Introduction: structural modeling

Joan Llull

Structural Empirical Methods for Labor Economics
(and Industrial Organization)

IDEA PhD Program

The founding members of the Cowles Commission defined econometrics as: "a branch of economics in which economic theory and statistical methods are fused in the analysis of numerical and institutional data". Today economists refer to models that combine explicit economic theories with statistical models as structural econometric models (Reiss and Wolak, 2007). In this sense, even a linear regression qualifies as a structural model, provided its parameters have a structural interpretation in terms of some specified economic model.

In this course we review different different structural empirical methods with applications in many fields of economics. We illustrate the different methods with applications mostly drawn from Labor Economics and, to a lesser extent, Industrial Organization. However, these methods are implementable essentially to any field of economics.

The course is divided in three parts, which are interconnected. In the first part, we review static and continuous choice models. All models reviewed here are, essentially, applications of methods already reviewed in the Microeconometrics course and in other courses in different semesters.

Chapter 1 reviews different methods for the estimation of production functions. This chapter is mainly an application of methods reviewed in previous courses for the analysis of continuous outcomes, including ordinary least squares, nonlinear least squares, maximum likelihood, panel data methods, and instrumental variables. The chapter is divided in two main sections. The first one reviews the estimation of production functions at the firm level. This part is very popular in the Industrial Organization literature. The second part is a review of some common issues experimented in the estimation of aggregate production functions in partial equilibrium setups. This part is central for Labor Economics and Macro. As an application, we review a recent working paper by Christoph Albert, Albrecht Glitz and myself on labor market assimilation of immigrants, which is based on the estimation of an aggregate production function by Nonlinear Least Squares.

In Chapter 2 we review some examples of papers that introduce the estimation of discrete choice (static) models, selection models, and dynamic (continuous choice) models. For the first case, we review a paper by Rebecca Diamond, published in the American Economic Review in 2016, in which she estimates a location choice model to explain the diverging location choices by skill in the 1980-2000 period. As an application of selection models in structural estimation, we present a review of the standard static female labor supply model based on the Handbook of Labor Economics chapter by Michael Keane, Kenneth Wolpin, and Petra Todd on structural estimation, published in 2011. Finally, to review the estimation of (equilibrium) continuous choice dynamic models, we review a classic paper by James Heckman, Lance Lochner, and Christopher Taber on human capital accumulation and the sources of wage inequality.

The second part provides a review of the main estimation methods for dynamic discrete choice structural models. Chapter 3 provides a review of the more standard full solution (normally based on maximum likelihood) methods for estimation. Chapter 4 introduces a class of estimators that allow the estimation of the model without need for solving for individual choices in each iteration of the estimation algorithm. These estimation techniques are orders of magnitude faster, and allow for the estimation of more complex models that would otherwise be unfeasible. As applications for these two chapters, we review two papers of mine that estimate equilibrium models of the labor market with immigration. The first one, published in the Review of Economic Studies in 2018, is based on full solution methods. The second one, currently work in progress, provides a related application that uses the methods developed in Chapter 4.

In the third part of the course we review models in which markets deviate from the standard perfect competition without frictions paradigm. The topics reviewed in this part are very popular in the Industrial Organization. In Chapter 5 we review the estimation of dynamic discrete games with incomplete information, popular in the context of oligopoly markets. In that chapter, after reviewing the main estimation methods for this class of models, developed as an application of those methods introduced in Chapter 4, we discuss the paper by Stephen Ryan on the environmental costs of regulation in the cement industry, published in Econometrica in 2012.

Finally, in Chapter 6 we deviate from the dynamic discrete choice paradigm to return to static continuous choice models. In particular, we provide an introduction to the structural estimation of auction models. Even though this approach is very popular in the Industrial Organization context, we review, as an application, a paper that is an application of Macro-Finance: an article by Aaron Barkley, Joachim Grogger, and Robert Miller on the estimation of dynamic treasury-bill auctions, forthcoming in the Journal of Econometrics.

Except for Chapters 1, 2, and 6, the main focus of this course is on a particular subset of structural models: dynamic discrete choice models. These models are a dynamic extension of the discrete choice models analyzed in the Microeconometrics course. From the perspective of a random utility model, the static models seen in Microeconometrics determine decisions though the comparison of current payoffs/utilities associated with each of the alternatives and make decisions accordingly. This approach might be limiting in some contexts. Many economics problems describe the behavior of forward-looking agents, that take into account how their decisions today affect tomorrow's outcomes. Several examples include: human capital formation and career path decisions, migration, investment decisions, machine replacements, smoking, marriage and fertility, social interactions, patenting a product, entry/stay/exit from a product market,... In the central chapters of this course, we model individual behavior by means of a *stochastic dynamic programming model* (DP). This way of modeling implies considering individuals as forward-looking and rational. The parameters to be estimated are structural in the sense that they describe agent's preferences and constraints of the DP, including beliefs about uncertain future events. We rely on the principle of *revealed preference* to estimate these parameters with (longitudinal) micro-data on individuals' choices and outcomes.

The seminal papers of Miller (1984), Wolpin (1984), Pakes (1986), and Rust (1987) show that estimating these dynamic discrete choice models is both feasible and important to answer key economic questions. This approach has important advantages. First, it brings close ties with theoretical models, which gives a very clear interpretation to the parameters in terms of the underlying economic theory behind them and inflicts discipline to the researcher in the need of explicitly specifying the role of each parameter and each unobservable variable in the model. Second, it is a very powerful tool for policy evaluation. Whenever a pilot trial experiment of a particular policy is not feasible (most of the cases), it allows to generate *counterfactual* exercises with the inclusion/exclusion/modification of the policy that is evaluated. And third, when the assumptions of the model are well justified, this approach allows us to extract deeper conclusions when data variability is too limited to identify the outcomes of interest in a reduced form or non-parametrically.

Obviously, these advantages come at a cost. The most important drawback of structural models is that they entail complex and computationally intensive solution/estimation algorithms. The complexity emerges from having to solve the individual dynamic optimization problem for each parameter evaluation (we will see alternatives that avoid solution of the model in estimation). The difficulty in the solution of the individual optimization problem is exacerbated because of the discrete nature of the data: we do not have Euler equations here. Also, there is a "curse of dimensionality" that implies that the computational cost increases exponentially with the number of state variables. The estimation methods based on the seminal work by Hotz and Miller (1993), which we introduce starting in Chapter 4, mitigate this concern, but at the cost of efficiency losses and potential small sample biases. The second problem is that identification relies heavily on functional form assumptions. These assumptions need to be strongly sustained by the researcher.