# Economics 468

**Fall 2010** **R. Davidson**

The principal reference for this course is the textbook *Econometric Theory and Methods* (ETM), Oxford University Press, ISBN 0-19-512372-7, by James MacKinnon and me. We have just learned that it now exists in a paperback edition, which is probably cheaper than the hardback.

The main secondary reference for the course is the textbook *Estimation and Inference in Econometrics*, Oxford University Press, ISBN 0-19-506011-3, also by James MacKinnon and me. This book treats things at a more advanced level than is needed for this course. *Estimation and Inference* has undergone around six printings – I've lost the exact count – and the successive reprintings allowed us to make a few corrections each time. My copy, sent to me by the publisher as an evaluation copy, is from the fourth printing. But even if you had the first printing, 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/`

All the chapters of ETM have numerous exercises at the end of the chapter. These are of various sorts. Some fill in gaps that are not explicitly handled in the main text, and are thus chiefly of theoretical interest. Some are for practice, and these usually include working with real data or else doing simulation experiments. Others are rather like the questions that may well be asked on exams. Some of the exercises are marked with a star. This means that solutions to them are available from the book's website, at `http://www.econ.queensu.ca/ETM/`. The starred exercises were selected either because they are very important or because they are rather difficult. Some students have asked for solutions to be made available for all the exercises. Not only would this get us into trouble with the publisher, but it would defeat the point of the exercises, which is to help people learn econometrics. If you think about it, I hope you will see that it's much better to struggle with an exercise and solve it than just to look the answer up "in the back of the book". In any case, answers to anything asked on assignments will be provided after the assignment is due.

A topic that may be new to many of you, but that figures in the early chapters, is the bootstrap. The bootstrap is also discussed in *Estimation and Inference*, but in the very last chapter. Please do not read the material on the bootstrap there! It is horribly out of date. The bootstrap is a simulation-based method, and, as such, is one illustration among many of the importance of simulation for modern econometrics.

There follows a set of remarks about the chapters of ETM. We won't cover all of these in the first term, but, for those intending to take the econometrics course next term, it may be useful to have this material as a reference.

**Chapter 1:** *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 1.3 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 2:** *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 2.4 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 2.6 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 3:** *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 4:** *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. Tests can be exact if the assumptions of the classical normal linear model are satisfied, but otherwise asymptotic theory allows us to construct approximate tests. In many cases, we can do better than these approximate tests by using the **bootstrap**; the elements of bootstrap testing are covered in this chapter.

**Chapter 5:** *Confidence Intervals.* Confidence intervals provide another way to conduct statistical inference. At a rather deep level, there is an equivalence between hypothesis tests and confidence intervals, but this equivalence is not always immediately obvious. Like hypothesis tests, confidence intervals can be exact, under the strong assumptions of the classical normal linear model, for instance, or approximate. Approximate confidence intervals can be based on asymptotic theory or on the bootstrap. Another topic introduced in this chapter, one of the first in which we abandon some of the classical assumptions, is heteroskedasticity consistent inference.

**Chapter 6:** *Nonlinear Regression.* In this chapter, we relax the "linear" part of the classical assumptions, by considering nonlinear models. In order to deal with these, we need to develop asymptotic theory more seriously than heretofore. We introduce the properties of **consistency**, **asymptotic normality**, and **asymptotic efficiency**, that we would like our estimators to possess. By seeking to fulfil this last requirement, we are led to the method of nonlinear least squares. We have a brief discussion of how NLS estimates can be computed, because this is not necessarily a straightforward matter, and the issues apply more generally to other sorts of nonlinear estimation. A concept that has its origins in the study of algorithms for nonlinear estimation is that of an **artificial regression**. The one used for nonlinear regression is known as the **Gauss-Newton regression**. It turns out that it has many more uses than those for which it was invented. In particular, it is possible to base almost all hypothesis testing on artificial regressions, for reasons that are best understood geometrically.

**Chapter 7:** *Generalized Least Squares and Related Topics.* We continue the process of relaxing the restrictive classical assumptions by considering models in which the error terms 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, we introduce some ideas related to **time series** that arise naturally from the study of serial correlation. Numerous tests are discussed, exact, asymptotic, and bootstrap, for both heteroskedasticity and serial correlation. The final section of this chapter deals with panel data, which are data where each observation is indexed by both a time index and a cross section index.

**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. It applies to both linear and nonlinear regressions, and it can be associated with an extension of the Gauss-Newton regression that can be used to implement hypothesis testing, both with white-noise errors and with heteroskedastic errors.

**Chapter 9:** *The Generalized Method of Moments.* This is indeed a very general method. It (we usually refer to it as **GMM**) subsumes all the other methods of estimation that we have treated up to this point as special cases. After working the method out in detail for the simple case of linear regressions, we provide some more general theory that can be applied to models of very many different sorts. In this chapter, we concentrate on the method, in particular **heteroskedasticity and autocorrelation consistent**, or **HAC**, estimators. In later chapters, the method is applied to various particular sorts of models. In terms of theory, this chapter contains a synthesis of all the principles of estimation and inference used in earlier chapters.

**Chapter 10:** *The Method of Maximum Likelihood.* This method can be thought of as an approach lying at the theoretically opposite extreme from GMM. It is a classical method, and imposes very restrictive assumptions on the models to which it can be applied. In fact, it is necessary to assume that the DGP is known up to a finite number of parameters than can be estimated by ML. Associated with the method, which is an estimation method, there is a classical approach to hypothesis testing, which is also laid out in this chapter.

**Chapter 11:** *Discrete and Limited Dependent Variables.* In this chapter, we use maximum likelihood to develop methods for use with dependent variables that are discrete (often binary), or else censored or truncated. In econometric practice, models involving such variables arise very frequently indeed. They require a certain number of extensions of the standard theory of Chapter 10, but remain reasonably simple to estimate, and they are very easy to bootstrap.

**Chapter 12:** *Multivariate Models.* We use instrumental variables when the explanatory variables we wish to use in our models are endogenous. That means that the model refers to more than one dependent variable. A complete model must provide a specification of the joint DGP for all of them. Such a model is called "multivariate", and the set of equations that define the model are called **simultaneous equations**. In principle, techniques for estimating simultaneous equations models, and doing inference on them, are applications of GMM, but there are enough implementation details for it to be necessary to treat these models separately. Long before GMM was invented, there was a classical theory of simultaneous equations models, for a long time considered one of the intellectual triumphs of the middle period of econometrics.

**Chapter 13:** *Methods for Stationary Time-Series Data.* This is the first of two chapters on time series. In this chapter, we limit ourselves to stationary series, and develop a good deal of methods that have been in use for quite a while in econometric analysis. Seasonality is one topic discussed in a fair amount of detail.

**Chapter 14:** *Unit Roots and Cointegration.* These rather strange names refer to phenomena of great importance when we deal with time series that are not stationary. The theory discussed in this chapter is relatively recent, and is still an active area of research. Mathematically, it involves different techniques from those used elsewhere in econometrics.

**Chapter 15:** *Testing the Specification of Econometric Models.* This chapter touches briefly on quite a number of topics. One is the concept of **artificial regressions**, a generalisation of the Gauss-Newton regression developed much earlier. Others include tests of nonnested hypotheses and nonparametric estimation. It is possible that we will treat selected topics from this chapter out of order rather than waiting for the end.

There is of course a website for the course provided by WebCT, but several years ago I lost patience with what seems to be another of those unwieldy monsters that consume vast amounts of time for the smallest and most trivial things. Therefore, what you will find there, apart from this outline, is a link to the site I maintain on my own computer:

`http://russell-davidson.arts.mcgill.ca/`

There I will put PDF files containing the assignments, and the answers to these, as the term advances. Data files for the assignments are available there as well, and whatever other materials, links, and so on, that seem suitable.

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, many of the exercises for this course 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, programs like Matlab, Gauss, and Ox, which are essentially matrix programming languages, have been growing in favour. The reason most frequently cited for this is that they run programs faster than the traditional packages, 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++.

Some time ago, I discovered to my distress that, despite what I have just said, many students had no access to any package that would do what they needed to do for this course. I therefore made available the software package that I wrote myself, primarily for the use of my students in France. The package is called ***Ects***, and the documentation was written in French. However, a group of students who took this course a few years ago undertook to translate most of it into English, and did a pretty good job. Both English and French versions of the documentation are on my website, as is the software itself. A new version, version 4.1, has reached stable status, and is also available as well as the version, 3.3, that has been available for some years now.

I don't really approve of proselytism in the course outline, but I'm going to indulge anyway, because I think that students may be misled more easily now than a decade ago as to how best to set about computing. Microsoft makes software for big business. That's where the big money is, not in academics. Thus, their products are *not at all* suitable for scientific computing, and that includes their rotten operating system. Here is a link to a paper by my coauthor James MacKinnon in which he provides very useful information about setting up your computer for serious work, as well as the less serious activities that people sometimes use computers for:

`http://www.econ.queensu.ca/pub/faculty/mackinnon/linux-review.pdf`

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.

In addition to the midterm exam, there will be five assignments for you in the course of the Fall term. The marking scheme is such that the final grade for the term will be

$$\text{Grade} = \max\left(F, \frac{F+M}{2}, \frac{F+(M+A)/2}{2}\right),$$

where $F$ is the grade on the final (Christmas) exam, $M$ that on the midterm, and $A$ a composite grade for the assignments. The main rule is that you cannot do worse on the entire term than what you get for the Christmas exam.

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 have been detected in recent years, not especially at McGill, but in other North American universities. So, please take seriously all the adminitions 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) 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/`.