Structural Estimation 1: The basics of Maximum Likelihood

Christine Braun

What is structural estimation?

- People have different definitions
  - Estimation of preference and technology parameters in a maximizing model
  - Structural parameters of a simultaneous equations model
- Main assumption: model parameters are policy invariant

Why take a structural approach?

- Counterfactuals
  - estimate the effects of new policy interventions
- Decompositions
  - analyze the effect of each part of an equilibrium change

Two very important points

1. Structural estimation and non-structural (reduced form) both have advantages and disadvantages $\Rightarrow$ they should be seen as compliments rather than substitutes!

    - "The Best of Both Worlds: Combining RCTs with Structural Modeling" by Todd and Wolpin, forthcoming JEL

2. You can write a good or bad paper of each kind! Neither approach is a guarantee for success.

Pros and Cons

	Pros	Cons
Structural	emphasis on external validity	tends to be more complicated lots of parameters
	mapping from parameters to implication is clearer	computationally costly
	forces you to think about DGP	more assumptions
Non-structural	emphasis on internal validity mapping from data to estimates is clearer	often silent on optimal policy often silent on mechanisms

Typical process for structural work

1. Pinpoint a policy question to be answered
   - when you are first starting research this is the hardest part! but don't give up :)
2. Write down the simplest model needed to simulate policy
3. Think about what data you will need to identify parameters
4. Estimate model
5. Simulate counterfactual policy

Estimation Methods

- Maximum Likelihood Estimation
- Generalized Method of Moments
- Indirect Inference

Maximum Likelihood

- y_i: outcome variable observed for each individual i (unemployment duration, wage)
- x_i: vector of observable characteristics of individual i (age, sex, education, ...)
- θ_0: vector of true parameters
- $\{y_i, x_i\}_{i=1}^N$: i.i.d. draws from $f(y_i|x_i;\theta_0)$
    - where does f come from?

Likelihood Function

- The likelihood function is the likelihood that the parameter vector θ produced the observed data

$$L(\theta; \{y_i, x_i\}_{i=1}^N) = f_{Y_1|x_1,\ldots,Y_N|x_N}(y_1, \ldots, y_n|x_1, \ldots, x_n; \theta)$$

$$L(\theta; \{y_i, x_i\}_{i=1}^N) = \prod_{i=1}^{N} f(y_i|x_i; \theta)$$

$$\mathcal{L}(\theta; \{y_i, x_i\}_{i=1}^N) = \sum_{i=1}^{N} \ln f(y_i|x_i; \theta)$$

Maximum Likelihood Estimator

- The maximum likelihood estimator (MLE) is

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \underbrace{\frac{1}{n} \sum_{i=1}^{N} \ln f(y_i | x_i; \theta)}_{Q_n(\theta)}$$

- Regularity Conditions

  A1. $Q_n(\theta) \to Q_0(\theta)$

  A2. Θ is compact and $\theta \in \Theta$

  A3. $\ln f(y_i | x_i; \theta)$ is continuous in θ

  A4. $Q_0(\theta)$ is uniquely maximized at θ_0:

  - This is identification and must be argued

- Under A1-A4 $\hat{\theta} \to \theta_0$

Asymptotic Distribution of MLE

- The Asymptotic Distribution of MLE is

$$\sqrt(n)(\hat{\theta} - \theta_0) \rightarrow N\big(0, (-H)^{-1}\big)$$

- $H = E[\nabla_{\theta\theta} \ln f(y_i|x_i;\theta)]$ is the hessian matrix

- $(-H)^{-1}$ is the variance-covariance matrix

- $\sqrt{diag[(-H)^{-1}]}$ is the vector of standard errors

A Simple Example: Model

- Model
  - unemployed workers receive job offers at rate λ
  - job offers are drawn from an exogenous wage distribution $G(w)$
  - jobs get destroyed at rate δ
  - workers discount at rate r

A Simple Example: Model

- Value functions and reservation wage

$$rU = b + \lambda \int_{w_R}^{\infty} E(w) - U \; dG(w)$$

$$rE(w) = w + \delta[U - E(w)]$$

$$w_R = b + \frac{\lambda}{r+\delta} \int_{w_R} w - w_R \; dG(w)$$

A Simple Example: Model

- What are the parameters of the model that we want to estimate?
  - λ: arrival rate of job offers
  - b: unemployment flow utility
  - r: discount rate
  - δ: separation rate
  - $G(w)$: wage offer distribution
    - we need more assumptions about the distribution
    - $G(w) \sim \ln N(\mu, \sigma)$

A Simple Example: Estimation

- We will work through two different data sets
  - Data 1: observe duration of unemployment spells
  - Data 2: Data 1 + employed worker's wages
- What can we identify given each data set?
  - estimation using each data set
- **Next Lecture:** Extensions and limitations

A Simple Example: Data 1

To get likelihood function we need to figure out $f(y_i|x_i;\theta)$

- y_i is what we observe
  - t_i, duration of unemployment spell
- x_i: observable characteristics, generally things y_i is a function of, apart from model parameters
  - here we don't have observables
- θ: the model's parameters that y_i is a function of
  - what determines unemployment duration in the model?

A Simple Example: Data 1

- In the model the rate at which workers find acceptable jobs is

$$\lambda[1 - G(w_R)]$$

- If λ is Poisson, then $\lambda[1 - G(w_R)]$ is Poisson

- Let $N(t)$ be the number of arrivals in interval length t

$$P(N(t) = n) = \frac{(\lambda[1 - G(w_R)])^n}{n!} \exp(-\lambda[1 - G(w_R)]t)$$

A Simple Example: Data 1

- What is the probability of observing an unemployment spell that lasted t_i periods?
  - Find the cdf

$$\begin{aligned} F(t_i; \lambda, w_R, F_w) &= P(T < t_i; \lambda, w_R, \mu, \sigma) \\ &= 1 - P(N(t_i) = 0) \\ F(t_i; \lambda, w_R, \mu, \sigma) &= 1 - \exp(-\lambda[1 - G(w_R; \mu, \sigma)]t_i) \end{aligned}$$

  - Take derivative to find pdf

$$f(t_i; \lambda, w_R, \mu, \sigma) = \lambda[1 - G(w_R; \mu, \sigma)] \exp(-\lambda[1 - G(w_R; \mu, \sigma)]t_i)$$

- Durations follow an exponential distribution

A Simple Example: Data 1

- What parameters is the likelihood a function of?
  - λ, $G(w; \mu, \sigma)$, and w_R
  - but $w_R(\lambda, b, r, \delta, \mu, \sigma)$
  - all of them!
- What can we identify with our data?

A Simple Example: Data 1

- What parameters is the likelihood a function of?
  - λ, G, and w_R
  - but $w_R(\lambda, b, r, \delta, \mu, \sigma)$
  - all of them!
- What can we identify with our data?
  - since the parameters all enter through one function (the hazard function)

$$h = \lambda[1 - G(w_R)]$$

    we can't identify the model primitives, only h.
- When is this enough?

A Simple Example: Data 1

- The likelihood function

$$\begin{aligned}\mathcal{L}(h; \{t_i i\}_{i=1}^N) &= \sum_{i=1}^N \ln f(t_i; h) \\ &= \sum_{i=1}^N \ln[h \exp(-h t_i)] \\ &= N \ln h - h \sum_{i=1}^N t_i\end{aligned}$$

- First order condition $\partial \mathcal{L} / \partial h = 0$ gives

$$\hat{h} = \frac{N}{\sum_{i=1}^N t_i}$$

A Simple Example: Data 1

- In this example it was easy to find max of a single parameter
  - this will rarely be the case!
- Let's take this to matlab to practice for harder cases
- What will we need to do
  1. write a function that takes as inputs h and the data $\{t_i\}$ and outputs the (negative) value of the likelihood function
  2. an algorithm that can numerically minimize a function: Fmincon

Syntax

```
x = fmincon(fun,x0,A,b)
x = fmincon(fun,x0,A,b,Aeq,beq)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon,options)
x = fmincon(problem)
[x,fval] = fmincon( __ )
[x,fval,exitflag,output] = fmincon( __ )
[x,fval,exitflag,output,lambda,grad,hessian] = fmincon( __ )
```

Description

Nonlinear programming solver.

Finds the minimum of a problem specified by

$$\min_x f(x) \text{ such that } \begin{cases} c(x) \leq 0 \\ ceq(x) = 0 \\ A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub, \end{cases}$$

Some useful commands

- Summing up elements in a vector v

$$sum(v)$$

- Element by Element multiplication $(2,1).*(2,1) = (2,1)$

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} .* \begin{bmatrix} 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 8 \\ 15 \end{bmatrix}$$

- Matrix multiplication $(2,1)*(1,2) = (2,2)$

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} * \begin{bmatrix} 4 & 5 \end{bmatrix} = \begin{bmatrix} 8 & 10 \\ 12 & 15 \end{bmatrix}$$

- Same syntax for division / vs ./ and exponents ^vs .^
- Anonymous Functions

$$f = @(x) \;\; x.\hat{}(0.5) + 3$$

$$f([1,4]) = [4,5]$$

Estimation in Matlab

- Use data1.csv
- File 1: SE1_main.m
  - read in data
  - create lower bound for parameter
  - pick an initial guess
  - estimate
- File 2: loglike1.m
  - inputs: parameter, data
  - output: negative log-likelihood value

A Simple Example: Data 1 Answer

- Estimate

$$\hat{h} = 0.3038$$

- Standard error

$$\hat{\sigma}_h = 0.0102$$

- Log-likelihood value

$$logL = -2.1914e + 4$$

Can we identify more than just the hazard rate?

- Data 2 also contains information on workers wages
- Let's solve for the new likelihood function
  1. likelihood of observing unemployment duration
  2. likelihood of observing wages

Likelihood of observing unemployment duration

- Conditional likelihood

$$f(t_i|u;\lambda, w_R, G) = \lambda[1 - G(w_R)] \exp(-\lambda[1 - G(w_R)]t_i)$$

- Un-conditional (full-information) likelihood

$$\begin{aligned} f(t_i, u; \lambda, w_R, \mu, \sigma, \delta) &= f(t_i|u;\lambda, w_R, \mu, \sigma) \times P(u; \delta, \lambda, w_R, \mu, \sigma) \\ &= \frac{\lambda[1 - G(w_R;\mu,\sigma)] \exp(-\lambda[1 - G(w_R;\mu,\sigma)]t_i)}{} \end{aligned}$$

Likelihood of observing unemployment duration

- Conditional likelihood

$$f(t_i|u;\lambda, w_R, G) = \lambda[1 - G(w_R)]\exp(-\lambda[1 - G(w_R)]t_i)$$

- Un-conditional (full-information) likelihood

$$\begin{aligned} f(t_i, u;\lambda, w_R, \mu, \sigma, \delta) &= f(t_i|u;\lambda, w_R, \mu, \sigma) \times P(u;\delta, \lambda, w_R, \mu, \sigma) \\ &= \frac{\lambda[1 - G(w_R;\mu, \sigma)]\exp(-\lambda[1 - G(w_R;\mu, \sigma)]t_i)\delta}{\delta + \lambda[1 - G(w_R;\mu, \sigma)]} \end{aligned}$$

Likelihood of observing an employed worker at w

- Conditional likelihood

$$f(w_i|e; w_R, \mu, \sigma) = \frac{g(w; \mu, \sigma)}{1 - G(w_r; \mu, \sigma)}$$

- Un-conditional (full-information) likelihood

$$\begin{aligned} f(w_i, e; \lambda, w_R, \mu, \sigma, \delta) &= f(w_i|e; \lambda, w_R, \mu, \sigma) \times P(e; \delta, \lambda, w_R, \mu, \sigma) \\ &= \frac{g(w; \mu, \sigma)}{[1 - G(w_R; \mu, \sigma)]} \end{aligned}$$

Likelihood of observing an employed worker at w

- Conditional likelihood

$$f(w_i | e; w_R, \mu, \sigma) = \frac{g(w; \mu, \sigma)}{1 - G(w_r; \mu, \sigma)}$$

- Un-conditional (full-information) likelihood

$$\begin{aligned} f(w_i, e; \lambda, w_R, \mu, \sigma, \delta) &= f(w_i | e; \lambda, w_R, \mu, \sigma) \times P(e; \delta, \lambda, w_R, \mu, \sigma) \\ &= \frac{g(w; \mu, \sigma)\lambda[1 - G(w_R; \mu, \sigma)]}{[1 - G(w_R; \mu, \sigma)]\{\delta + \lambda[1 - G(w_R; \mu, \sigma)]\}} \end{aligned}$$

Log-Likelihood Function

- Dummy variable needed

$$\mathbb{1}(u_i) = 1 \;\text{ if person } i \text{ is unemployed}$$

- Likelihood function

$$L(\lambda, w_R, \mu, \sigma, \delta; w_i, t_i, u_i) = \prod_{i=1}^{N} f(w_i, e; \lambda, w_R, \mu, \sigma, \delta)^{1-\mathbb{1}(u_i)} \times f(t_i, u; \lambda, w_R, \mu, \sigma, \delta)^{\mathbb{1}(u_i)}$$

- Log-Likelihood function

$$\mathcal{L}(\lambda, w_R, \mu, \sigma, \delta; w_i, t_i, u_i) = \sum_{i=1}^{N} [1 - \mathbb{1}(u_i)] \ln f(w_i, e; \lambda, w_R, \mu, \sigma, \delta) + \mathbb{1}(u_i) \ln f(t_i, u; \lambda, w_R, \mu, \sigma, \delta)$$

Log-Likelihood Function

- Log-Likelihood Simplified

$$\mathcal{L} = \sum_{i=1}^{N} \ln \lambda + \ln[1 - G(w_R; \mu, \sigma)] - \ln\{\delta + \lambda[1 - G(w_R; \mu, \sigma)]\}$$
$$+ \mathbb{1}(u_i) \ln \delta - \mathbb{1}(u_i)\lambda[1 - G(w_R; \mu, \sigma)]t_i$$
$$+[1 - (u_i)] \ln g(w; \mu, \sigma) - [1 - (u_i)] \ln[1 - G(w_R; \mu, \sigma)]$$

- What parameters are we trying to identify

$$\lambda, r, \; b, \; \mu, \; \sigma, \; \delta$$

Identification

- λ, μ, σ, δ show up "explicitly" in likelihood function
- r and b show up only through a scalar constant w_R
- Flinn & Heckman (1982) propose

$$\hat{w}_R = \min_i w_i$$

the estimate of the reservation wage is the lowest wage observed. This is an extreme value estimator and converges to the true lower bound at rate N.

Identification

- Using $\hat{w}_R$ maximize log-likelihood function to get $\hat{\lambda}$, $\hat{\mu}$, $\hat{\sigma}$, $\hat{\delta}$
- Then given consistant estimators we have

$$\hat{w}_R = b + \frac{\hat{\lambda}}{r + \hat{\delta}} \int_{\hat{w}_R}^{\infty} w - \hat{w}_R \; dG(w; \hat{\mu}, \hat{\sigma})$$

- This gives a locus of points (r, b) consistent with the reservation wage
- The model is fundamentally under-identified
- Typically we fix r and then get an estimate of b

Estimation in Matlab

- Use data2.csv: [unemployed, duration, wage]
- File 1: SE_main.m
  - read in data
  - find the consistent estimator of w_R
  - create lower bound and initial guess
  - estimate
  - solve for b using $r = 0.05$
- File 2: loglike2.m
  - inputs: parameters, data, $\hat{w}_R$
  - output: negative log-likelihood value

A Simple Example: Data 2 Answer

- Reservation wage

$$\hat{w}_R = 8.8253$$

- Estimates and standard errors from likelihood estimation

Parameter	Estimate	Standard Error
λ	0.2820	0.0127
μ	2.2339	0.0119
σ	0.3794	0.0043
δ	0.0225	0.0009

- Log-likelihood value

$$logL = -2.7602e + 4$$

- Estimate of b

$$\hat{b} = 0.6497$$