9) and must be submitted via [Inspira](#). See [Inspira Submission Requirements](#) for more details on how to submit. The assignment will consist of two

parts. These will be equally-weighted and assessed according to the same [marking criteria](#) as detailed below. The two parts are as follows:

Part A: Problem Set Questions

Part A will consist of three questions drawn from your weekly [problem sets](#), *half as many as last year*. The logistics for Part A will be identical to last year: the questions will be announced in Week 8 and will be due in Week 9.

Part B: A Mini-project of Your Choice

Part B will consist of “mini-project” of your choice that you will complete between weeks 3 and 9 of term. For full details, see the [Mini-Project FAQs](#) below. Because you choose the mini-project, you can work on something that is intrinsically interesting to you. Ideally the topic will be relevant to your MPhil thesis; you can *kill two birds with one stone*. And because you will complete your mini project during the term, you will have the opportunity to get help and feedback from me and your GTAs at lectures and the weekly [drop-in surgeries](#).

Academic Integrity

You are allowed, and indeed encouraged, to discuss course problems and assignments with your classmates and GTAs, but you are not allowed to directly copy code or results from another student. The work that you submit for assessment must be your own, even if it incorporates suggestions from your classmates and GTAs. You are likewise allowed to use large language models (LLMs) as a tool to help you learn R, find errors in code that doesn’t run correctly and so on. In other words, you are free to consult an LLM *in the same way* that you are free to consult your classmates and GTAs. I would be remiss, however, if I failed to sound a note of caution. ChatGPT4 is an incredibly valuable tool when used *appropriately*. While it is capable of generating convincing solutions to my problem set questions entirely on its own, it is as yet incapable of generating solutions that would *receive a passing grade* according to my [Marking Criteria](#). Check yourself before you wreck yourself.

Lecture Slides

- **Week 1**
 - [Crash Course in R Programming](#), Solutions
- **Week 2**
 - [Getting Started with dp1yr](#) , Solutions
 - [Getting Started with ggplot2](#) , Solutions
 - [Monte Carlo Simulation Basics](#), Solutions
- **Week 3**
 - [Research Plumbing I](#), Solutions

- [Linear Regression](#), Solutions
- [Running a Simulation Study](#), Solutions
- **Week 4**
 - [The Multivariate Normal Distribution](#), Solutions
 - [Instrumental Variables](#), Solutions
 - [Research Plumbing II](#), Solutions
- **Week 5**
 - [Logistic Regression](#), Solutions
 - [Statistical Inference - Defense Against the Dark Arts](#), Solutions
 - [Heteroskedasticity and Clustering](#), Solutions
- **Week 6:**
 - [Panel Data Basics](#), Solutions
 - [Selection-on-observables](#)
 - [DAGs and Bad Controls](#)
- **Week 7:**
 - [IV and Local Average Treatment Effects](#)
 - [Regression Discontinuity](#)
 - [Difference-in-Differences](#)
- **Week 8:**
 - [Mini-project Help Session](#)
 - [Mini-project Help Session](#)
 - [Mini-project Help Session](#)

Draft Book

When I first taught this course back in 2022, I started writing a book to accompany it. This turned out to be a tall order, but I did manage to produce ten draft chapters. You can view them at <https://empirical-methods.com>. Based on my experiences teaching version 1.0 of core ERM, I decided to make a number of changes to the course. While much of the material in my draft book remains relevant, my lecture slides will be the final authority on the course material in the 2024 version of core ERM. I hope to rework the book before next year’s version of core ERM.

Re-sits

Barring a serious personal issue that affects your studies, there is no reason why you should fail core ERM. If you attend class, participate actively, solve the weekly problem sets, and get help at the [drop-in surgeries](#) as needed, you will develop all of the skills needed to complete the course assignments to the [appropriate standard](#). If for some reason you *do fail* core ERM, you will be given an opportunity to complete a re-sit assignment that is due in early September. The summer re-sit assignment will be different from the original course assignments and, as it will not include

any problem set questions, you may well find it more challenging. Working on a re-sit assignment is certain to spoil your summer, so take my advice and do what's necessary to **pass the first time around**.

Problem Sets

There will be a problem set for each of weeks 1–7 of TT. Your final assessment will feature **three questions** taken verbatim from these problem sets. In light of the rocky start we've gotten this term thanks to the microeconomics exam fiasco, I have released sample solutions to the problem set questions from Weeks #1 and #2. I will not release solutions to the other problem set questions. You are allowed, and indeed encouraged, to discuss these questions with your classmates and GTAs, although you are not allowed to submit code or results directly copied from another student as part of your final assessment.

- Week 1: [Making Change, Sieving Primes](#)
- Week 2: [Bias in the Labor Market, Growth Accounting, Desert Island Monte Carlo](#)
- Week 3: [Quarto/RMarkdown, Football & Market Efficiency, Optimal Stopping](#)
- Week 4: [Bivariate Normal Simulation, Colonial Origins, Lead Cleanup](#)
- Week 5: [Contaminated Wells in Bangladesh, Filedrawer / Publication Bias, Robust SEs,](#)
- Week 6: [Airfare Panel, NSW versus CPS](#)
- Week 7: [Minimum Legal Drinking Age, Minimum Wage](#)

Mini-project FAQs

What is a mini-project?

It is a small project project of your own choosing, roughly equivalent to three problem set questions in terms of the time required to complete it. You will complete your mini-project between weeks 3 and 9 of term and submit it as part of your final assessment.

What can I submit as my mini-project?

You mini-project should be a *replication* (or partial replication) of a reputable paper in economics or a closely related field. For an **applied paper** this would involve obtaining the data for the paper, and writing R code to clean the data and reproduce a few key tables and figures from the paper, e.g. those containing summary statistics and main results. For an **econometrics paper** it would involve programming an econometric procedure that isn't already built into R and using your code to replicate all or part of the simulation results from the paper.

Can I choose any paper that I want for my mini-project?

There are four key rules for choosing a paper for your mini-project:

1. Each student must choose a different paper.
2. There should not *already* be R replication code for your chosen paper available online. (It's fine if there's code in a different language, e.g. STATA.)
3. For an applied paper, the data you need must be available online and not come with special access restrictions.
4. Your project must be approved by me or your GTAs before you start working on it (see below).

Subject to these caveats, you can choose any paper that you like for your mini-project. That said, we *strongly suggest* considering a paper from one of the two sources listed in the next section.

Help! I don't know how to find a paper to replicate!

Not to worry. Here are two excellent resources that you can use to find a paper to replicate:

- <https://www.socialsciencereproduction.org/>
- <https://i4replication.org/>

If you're stuck you are also welcome to come to the drop-in surgeries to ask your GTAs for suggestions on good papers to replicate.

How will I know if my project is good enough?

Before you start working in earnest on your mini-project we will have a short discussion with each of you individually to make sure that your project is neither overly ambitious nor overly simple. If we are happy with your project, we will **approve** it. If it is not quite right, we will suggest some changes that you can make to get it approved or help you choose an alternative project. While you are working on your project, you can ask for help and feedback from me and your GTAs. You can even say: "here's what I've done, does this look good to you?" and we will be happy to give you feedback. The drop-in surgeries offered by your GTAs are a great opportunity to get help with your mini-project. I may also allocate some time during lectures to provide feedback on the mini-projects.

How will the project approval process work?

We will post a sign-up sheet online shortly where you will be able to list the paper you want to replicate and check that no other student has chosen the same one. On the sign-up sheet there will also be a place for you to supply a short description of the parts of the paper you intend to replicate. Based on this information, your GTAs and I may be able to approve your project as-is, without the need for a meeting. If not, we'll ask you to email us a bit more detail about your project and schedule a short meeting to discuss further. We will aim to complete the project approval process as quickly as possible and by the beginning of week 6 by the latest.

In what format should I submit my mini-project?

You should submit output generated from RMarkdown/Quarto file, containing your code and output, as well as a short written report that explains what you did and what you found. I will provide further details in lecture.

How will my mini-project be assessed?

The same marking criteria will apply to the written report as to the problem set questions. Each of the two parts of your final submission will be assessed on a pass/fail basis, using the marking criteria from the course website. You must submit both parts to an appropriate standard to pass the course. The re-sit policy from the course website applies to the entire final assessment, including both parts.

Can I work with a partner on my mini-project?

Each student will complete a mini-project based on a different paper: these are individual assignments. The same academic integrity policy applies to mini-projects as to problem set questions. In particular: you are welcome to discuss your work with your GTAs, your classmates, but the work that you submit must be your own.

Marking Criteria

Assignments in Core ERM will be graded pass/fail based on five criteria. Criteria 1–3 are all-or-nothing and *necessary* to pass. Criteria 4 and 5 allow for partial marks.

1. **Clean Code:** Your R code must adhere to the [tidyverse style guide](#). It should be clean, easy to read, and appropriately commented.
2. **Correct Code:** Your R code must be syntactically correct, i.e. it must run without errors. This will be assessed based on your ability to successfully knit an RMarkdown/Quarto file with your results: your file will not knit unless the code is correct. More details on RMarkdown/Quarto will be provided in lectures.
3. **Formatting & Typesetting:** For a given assignment, you will submit a single pdf document constructed from one or more underlying RMarkdown/Quarto reports incorporating your code and detailing your solutions to the questions on the assessment. Your write-ups should be clearly formatted using appropriate [markdown](#) commands. Any mathematical formulas that you incorporate should be clearly and cleanly typeset using appropriate [LaTeX](#) commands.
4. **Completeness:** To pass a given question on a course assignment, your answer must at a minimum be substantially complete. Partial solutions only receive partial marks, regardless of quality.
5. **Quality:** To pass a given question on the assignment, your answer must be substantially correct. Poorly explained or substantially incorrect answers will only receive partial marks. *If I can't tell that you understand what you're doing, you will not pass a given question.*

Inspira Submission Requirements

These will be posted soon.

Auditing Core ERM

Between 80 and 90 students take Core ERM for credit each year, but the [MRB](#) lecture theatre seats 120. Provided that there's space left in the room, any member of the university is most welcome to attend my lectures without asking for permission in advance. I ask only that you respect the following guidelines. First, please sit in the back row if you're auditing so that I can more easily gauge attendance etc. Second, the Drop-In Surgeries are only for students who are taking the course for credit. Third, Core ERM lectures are fairly interactive: I and the GTAs will circulate to help students who encounter difficulties while working on the exercises. I won't go so far as to say that we *won't help you* if you're auditing, but we will need to prioritize the students who are taking the course for credit.