

ECON0108: Topics covered in lectures 1-5, 2024-5

In this 5-lecture sequence, econometric models are motivated as sets of restrictions defining admissible data generating processes. We see data and would like to know: which particular process delivered that data; with large amounts of data could all but one admissible process be ruled out as the data generator; could it be that a model is misspecified in the sense that no process it admits can generate the data we see? We set up a framework in which such questions can be answered.

We study: econometric model construction, the identifying power of models, analogue estimation built on identifying correspondences, strongly and weakly restrictive models built on parametric, semi- and non-parametric restrictions, overidentification and the falsifiability of restrictions, and partial identification. The focus is almost entirely on structural econometrics.

Economic understanding, argument and theory deliver some of the restrictions employed in structural econometric models. Other restrictions have a less tangible basis. Less restrictive models are more credible, so we study which restrictions are essential if information about interesting economic magnitudes is to be got from economic data. Central in this endeavour is the concept of the identifying power of an econometric model.

In these five lectures we make a careful study of **identification** first considering the concepts involved and then methods for determining whether a model and a probability distribution of observable variables identify a particular structural feature. Since parametric restrictions are rarely credible, we study the identifying power of semi- and non-parametric models, first considering models with additive unobservable variables and then non-additive models.

Weakly restrictive models may admit the possibility that more than one structure generates the probability distribution of the random variables whose realizations we see. In this case there may be partial identification of structural features and functionals of those features. We study methods for characterizing identified sets of economic magnitudes.

Identifying correspondences link features of models to features of the probability distributions that structures admitted by models generate. Replacing features of distributions by estimates in these identifying correspondences leads to **analogue estimators** of structural features. We introduce and motivate OLS, GLS, IV and other estimators in this way. In doing so we provide a review of basic econometric estimators. When there is partial identification, it is necessary to consider estimators of and inference on identified sets.

The way in which sources of stochastic variation feature in models and the properties of stochastic variation are important considerations. Human tastes and experience are diverse so it is interesting to consider how multidimensional latent variables can be included in econometric models. This requires consideration of nonlinear models. We consider the nature of stochastic variation in econometric models.

Economic processes deliver many endogenous outcomes and there is interest in understanding the structural equations, correspondences, and inequalities whose simultaneous satisfaction produces values of those outcomes. Econometric models allowing endogenous explanatory variables are a central topic in the course and we study a variety of identifying models.

We show how in models involving a scalar non-additive latent variable, **quantile regression functions** can carry information about structural features.

A leading case in which non-additive error models naturally arise is when **discrete responses** are observed. We study parametric and semiparametric models for discrete outcomes and

consider their identifying power when there is endogeneity. In these cases, most of the point identifying models used in practice are either complete models or employ conditional independence restrictions under which endogeneity is removed once there is conditioning on certain observed variables. We will consider how these restrictions can be relaxed using partially identifying instrumental variable models.

We study a general class of **incomplete models** in which multiple values of unobservable variables can deliver a particular value of endogenous variables and provide examples of such models in the study of auctions, empirical IO and labour markets. We give a result on the identifying power of these, generally partially identifying, models.

Extensive notes and references will be distributed via the course Moodle page where you can find links to some key articles, exercises, solutions, slides and videos.

The topics we will study, and the planned order of attack are as follows.

Lecture 1. Structural econometrics. The purpose and history of identification analysis. Identification – concepts, definitions. Identification in parametric and semiparametric models. Econometric estimators as analogue estimators, OLS 2SLS, ML, GMM.

Lecture 2. Complete models. Linear simultaneous equations models. Reduced form. The classical analysis of identification, rank and order conditions. Triangular models, linear, nonlinear and nonparametric, additive errors. Control functions.

Lecture 3. Quantiles and quantile regression. Triangular models with nonadditive errors. Conditional independence restrictions. Treatment effect models. Incomplete models. Instrumental variable models with additive unobservables.

Lecture 4. Incomplete models. Instrumental variable (IV) models. IV methods in additive error models, flexible parametric specifications and ill-posedness. Nonparametric IV models with non-additive errors. Application to modelling returns to schooling.

Lecture 5. Generalized IV (GIV) models. GIV models for discrete outcomes. Partial identification. Application to modelling female labour force participation.

Andrew Chesher, September 25th, 2024