Lecture 1: Introduction to structural estimation

Short course "Dynamic programming and structural estimation"

Fedor Iskhakov
Australian National University

Higher School of Economics
St. Petersburg
January-February 2020

About the short course

Dynamic programming and structural estimation

- Numerical techniques for dynamic modeling in economics
- Two classic models:
  1. Rust model of engine replacement, discrete choice
  2. Deaton model of consumption and savings, continuous choice
- Overview of estimation methods:
  1. A battery of standard methods for discrete choice
  2. Method of simulated moments for continuous choice
- Lectures + Lab sessions for hands-on experience

About the lecturer

Fedor Iskhakov | Исхаков Фёдор Валентинович
Professor of Economics
PhD 2009, к.э.н. 2006

Australian National University (Canberra)

- applied micro-econometrician and computational economist
- structural estimation of (primarily) dynamic models
- single agent, equilibrium and models of strategic interaction
- applications in labor economics and industrial organization (IO)

fedor.iskh.ru
fediskhakov@gmail.com

Schedule

1. Lectures
   - On Tuesdays Jan 21, 28, Feb 4, 9:00-12:00 (2 пары)
   - Location: this room
2. Labs/Seminars
   - On Thursdays Jan 23, 30, Feb 6
   - First week 9:00-12:00 for **group A**, 13:00-16:00 for **group B**
   - Location: computer lab

Technical aspects

- Programming language: Python
- Course materials: GitHub
  https://github.com/CompEconCourse
- Personal computers/laptops are encouraged

1. Register on GitHub.com if not yet
2. Set up local Git software and a GUI for it
   (for example, SourceTree or GitHub Desktop)
3. Set up computing environment following
   https://python.quantecon.org/getting_started.html

Will make sure to get everybody on the same page on Thursday

Student introductions

1. Your name
2. Your research interests
3. Your prior experience with computer programming

Road map for the rest of the day

Lecture 1:

1. What is dynamic structural estimation?
2. Reminder of static discrete choice model
3. Dynamic discrete choice:
   Rust engine replacement model of Harold Zurcher

Next in Lecture 2:

4. MLE estimation of dynamic discrete choice model

Structural econometrics

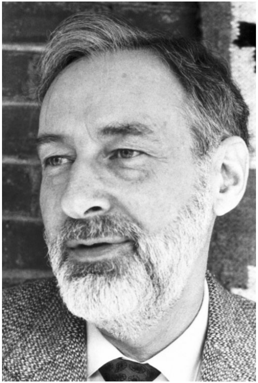
Econometrics is a branch of statistics focused on economic measurement, prediction, and the development and testing of economic theories

- The term "structural" arose in the 1940s
- Due to Trygve Haavelmo, Tjalling Koopmans, Jacob Marschak and other founders of the Cowles Commission.
- Appears in the 1949 *Econometrica* paper by Koopmans "Identification Problems in Economic Model Construction"

> Statistical inference, from observations to economic behavior parameters, can be made in two steps: inference from the observations to the parameters of the assumed joint distribution of the observations, and inference from that distribution to the parameters of the structural equations describing economic behavior.

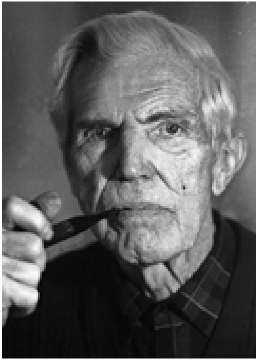
Founding fathers

Tjalling Koopmans



Jacob Marschak

Trygve Haavelmo



Reduced-form vs structure

- We use the term *reduced-form model* to denote an econometric model that may not necessarily be derived from, or tightly linked to a particular economic model or theory
- Whereas a *structural model* is an econometric model explicitly derived from an economic model or theory and we want to estimate/infer the parameters to test/evaluate the theory
- Example: linear supply and demand simultaneous equations model

$$\begin{array}{lll} \text{structural model} & Y\Gamma = XB + U & \theta = (\Gamma, B, \Sigma) \\ \text{reduced-form model} & Y = X\Pi + V & \Pi = B\Gamma^{-1} \end{array}$$

- θ are the structural parameters, (Π, Ω) are the reduced-form parameters. $\Pi = B\Gamma^{-1}$ and $\text{cov}(V) = \Omega = \Gamma'^{-1}\Sigma\Gamma^{-1}$.

The identification problem

- The reduced-form equation $Y = X\Pi + V$ holds regardless whether the underlying structural model is correct or not.
- Π and Ω parameters can be estimated by OLS.
- *Identification problem:* when it is possible to invert the mapping from the reduced-form parameters to the structural parameters?
$$\theta = (\Gamma, B, \Sigma) \longleftrightarrow (\Pi, \Omega)$$
- In general, there are more structural parameters than reduced-form parameters and we need additional *exclusion restrictions* and/or *cross-equation restrictions* to identify θ.
- Early work at the Cowles Foundation determined necessary and sufficient conditions for identification and related these to maximum likelihood estimation and instrumental variables estimation, i.e. that the restrictions required for identification give rise to the instrumental variables required for estimation.

Why do structural estimation?

- *Counterfactual analyses and predictions* aka *policy simulations*
- The general idea of the Cowles Commission approach to inference is that *structure is policy-invariant* which do not change if *policies* change (e.g. taxes, regulations, other government/firm policies).
- Let θ represent the structure (e.g. preference, production function parameters, etc). and let π represent policy variables under the *status quo*
- Suppose there are "endogenous variables" y (firm outputs or prices or consumer choices/demand) and observable "exogenous variables" x (i.e. weather) as well as ϵ unobserved exogenous variabes. The economic model predicts a reduced form relationship that could be written as

$$y = f(x, \epsilon, \theta, \pi)$$

Why do structural estimation?

- Think of reduced-form estimation as the need to specify a distribution of unobservables $g(\epsilon)$ and just use flexible non-parametric methods to estimate the conditional probability of y given x:

$$P(y \in B|x) = \int I\{f(x, \epsilon, \theta, \pi) \in B\} g(\epsilon) d\epsilon$$

- While $P(y|x)$ may provide a good forecast of behavior/endogenous outcomes under the *status quo* policy π it will not provide a good prediction under a counterfactual policy π'.
- Further it will not tell us a great deal about *welfare* and *distributional consequences*.
- Structural estimation and identification attempts to invert and use the reduced form $P(y|x)$ to map back to the *structure* $[\theta, g, f]$ separately from the policy variables π.

Policy forecasting using structural models

- The economic model can enable us to assess welfare effects and distributional effects of a policy change from π to π' and results in a counterfactual prediction

$$P(y \in B|x, \theta, g, \pi') = \int I\{f(x, \epsilon, \theta, \pi') \in B\} g(\epsilon) d\epsilon$$

- This is the critical advantage of structural estimation: the ability to do counterfactual forecasts and simulations!
- Contrast this with reduced-form methods. In general, the reduced form will shift when the policy shifts, so

$$P(y \in B|x, \theta, g, \pi) \neq P(y \in B|x, \theta, g, \pi')$$

and thus the reduced-form relationship we estimate under the *status quo* is not likely to hold, and provide accurate forecasts of behavior/outcomes under an alternative policy π'

6

Static Discrete Choice Models

- Problem: a decision maker in state x chooses an alternative d from a finite set $D(x)$ of possible alternatives to maximize $u(x, d)$
- Economic approach: agent uses a *decision rule* $d^*(x) = argmax_{d \in D(x)} u(x, d)$
- If we know the person's state x and utility function $u(x, d)$ and choice set $D(x)$, then we can *perfectly predict the choice* $d^*(x)$

Static Discrete Choice Models

- *Probabilistic choice theory* The choice d^* is not perfectly predictable because it depends on *private information* ϵ that the agent observes that we (as the econometrician) do not observe
- Decision rule is $d^*(x, \epsilon) = argmax_{d \in D(x)}[u(x, d) + \epsilon(d)]$ which is a *random variable* that is not perfectly predictable.
  RUM = *Random Utility Model*
- Define the *conditional choice probability*
  $P(d|x) = \text{Prob}\{d^*(x, \epsilon) = d|x\} = \int_\epsilon I\{d^*(x, \epsilon) = d\} q(\epsilon|x)$

Probabilistic choice theory

- Early work by psychologists such as Thurstone *The measurement of values* and Duncan Luce *Individual choice behavior* both published in 1959
- $P(d|D,x)$ conditional probability of choosing alternative $d \in D(x)$ for a subject with observed characteristics x (x might also capture *attributes* of the alternatives in $D(x)$)

Independence from Irrelevant Alterantives (IIA) Axiom

If $B(x) \subset D(x)$ and $d \in B(x)$ then

$$P(d|D,x) = P(d|B,x)P(B|D,x), \text{ where}$$

$$P(B|D,x) = \sum_{d \in B(x)} P(d|D,x)$$

The multinomial logit model

Luce's Theorem

If IIA holds, then there exist non-negative weights $v(x, d)$, $d \in D(x)$ such that

$$P(d|D, x) = \frac{exp\{v(x, d)\}}{\sum_{d' \in D(x)} \exp\{v(x, d')\}}$$

- These choice probabilities are known as the multinomial logit model (MNL)

The IIA property of the MNL model

- Alternative form of (IIA) Axiom:
  The odds of selecting one alternative d relative to another alternative d' are independent of the composition of the choice set $D(x)$ and the utilities of alternatives other than d and d'
- The MNL model satisfies the IIA Axiom:

$$\frac{P(d|x)}{P(d'|x)} = \frac{\exp\{u(x,d)\}}{\exp\{u(x,d')\}}$$

- Luce (1959) proved that the MNL is the *only choice probability* that satisfies the IIA axiom.
- Marshak (1960) provided a RUM interpretation of the MNL model using the Type 1 extreme value (Gumbel) family of probability distributions.

Testable implications of the IIA property

- Red bus/blue bus paradox (Debreu) Suppose a person can commute to work by "blue bus" b or car c and $u(x, b) = u(x, c)$. MNL $\Rightarrow P(b|x) = P(c|x) = .5$
- Now suppose we do an experiment introducing a third artificial alternative red bus r: $u(x, r) = u(x, b) = u(x, c)$. Then the MNL predicts $P(r|x) = P(b|x) = P(c|x) = 1/3$. Yet, a more reasonable prediction is that the choice probabilities should be $P(r|x) = P(b|x) = .25$, $P(c|x) = .5$.
- The independence of the random errors $\{\epsilon(r), \epsilon(b), \epsilon(c)\}$ is questioned here. The observed attributes of two bus alternatives are essentially identical and thus we would expect that $\epsilon(b)$ and $\epsilon(r)$ are either identical or highly correlated.
- $\exists$ extensions of the MNL model to deal with this problem.

Econometric estimation of MNL models

- Data on individual states and choices, $\{(x_i, d_i) | i = 1, \ldots, N\}$
- McFadden showed how to estimate unknown parameter θ by *maximum likelihood* (MLE)

$$\hat{\theta}_N = \underset{\theta}{argmax} \log \left(\prod_{i=1}^{N} \frac{\exp\{u(x_i, d_i, \theta)\}}{\sum_{d' \in D(x_i)} \exp\{u(x_i, d', \theta)\}} \right)$$

- McFadden showed that $\log(L_N(\theta))$ is *globally concave in* θ when $u(x, d, \theta)$ is *linear in parameters*
- Nowadays there is no reason to restrict $u(x, d, \theta)$ to be linear: modern "hill climbing" algorithms can also estimate $\hat{\theta}_N$ for more general specifications. The main complication is that $\log(L_N(\theta))$ is not necessarily concave in θ, creating the possibility of *multiple local optima*.

CCPs and Hotz-Miller Inversion

CCPs = Conditional Choice Probabilities = $P(d|x)$, probability an individual with characteristics x chooses alternative $d \in D(x)$.

- CCPs can be recovered from the data $\{(x_i, d_i)|i = 1, \ldots, N\}$ using non-parametric estimation or flexible models, and thus can be thought of as "observed"
- The CCPs may or may not be consistent with the underlying model of static discrete choice
- Is it possible to find some utility function to "rationalize" any CCP $P(d|x)$?

CCPs and Hotz-Miller Inversion

- The answer is YES
- Define $u(x, d) = \log(P(d|x))$. Then for MNL model we have

$$
\begin{aligned}
P(d|x) &= \frac{\exp\{\log(P(d|x))\}}{\sum_{d' \in D(x)} \exp\{\log(P(d'|x))\}} \\
&= \frac{P(d|x)}{\sum_{d' \in D(x)} P(d'|x)} = P(d|x).
\end{aligned}
$$

Hotz-Miller Inversion Theorem

There is a one-to-one mapping between CCPs and normalized utility functions

$$
\{P(d|x) | d \in D(x)\} \longleftrightarrow \{u(x, d) - u(x, d_0) | d \in D(x)\}
$$

where $d_0 \in D(x)$ is some fixed element in the choice set.

Implications for identification

- The CCP $\{P(d|x)|d \in D(x), x \in X\}$ is the *reduced-form object* — it can be estimated regardless of whether the underlying RUM holds or not, and hence treated as "known"
- The utility function $\{u(x,d)|d \in D(x), x \in X\}$ and the distribution of random components $\{q(\epsilon|x)|x \in X\}$ are the *structural objects*
- The RUM model is identified if there is a one-to-one mapping between the structure and reduced form

$$\{P(d|x)|d \in D(x), x \in X\} \longleftrightarrow \{[u(x,d), q(\epsilon|x)]|d \in D(x), x \in X\}$$

- Unfortunately, Hotz-Miller Inversion Theorem tell us that the MNL is *non-parametrically unidentified* — i.e. without any further restrictions, for any distribution $q(\epsilon|x)$ of random utility components, we can find at least one utility function $\{u(x,d)|d \in D(x), x \in X\}$ that rationalizes the CCPs, $P(d|x)$.

Overview of where we are going

- Add dynamics. Fundamental tool: Dynamic Programming
- Structural estimation of dynamic models: dynamic discrete choice model
- We use *model* as laboratory for analyzing *causality* and *counterfactual predictions*
- DP is powerful due to its flexibility and breadth: it provides a framework to study decision making over time, under uncertainty, and can accommodate *learning,* strategic interactions between agents (game theory) and market interactions (equilibrium theory).