Lecture 1 Toolbox/review: GMM, ML, SML, SMM, indirect inference

Ciprian Domnisoru

Aalto University

This course: Structural models

- Data generating process. Assume data was generated by a model
- Model has a grounding in economics. Explicit assumptions about economic actors' objectives, choices, economic environment and information set:
  - Individuals solving dynamic optimization problems
  - individuals and firms make consumption and production choices
  - individuals optimally contract with each other
  - individuals or firms play a noncooperative game using equilibrium strategies.
- Interest in recovering the parameters θ that determine choices/ the data generating proces: preferences, technology parameters, financing frictions, agents' bargaining power.
- Typical analysis, after recovering the estimated parameters: model fit (data vs. static model vs. dynamic model); counterfactual analyses; welfare analysis
- Reduced form ("treatment effects"): What is the (causal) effect of X on Y?
- Some similarities with calibration, but focus on **estimation** of parameters

A brief history

- Econometrics: a branch of statistics focused on economic measurement, prediction, and the development and testing of economic theories.
- The term 'structural' is credited to Trygve Haavelmo, Tjalling Koopmans and Jacob Marschak at the Cowles Commission. It appears in the 1949 Econometrica paper by Koopmans "Identification Problems in Economic Model Construction"
- Koopman's essay, *Measurement without theory*, criticized the "decision not to use theories of man's economic behavior, even hypothetically" because the absence of theory "limits the value to economic science and to the maker of policies"

A brief history

- Structure: Marshak (1953) defines a structure "(1) a set of relations describing human behavior and institutions as well as technological laws and involving in general, nonobservable random disturbances and nonobservable random errors in measurement; (2) the joint probability distribution of these random quantities"
- Lucas Critique: "any change in policy will systematically alter the structure of econometric models"
- Structural estimation/inference applies nowadays to all types of models including behavioral models that relax the assumptions of rationality, optimization and equilibrium.
- When not all elements of the structure are identified, we say the model is *partially identified*. Sometimes we cannot *point identify* parameters but can provide bounds on the parameters and say the parameters are *set identified*.

This course: Structural methods

- 1. Specify an economic model where a decisionmaker is assumed to derive value from choices $V(c, \mathbf{x}, \theta)$.
- 2. Derive a statistical model from $V(c, \mathbf{x}, \theta)$ by introducing randomness: mapping between the distribution of choices (or moments of this distribution) and the structural parameters θ.
    1. GMM
    2. MLE
    3. SMM
    4. SMLE
    5. Indirect inference
    6. Two-step methods used to solve dynamic discrete choice methods

This course: Structural methods

1. Lectures 1-4: GMM, MLE, SMM, (SMLE), Indirect inference. Assignment 1
2. Lecture 5: EM algorithm. Assignment 2.
3. Lectures 6-11: Two-step methods used to solve dynamic discrete choice methods. Assignment 3.
4. Last lecture: student presentations.
5. Grades: 33% Homework, 33% Presentation, 33% Replication notes.

SMM paper examples

SMM has been used to estimate

- Models of job search (Flinn and Mabli, 2008)
- Educational and occupational choices (Adda et al., 2011, 2013)
- Household choices (Flinn and Del Boca, 2012)
- Stochastic volatility models (Andersen et al., 2002; Raknerud and Skare, 2012)
- Dynamic stochastic general equilibrium models (Ruge-Murcia, 2012)

Dynamic discrete choice paper examples

- Occupational Choice (Keane and Wolpin, JPE 1997)
- Retirement (Rust and Phelan, ECMA 1997)
- Brand choice and advertising (Erdem and Keane, MaScience 1996)
- Choice of college major (Arcidiacono, JoE 2004)
- Individual migration decisions (Kennan and Walker, ECMA 2011)
- High school attendance and work decisions (Eckstein and Wolpin, ECMA 1999)
- Sales and dynamics of consumer inventory behavior (Hendel and Nevo, ECMA 2006)
- Advertising, learning, and consumer choice in experience good markets (Ackerberg, IER 2003)
- Route choice models (Fosgerau et al, Transp. Res. B)
- Fertility and labor supply decisions (Francesconi, JoLE 2002)
- Residential and Work-location choice (Buchinsky et al, ECMA 2015)

Methods used to estimate dynamic discrete choice methods

- Rust (1987): MLE using Nested-Fixed Point Algorithm (NFXP)
- Hotz and Miller (1993): CCP estimator - (two step estimator)
- Keane and Wolpin (1994): Simulation and interpolation
- Rust (1997): Randomization algorithm (breaks curse of dimensionality)
- Aguirregabiria and Mira (2002): Nested Pseudo Likelihood (NPL).
- Bajari, Benkard and Levin (2007): Two step-minimum distance (equilibrium inequalities).
- Arcidiacono Miller (2002): CCP with unobserved heterogeneity (EM Algorithm).
- Norets (2009): Bayesian Estimation (allows for serial correlation in ε)
- Su and Judd (2012): MLE using constrained optimization (MPEC)

Labor and MLE Application: “An empirical investigation of the Option Value of College Enrollment", Kevin Stange

Preliminaries: Heckman, Lochner and Todd (2006)

- In common usage, the coefficient on schooling in a regression of log earnings on years of schooling is often called a rate of return. In fact, it is a price of schooling from a hedonic market wage equation. It is a growth rate of market earnings with years of schooling and not an internal rate of return measure [..] coefficient on schooling as a rate of return derives from a model by Becker and Chiswick (1966). It was popularized and estimated by Mincer (1974) and is now called the Mincer model.
- We explore the importance of Mincer's implicit stationarity assumptions, which allowed him to use cross-section experience - earnings profiles as guides to the life cycle earnings of persons. In recent time periods, life cycle earnings -education-experience profiles differ across cohorts. Thus cross-sections are no longer useful guides to the life cycle earnings or schooling returns of any particular individual.

“An empirical investigation of the Option Value of College Enrollment", Kevin Stange

- Common to analyze educational decisions as investments (Becker, Mincer): weighing short-term costs against future benefits, choosing the level of schooling that maximizes welfare.
- In a static framework, hard to rationalize why dropout rates are so high, given the high premiums for graduating from college in the US.
- Structural model in which schooling decisions are sequential, academic ability is learned through grades.

“An empirical investigation of the Option Value of College Enrollment", Kevin Stange

- Solves for the parameters of the model. Simulates educational outcomes and welfare using the model and compares this to the counterfactual wherein individuals commit to an educational outcome before enrolling in college: difference between the two indicates the “option value" of college enrollment: individuals are not bound to remain enrolled in college following an adverse shock revealed after enrollment.
- Models that ignore the **option value** will necessarily understate the incentive to enroll and mischaracterize the social desirability of college dropout.
- Counterfactuals relevant for school tracking policies: “forcing students to committ ex ante makes educational outcomes more polarized by background and reduces welfare, particularly for students at the margin. [...] weighted against any efficiency gains resulting from greater specialization."

Model

- Estimates a **dynamic discrete choice** model of college attendance.
- In the first two periods, choosing among no college, 2-year college, and 4-year college.
- In the second two periods, choosing among no college and 4-year college
- No college is a **terminal choice** (no more decisions are made)
- In this model, the simplification is that school-leaving is irreversible. The labor market is an **absorbing state.**
- It is assumed that people know their expected discounted income at each of the 5 "exit nodes"
- Uncertainty on:
  1. **preference shocks** ("factors — getting ill, having a parent lose a job, having a winning football team — **that are not expected to persist over time**")
  2. grades
  3. wages (specific realization is unknown ex ante)
- Individuals learn about ability through course grades.

Model

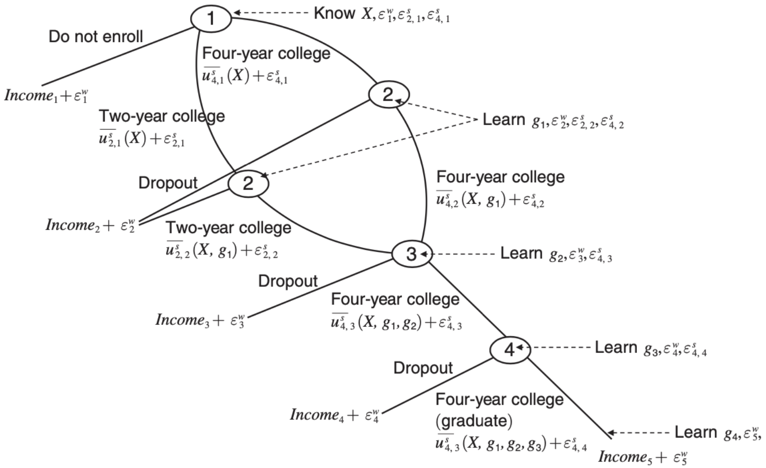

FIGURE 4. FULL EMPIRICAL DYNAMIC MODEL

Learning

- A represents ability and affects the cost of going through college.
- $\bar{u}_t^s = \alpha_m + \alpha_A E_t[A] - \alpha_d Dist_t - Tuition_t + \alpha_{2year} 1[2year]$
- Students learn A through their college grades, with a prior as of the end of high school that depends on HSGPA, AFQT, and parental education
- $E_1[A] = \gamma_m + \gamma_g HSGPA + \gamma_t AFQT + \gamma_p ParentED$
- α_m and γ_m are random effects assumed to have a discrete bivariate distribution with M=3 points of support. These random effects denote "types" : type is known to the individual, but unknown to the econometrician.
- $E_t[A] = k_t E_1[A] + (1 - k_t) GPA_{t-1}$, where:
- $GPA_{t-1} = \frac{1}{t-1} \sum_{k=1}^{t-1} g_k$ is the cumulative GPA up to t-1.

Solving the model

- Solving the model consists of finding the value functions for each alternative at each point in time.
- At node 4 (end of third year), the student has observed g1, g2, g3 and has to compare:

$$\begin{aligned}
\text{value of dropout} &= V_4^w = Income_4 + \epsilon_4^w \\
\text{value of continuing} &= V_4^s = \overline{u}_4^s + \epsilon_4^s + Income_5 + E[\epsilon_5^w] \\
&= \alpha_m + \alpha_A E_4[A|GPA_3, X] - \alpha_d Dist_4 - Tuition_4 + \epsilon_4^s + Income_5 + c.
\end{aligned}$$

- At node 3 (end of 2nd year), the student has observed g2 and has to compare

$$\begin{aligned}
\text{value of dropout} &= V_3^w = Income_3 + \epsilon_3^w \\
\text{value of continuing} &= V_3^s = \overline{u}_3^s + \epsilon_3^s + E_3 E \max[V_4^w, V_4^s|GPA_3, X] \\
&= \alpha_m + \alpha_A E_3[A|GPA_2, X] - \alpha_d Dist_3 - Tuition_3 \\
&\quad + E_3 E \max[V_4^w, V_4^s|GPA_3, X] + \epsilon_3^s
\end{aligned}$$

Solving the model

- Since g_3 is unknown at node 3, $E_3 Emax[V_4^w, V_4^s \mid GPA_3, X]$ requires a probability distribution for GPA_3 conditional on GPA_2:

$$E_3 E\max[V_4^w, V_4^s | GPA_3, X] = \int E\max[V_4^w, V_4^s | GPA_3, X]\pi(GPA_3, X | GPA_2, X)$$

  where π gives the transition probabilities from GPA2 to GPA3.
- To simplify, Stange assumes that cumulative GPA lies on a discrete grid. Assuming g3 is normally distributed allows him to use simple expressions for the probabilites of alternative values of GPA3 conditional on GPA2.
- After similar calculations at nodes 2 and 1, the model delivers a likelihood for the joint distribution of GPA outcomes and dropout decisions.

Solving the model

- Under the assumption that the preference shocks are not serially correlated and are drawn from an extreme value distribution with location parameter zero and scale parameter τ, this expectation has a closed-form representation:

$$E_t\left[\max(V^w_{i,t+1}, V^s_{i,t+1})\right] = \int \left[\tau\lambda + \tau \log\left\{\exp\left(\frac{1}{\tau}\overline{V}^s_{i,t+1}(g_{i,t})\right) + \exp\left(\frac{1}{\tau}\overline{V}^w_{i,t+1}\right)\right\}\right] \cdot \Pi\left(dg_{i,t} | X_i, \{g_{i,1}...g_{i,t-1}\}\right)$$

- In general, given the Type I extreme value assumption on the error terms, $V(s) = \int max_c[v(s,c) + \varepsilon(c)]G(\varepsilon)d\varepsilon$ can be expressed as $V(s) = ln[\sum_{j=1}^{J} exp(v_j)] + \gamma$, where $\gamma = 0.577216$ is Euler's constant.

Solving the model

- Without heterogeneity:

$$\text{Period 1: } L_i^1 = \Pr(S_{i,2,1} = 1)^{S_{i,2,1}} \Pr(S_{i,4,1} = 1)^{S_{i,4,1}} \Pr(S_{i,1} = 0)^{1-S_{i,1}};$$

$$\text{Periods 2 to 4: } L_i^2 = \prod_{t=2}^{4} \Pr(S_{i,t} = 1)^{S_{i,t}} \Pr(S_{i,t} = 0)^{1-S_{i,t}};$$

$$\text{Grades: } L_i^3 = \prod_{t=1}^{4} \Pr(g_{i,t}),$$

- Thanks to the Type I extreme value, these probabilities have logit closed form solutions.
- With heterogeneity, integrate likelihood contribuion for an individual i over the joint distribution of the random effects γ_m and $\alpha_{m,j}$. This distribution is assumed to have M mass points, each with probability p_m (which also needs to be estimated): $L_i = \sum_{m=1}^{M} p_m L_{im}$

Coefficient estimates

TABLE 1—ESTIMATES OF STRUCTURAL PARAMETERS

	No learning		Learning	
	One type	Three types	One type	Three types
	(1)	(2)	(3)	(4)
Utility parameters				
Constant (2yr)	−2.911	−4.346	−2.569	−3.187
	(0.150)	(0.442)	(0.121)	(0.378)
Constant (4yr)	−2.588	−3.765	−2.220	−2.845
	(0.137)	(0.391)	(0.105)	(0.332)
$E[A_i]$	0.707	1.242	0.591	1.009
	(0.049)	(0.161)	(0.039)	(0.154)
Distance (100)	0.121	0.277	0.139	0.220
	(0.034)	(0.074)	(0.034)	(0.065)
τ	0.511	0.780	0.513	0.642
	(0.022)	(0.074)	(0.023)	(0.070)
Grade parameters				
Constant (gpa)	1.192	0.835	0.802	0.659
	(0.056)	(0.102)	(0.072)	(0.087)
High school GPA	0.383	0.394	0.436	0.523
	(0.019)	(0.026)	(0.025)	(0.029)
AFQT	0.411	0.702	0.581	0.695
	(0.039)	(0.082)	(0.057)	(0.072)
Parent BA	0.206	0.297	0.281	0.336
	(0.017)	(0.033)	(0.026)	(0.033)
$E[A \mid X]$ period 2			0.482	0.528
			(0.030)	(0.034)
$E[A \mid X]$ period 3			0.319	0.343
			(0.038)	(0.046)
$E[A \mid X]$ period 4			0.188	0.206
			(0.046)	(0.057)
SD (GPA)	0.645	0.478		
	(0.008)	(0.007)		
SD (GPA year 1)			0.657	0.617
			(0.014)	(0.016)
SD (GPA year 2)			0.534	0.521
			(0.013)	(0.013)
SD (GPA year 3)			0.526	0.520
			(0.014)	(0.014)
SD (GPA year 4)			0.547	0.545
			(0.016)	(0.016)

(*Continued*)

Coefficient estimates

Table 1—Estimates of Structural Parameters (*Continued*)

	No learning		Learning	
	One type	Three types	One type	Three types
Type-specific parameters				
Constant (gpa)-T2		0.634		0.256
		(0.024)		(0.088)
Constant (2yr)-T2		0.124		0.603
		(0.290)		(0.196)
Constant (4yr)-T2		−0.244		−2.387
		(0.185)		(0.519)
Probability T2		0.174		0.075
		(0.022)		(0.011)
Constant (gpa)-T3		−0.889		−0.536
		(0.042)		(0.067)
Constant (2yr)-T3		−0.201		−1.646
		(0.245)		(0.496)
Constant (4yr)-T3		0.271		−0.441
		(0.088)		(0.120)
Probability T3		0.359		0.625
		(0.030)		(0.040)
Observations	2,055	2,055	2,055	2,055
lnL (total)	6,328	5,844	5,888	5,719

Notes: Utility is in units of \$100,000. Income specification (1) from online Appendix Table B1 was used to generate counterfactual income estimates. Standard errors (in parentheses) were calculated from the inverse of the numerical Hessian. Specifications (3) and (4) uses 17 GPA categories for Emax approximation $(0.0, 0.25, 0.50, \ldots, 4.0)$.

Counterfactuals

- Educational choices and welfare are simulated under alternative assumptions about individuals? information set (including first best, which implies knowledge of all future shocks)
- Since the option value is a highly nonlinear and complicated function of the parameters, the author relies on simulations to compute the confidence intervals.
  1. Each observation is replicated 100 times
  2. For each simulated observation, preference and grade shocks are drawn from the EV(1) and normal distributions.
  3. An unobserved type is assigned based on the estimated probabilities.
  4. Optimal choices for each individual are calculated by utility comparisons, incorporating shocks.

Welfare analysis: Option value results

Table 2—Estimated Option Value, by Expected Academic Ability

$E[A_i \mid X_i]$	Option value ($1,000)		Option value as percent of total value of enrollment in dynamic scenario		Option value as percent of welfare loss between full information and static scenarios	
	Estimate	90 percent confidence interval	Estimate	90 percent confidence interval	Estimate	90 percent confidence interval
1.0	0.3	[0.21, 1.67]	7%	[4%, 19%]	3%	[2%, 12%]
1.5	3.2	[2.76, 5.05]	25%	[22%, 29%]	12%	[11%, 17%]
2.0	16.8	[13.39, 20.48]	35%	[31%, 38%]	27%	[25%, 29%]
2.5	25.0	[18.05, 31.2]	19%	[17%, 23%]	32%	[29%, 34%]
3.0	16.6	[12.67, 21.27]	6%	[5%, 8%]	28%	[25%, 31%]
3.5	12.2	[4.68, 19.16]	3%	[1%, 5%]	24%	[9%, 32%]
All	14.9	[11.43, 18.09]	14%	[12%, 16%]	27%	[25%, 29%]

Notes: For a given parameter vector, option value is calculated as the average welfare difference between the static and dynamic scenarios when the type, shocks, and choices of each observation are simulated 100 times. Confidence intervals are computed by performing this option value calculation for 200 different draws of the parameter vector from its estimated distribution.