Economics 662

Fall 2024 **Russell Davidson**

Textbooks

The principal reference for this course used to be the textbook *Econometric Theory and Methods* (ETM), Oxford University Press, ISBN 0-19-512372-7, by James MacKinnon and me. However, it is now much better to rely mainly on the new ebooks *Foundations of Econometrics Part 1* and *Part 2*. These are partially based on ETM, but are shorter, since they contain little more than the material needed for this course. Some of you may be interested to know that ETM itself has been translated into Chinese and Russian. More information about this book can be found at

`https://qed.econ.queensu.ca/ETM/`

Both the ETM textbook, and also the older textbook, *Estimation and Inference in Econometrics* (EIE), Oxford University Press, ISBN 0-19-506011-3, by James MacKinnon and me as well, are available at the above website as PDF files, and can be accessed or downloaded completely free of charge. We ask you to respect the copyright, which is held by us, the two authors, since 2021.

Be warned that the book referred to as EIE treats things at a more advanced level than is needed for this course. It may well still serve as a useful reference for certain things.

If you have, or can find, a hard copy of either book, please note that both of them have undergone a number of printings, with a few corrections incorporated each time. Even if you have the first printing of either book, that would serve perfectly well, since all corrections, right from the beginning, are available on the book homepage, at

`http://qed.econ.queensu.ca/pub/dm-book/`

The versions available on the website are of course the most up to date.

Software

Econometrics is primarily an applied discipline. The theory is rich and satisfying, but it is just an exercise in navel contemplation if it is not used to guide empirical practice. There are two aspects of this practice that I would like to get across. The first is the use of computers in order to implement the various estimation and testing procedures that theory provides us with. Econometrics software is no harder to use than most other sorts of software, although it is highly desirable to get out of the "point and click" frame of mind. Most packages require a certain amount of programming. However, many econometrics packages are so powerful that, if you manage to get them to read in your data, they will do all the rest, with no further thinking on your part. Since thinking is a desirable activity in students (and others), many of the exercises on the assignments and in the exercises to be found at the end of each chapter of the *Foundations* texts, or of ETM, will require you not to make use of all the features of sophisticated packages, in order that you may see more clearly how those features can be built up from simpler ones.

For many years, even decades, now, packages like TSP and Shazam have epitomised econometric software. Many such packages exist, and they provide a convenient way to perform most of the numerical operations useful for econometrics. In recent years, however, programs like Matlab, Gauss, and Ox, which are essentially matrix programming languages, have been growing in favour. I must also mention R, a statistical package that has many aficionados. The reason most frequently cited for this move away from the traditional packages is that the modern packages run programs faster, and this is certainly true for programs involving lots of simulations. You should be warned that, if you want to do "industrial strength" simulations, even these programs waste a good number of CPU cycles compared with programs written in a lower-level language like Fortran or C++ .

This year, at least, the recommended software packages for econometrics are, first, MatLab, a commercial product, but freely available to people at McGill on account of a site licence, and, second, Python, a general-purpose interpreted programming language. Python is free software (prefer Python 3 to Python 2) and has a vast number of libraries available to it for many things, including econometric calculations.

Some people seem to think that good empirical practice is all there is to econometrics. In one sense, that is so, but my experience has shown me that there is no good empirical practice without a good mastery of the underlying theory. It can be tempting to think of econometrics as a set of cookbook recipes, especially as so many of these recipes are made available by modern software. But it is all too easy to apply recipes wrongly if you do not understand the theory behind them. (This remark also applies to the cooking of food!) Thus the second vital aspect of econometric practice is *understanding* what data are telling you. Although I can make you do exercises that should make you competent in the implementation of a number of procedures, no one can (directly) teach you how to interpret the results of these procedures. Such interpretation is more an art than a science, and can therefore be taught best by example. Unfortunately, we do not have too much time for that. But even if some of the exercises you will be given in the assignments use simulated rather than real data, I will try to make you think of how your results can be interpreted. Making a practice of that may well save you from purely formal errors in the exercises.

Detailed description

I plan to cover the entirety of *Foundations of Econometrics Part 1*, although some of it will be covered rather quickly, and as much of Part 2 as we have time for certainly not all of it. The contents of Part 1 are briefly described below.

Chapter 1: *Models in Science*. This is a completely new chapter. It contains a somewhat philosophical discussion, and it may well be that some people will disagree with the point of view adopted here. Textbooks and scientific papers normally do not broach philosophical questions, and some scientists have been known to express the opinion that philosophy is a waste of time for practising scientists. Obviously we disagree, but here we give fair warning that people with a different cast of mind may omit this preliminary chapter, and lose nothing of conventional econometrics by so doing.

Chapter 2: *Regression Models*. A chapter containing a brief review of regressions, along with a few reminders of things from statistics and probability theory that will be needed later. Very important is the subsection of Section 2.3 entitled "Simulating Econometric Models". The ideas in that subsection are essential for understanding the bootstrap and much else, and are not too easily found elsewhere. There are also sections dealing with various useful aspects of matrix algebra. You may find that this is well-known stuff, except perhaps for some of the material on partitioned matrices.

Chapter 3: *The Geometry of Linear Regression*. In this chapter, statistical issues are set aside in order to discuss ordinary least squares as a purely formal procedure. The chapter begins with some straightforward geometry, and introduces the concept of vector, or linear, spaces. The most important concept introduced here is that of **orthogonal projection matrices**, which are an indispensable tool in developing econometric theory. Section 3.4 presents the important **FWL theorem** (FWL = Frisch-Waugh-Lovell). We will spend a little time on this exceedingly useful result. As an application of the theorem, Section 3.6 deals with the phenomenon of leverage, whereby some observations in a sample have much more influence on OLS parameter estimates than others. This is a topic of great importance that is not treated in all econometrics texts.

Chapter 4: *The Statistical Properties of Ordinary Least Squares*. In this chapter, the most fundamental concept of econometric theory is introduced, that of a **data-generating process**, or **DGP**. This concept allows us to define the almost equally important concept of a statistical or econometric **model**, and how such models are **specified**. Pretty much the simplest regression model is what we call the **classical normal linear model**, and this is introduced in this chapter. Much of the material in later chapters allows us to relax the very restrictive assumptions that are made in specifying this model. Although it is restrictive, it is only for this model that the beautiful, exact, results of classical regression theory are true. Some of these results are developed here, including the **Gauss-Markov theorem**, which proves the efficiency of the OLS estimator under classical assumptions. In this chapter, we also introduce some concepts, most importantly that of **probability limits**, needed for **asymptotic theory**, the approximate theory used for more general models, for which the exact classical results do not hold. Using this theory, we can show that the OLS estimator is **consistent** under much weaker conditions than the classical ones.

Chapter 5: *Hypothesis Testing in Linear Regression Models*. The two chief activities in econometrics are estimation and inference. Estimation of linear regressions is easy – just use OLS – and so this chapter is devoted to inference. The easiest approach to inference is hypothesis testing, and that is the topic of this chapter. After defining the basic concepts that underlie hypothesis testing, we develop tests for linear regressions using the geometric ideas of Chapter 2 for the special case of the classical normal linear model. For it, there exist exact results about the distribution of the test statistics, but, more generally, it is necessary to have recourse to asymptotic theory. This chapter contains two sections that are new in this book, one on performing tests of several hypotheses simultaneously, and one on pretesting.

Chapter 6: *Confidence Sets and Sandwich Covariance Matrices*. This chapter continues the story of statistical inference. The first three sections cover confidence intervals and confidence regions; the next three deal with covariance matrix estimation in some circumstances in which the disturbances are not independent and identically distributed. The robust covariance matrix estimators discussed are the HCCME (heteroskedasticity-consistent covariance estimator), HAC (heteroskedasticity and autocorrelation consistent) estimators, and the CRVE (cluster-robust variance estimator). The next section deals with the important difference-in-differences technique of estimation, and the final section of this chapter explains the delta method, as a way to estimate the covariance matrix of a set of nonlinear functions of parameter estimates.

Chapter 7: *The Bootstrap.* This chapter is not entirely new, since the bootstrap is mentioned in numerous places in ETM. Here, however, we have collected the material on bootstrapping in ETM, and added a good deal more. The bootstrap is mentioned again in the following two chapters, but it seemed important not to postpone consideration of the bootstrap until after treating all the other topics in the book. In this chapter, we start with bootstrap hypothesis tests, along with bootstrap P values, and proceed to the study of bootstrap confidence sets. Bootstrapping is also discussed in cases in which the robust covariance matrix estimators of the previous chapter should be used.

Chapter 8: *Instrumental Variables Estimation.* In economics, typically "everything depends on everything else". In econometric parlance, almost all economic variables are endogenous. Their endogeneity means that they cannot be used as explanatory variables in regression models estimated by least squares, and this is not something that asymptotic theory can get around. A new estimation method is needed, and we are led to the study of instrumental variables. Instrumental variables estimation is a very natural generalization of the **method of moments** that was used earlier to justify least squares. For estimation by instrumental variables (IV), techniques of inference are considerably different from those previously presented for least squares. The special cases of tests for over-identifying restrictions, and Durbin-Wu-Hausman (DWH) tests, each receive a section, as does bootstrapping models estimated by IV.

Chapter 9: *Generalized Least Squares and Related Topics.* We continue the process of relaxing the restrictive classical assumptions by considering models in which the disturbances may have a more complicated specification, in particular by being heteroskedastic, or serially correlated, or both. The Gauss-Markov theorem does not apply in such cases, and so, although the OLS estimator remains consistent, it is no longer asymptotically efficient. The **GLS** estimator replaces the OLS estimator as an asymptotically efficient estimator if the pattern of heteroskedasticity or serial correlation is known, or if it can be consistently estimated, in which case we use the **feasible GLS** estimator. In this chapter, there is discussion of heteroskedasticity, both testing for its absence, and efficient estimation when the pattern of heteroskedasticity is known. It is natural that this should be followed by discussion of autocorrelation. Tests, both old and relatively new, with a null hypothesis of no autocorrelation are presented, and then techniques of estimating models with autocorrelated disturbances. A final section applies the ideas of GLS to an introductory treatment of panel data models.

Method of Evaluation

Although the course will be delivered in person, the method of evaluation has been affected by my experience when it was online. There will be some assignments, just how many is hard to predict, perhaps around four, a midterm exam and a final exam. In the past, it was not possible for me to ask you to work on the computer during exams, that is now quite possible. I will therefore take advantage of this and the midterm exam will contain practical work involving computing.

The formal weights will be one quarter for the assignments, one quarter for the midterm, and one half for the final. However, no student will get a lower grade than the one from the final. Thus the course grade will be the greater of the final mark and the weighted average of the assignments, midterm, and final, with the specified weights.

Generative AI

The university now wants all course outlines to specify rules concerning the use of generative AI for a course. As I explained in class, you are welcome to use AI for help with assignments or take-home exams. The only qualification is that each such use must be acknowledged, just as sources are acknowledged in any research paper. You may use any suitable format for the acknowledgement, but please do acknowledge all uses you may make of generative AI, such as Chat-GPT for instance.

Academic Honesty

You'll have seen the following in all of your course outlines, because the McGill Senate requires that it should appear in all of them. I used to think of it as a pure formality, but a disturbing number of cases of plagiarism were detected the year that the course was exclusively online, with students trying to take advantage of the possibility of doing a take-home exam by copying from material found on the Internet. Be warned! If you are caught plagiarising, you are exposed to disciplinary action at the level of the University. So, please take seriously all the admonitions in the following text.

1) Right to submit in English or French written work that is to be graded [approved by Senate on 21 January 2009]: In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded. This right applies to all written work that is to be graded, from one-word answers to dissertations.

2) According to Senate regulations, instructors are not permitted to make special arrangements for final exams. Please consult the calendar, section 4.7.2.1, General University Information and Regulations, at http://www.mcgill.ca.

3) Academic Integrity statement [approved by Senate on 29 January 2003]: McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures.

   (see http://www.mcgill.ca/students/srr/honest/ for more information).

Et en français:

L'université McGill attache une haute importance à l'honnêteté académique. Il incombe par conséquent à tous les étudiants de comprendre ce que l'on entend par tricherie, plagiat et autres infractions académiques, ainsi que les conséquences que peuvent avoir de telles actions, selon le Code de conduite de l'étudiant et des procédures disciplinaires (pour de plus amples renseignements, veuillez consulter le site http://www.mcgill.ca/students/srr/honest/)