Introducing the Estimators with Examples

Lecture Notes

Overview

- Before diving into statistical theory, we introduce estimators through examples from our prototype models
- Three workhorse methods:
  1. **Maximum Likelihood (MLE)**
  2. **Generalized Method of Moments (GMM)**
  3. **Minimum Distance**

Extremum Estimators

- All three methods are **extremum estimators**: estimators characterized as solutions to optimization problems
- **Definition**: $\hat{\theta}$ is an extremum estimator if:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} Q_N(\theta)$$

  where $\Theta \subset \mathbb{R}^p$
- The objective function $Q_N(\theta)$ differs across methods but the optimization structure is common

Key Properties to Establish

- For each estimation approach, we want to establish:
  1. **Consistency**: Does $\hat{\theta} \to \theta_0$ as $N \to \infty$?
     - Does our estimate approach the "true" parameters as we collect more data?
  2. **Inference**: What is the sampling distribution of $\hat{\theta}$?
     - How uncertain are we about our estimates?
     - Can we place reasonable bounds on the correct answer?

The Generalized Roy Model

Model Setup

- Selection equation (linear form):

$$D = \mathbf{1}\{\gamma_0 + \gamma_1 X + \gamma_2 Z - V \geq 0\}$$

- Outcome equations:

$$Y_D = \beta_{D,0} + \beta_{D,1} X + U_D$$

- Distributional assumption: $V \sim \mathcal{N}(0, 1)$

Two-Step Estimator

- Our identification argument suggested a **two-step** procedure:

Step 1: Estimate Selection by Maximum Likelihood

- The probit MLE:

$$\hat{\gamma} = \arg\max_{\gamma} \frac{1}{N} \sum_{n=1}^{N} D_n \log(\Phi(\mathbf{w}_n \gamma)) + (1 - D_n) \log(1 - \Phi(\mathbf{w}_n \gamma))$$

- Where $\mathbf{w}_n = [1,\ X_n,\ Z_n]$
- $\Phi(\cdot)$ is the standard normal CDF

Step 2: Estimate Outcomes with Selection Correction

- Estimate outcome equations by OLS using a **selection correction** (Heckman correction)
- Uses $\hat{\gamma}$ from first stage to construct correction term

Theoretical Challenge

- This is a **two-step estimator**: second stage relies on first-stage estimates
- We need to develop theory for how first-stage estimation error affects second-stage inference
- Standard errors must account for the two-step nature

The Search Model

Setup with Measurement Error

- Observe wages with small *known* measurement error:

$$\log(W^o_{n,t}) = \log(W_{n,t}) + \zeta_{n,t}$$

- Where $\zeta_{n,t} \sim \mathcal{N}(0, \sigma^2_\zeta)$ and $\sigma_\zeta = 0.05$
- Parametric assumption: W is log-normally distributed with mean μ and variance σ^2_W

Parameters to Estimate

- Directly estimated:

$$\theta = (\mu, \sigma^2_W, h, \delta, w^*)$$

- Derived parameters: λ and b (inverted using model structure)
  - b comes from the reservation wage equation

The Log-Likelihood

- Let $X_n = (W^o, t_u, E)$ denote the data (observed wage, unemployment duration, employment status)
- Log-likelihood of a single observation:

$$l(X;\theta) = E \times \int f_{W|W>w^*}(\log(W^o) - \zeta)\phi(\zeta;\sigma_\zeta)d\zeta + (1-E) \times [\log(h) + t_u \log(1-h)]$$

- The truncated wage density:

$$f_{W|W>w^*}(w) = \frac{\phi(w;\sigma_W)}{1 - \Phi(w^*/\sigma_W)}$$

The MLE

- Maximum likelihood estimator:

$$\hat{\theta} = \arg\max_\theta \frac{1}{N} \sum_n l(X_n;\theta)$$

Derived Parameter Estimates

- MLE estimate of job arrival rate:

$$\hat{\lambda} = \hat{h}/(1 - \widehat{F}_{W|W>w^*}(\hat{w}^*))$$

- MLE estimate of unemployment value:

$$\hat{b} = w^* - \frac{\hat{\lambda}}{1 - \beta(1-\hat{\delta})} \int_{\hat{w}^*} (1 - \widehat{F}_{W|W>w^*}(w))dw$$

Theoretical Questions

- Need to characterize asymptotic properties of both:
  - Direct estimates $\hat{\theta}$
  - Derived estimates $\hat{b}$ and $\hat{\lambda}$ (functions of $\hat{\theta}$)
- The Delta Method will be key for derived parameters

Implementation Notes

- When passing data to the likelihood, use **typed data structures** (NamedTuples) rather than DataFrames
- DataFrames have untyped columns which can dramatically slow optimization
- Use transformations (logit, exp) to enforce parameter constraints during optimization

---

The Labor Supply Model

Setup

- Cross-sectional observations: $(W_n, H_n, C_n, \mathbf{z}_n)$

- Labor supply equation:

$$\log(H) = \mu - \psi \log(W) - \psi\sigma \log(C) + \epsilon$$

- Instruments $\mathbf{z}_n$ move consumption and labor supply
- Key assumption: $\mathbb{E}[\epsilon\ |\mathbf{z}] = 0$

Moment Conditions

- The exclusion restriction implies:

$$\mathbb{E}[\epsilon \mathbf{z}] = 0$$

- Parameters: $\theta = (\mu, \sigma, \psi)$

Sample Moment

- Define the sample moment:

$$g_N(\theta) = \frac{1}{N} \sum_N \left(\log(H_n) - \mu - \psi \log(W) - \psi\sigma \log(C)\right) \mathbf{z}_n$$

The GMM Estimator

- GMM minimizes a quadratic form in the moments:

$$\hat{\theta} = \arg\min_\theta g_N(\theta)' \mathbf{W}_N g_N(\theta)$$

- $\mathbf{W}_N$ is a symmetric, positive definite **weighting matrix**
- Since system is linear, this is a quadratic minimization with known closed-form solution
- But the theory we develop will be more general (nonlinear GMM)

Key Questions for GMM

1. What choice of $\mathbf{W}$ gives the "best" estimator?
    - Need to define what "best" means (efficient GMM)
2. Can we implement optimal weighting in finite samples?
    - Feasible efficient GMM

The Savings Model

Focus: Income Process Estimation

- Save preference parameter estimation for simulation chapter
- Identification insight: match implied variances and covariances
- When moments outnumber parameters: **over-identification**
- Estimate by **minimum distance**

The Income Process

- AR(1) transitory component:

$$\varepsilon_{n,t+1} = \rho\varepsilon_{n,t} + \eta_{n,t}, \qquad \eta_{n,t} \sim \mathcal{N}(0, \sigma_\eta^2)$$

- Extended specification with permanent heterogeneity:

$$\log(y_{n,t}) = \mu_t + \alpha_n + \varepsilon_{n,t}$$

- Where $\alpha_n \sim (0, \sigma_\alpha^2)$ is an individual fixed effect
- Initial condition: $\varepsilon_0 = 0$

Implied Dynamics

- With $\varepsilon_0 = 0$:

$$\varepsilon_t = \sum_{s=1}^{t} \rho^{t-s}\eta_s$$

- Parameters to estimate: $\theta = (\rho, \sigma_\alpha^2, \sigma_\eta^2)$

Moment Restrictions

- Define $\epsilon = \log(y) - \mu_t = \alpha + \varepsilon$
- Variance at age t:

$$\mathbb{V}[\epsilon_t] = \sigma_\alpha^2 + \frac{(1-\rho^{2(t-1)})}{1-\rho^2}\sigma_\eta^2$$

- Recursive variance relationship:

$$\mathbb{V}[\epsilon_{t+1}] = \sigma_\alpha^2 + \rho^2\mathbb{V}[\varepsilon_t] + \sigma_\eta^2$$

- Autocovariance:

$$\mathbb{C}(\epsilon_t, \epsilon_{t+s}) = \sigma_\alpha^2 + \rho^s\mathbb{V}[\epsilon_t]$$

Two Alternative Moment Vectors

Option 1: Variance Profile

$$\mathbf{v} = [\mathbb{V}[\epsilon_1],\ \mathbb{V}[\epsilon_2],\ \ldots,\ \mathbb{V}[\epsilon_T]]'$$

Option 2: Variances and Covariances

$$\mathbf{c} = [\mathbb{V}[\epsilon_t],\ \mathbb{V}[\epsilon_{t+1}],\ \mathbb{C}(\epsilon_t, \epsilon_{t+1}),\ \ldots,\ \mathbb{C}(\epsilon_t, \epsilon_{t+K})]'$$

The Minimum Distance Estimator

- Let $\mathbf{v}(\theta)$ be model-implied moments
- Let $\hat{\mathbf{v}}$ be sample estimates
- Minimum distance estimator:

$$\hat{\theta} = \arg\min_{\theta} \left(\hat{\mathbf{v}} - \mathbf{v}(\theta)\right)' \mathbf{W} \left(\hat{\mathbf{v}} - \mathbf{v}(\theta)\right)$$

- $\mathbf{W}$ is a positive definite weighting matrix

Key Question

- How does choice of $\mathbf{W}$ affect precision?
- Is there an "optimal" choice? (Same question as GMM)

STOP FOR DISCUSSION We identify income risk parameters by matching variance growth with age—attributing all growth to income risk. • Are you comfortable with this identification strategy? • Suppose we use this model to evaluate welfare gains from redistributive taxes. How important will these parameters be for how agents value social insurance? • What other sources could explain variance growth with age?

The Entry-Exit Model

Key Insight

- Choice probability $p(x, a, a')$ is directly observable in the data
- Encodes information about underlying payoff parameters

Payoff Specification

- Payoff from operating ($d = 1$):

$$u_1(x, a, d') = \phi_0 + \phi_1 x - \phi_2 d' - \phi_3(1 - a)$$

- Payoff from not operating ($d = 0$):

$$u_0(x, a) = \phi_4 a$$

- Parameters: $\phi = (\phi_0, \phi_1, \phi_2, \phi_3, \phi_4)$

Estimation Approach 1: Minimum Distance

Model-Implied Choice Probabilities

- For each state (x, a, a'), the model implies:

$$p(x, a, a'; \phi, \beta) = \frac{\exp(v_1(x, a, a'; \phi, \beta))}{\exp(v_0(x, a, a'; \phi, \beta)) + \exp(v_1(x, a, a'; \phi, \beta))}$$

- Where v_0 and v_1 are choice-specific value functions (depend on equilibrium solution)

Data Structure

- Cross-section: $(X_m, D_m, A_m, A'_m)_{m=1}^M$ for M markets
- X takes discrete values in support $\mathcal{X}$

Empirical Choice Frequencies

- For each state (x, a, a'):

$$\hat{p}(x, a, a') = \frac{\sum_m D_m \mathbf{1}\{X_m = x, A_m = a, A'_m = a'\}}{\sum_m \mathbf{1}\{X_m = x, A_m = a, A'_m = a'\}}$$

The Estimator

- Stack probabilities: $\mathbf{p}(\theta)$ model-implied, $\hat{\mathbf{p}}$ empirical
- Minimum distance estimator:

$$\hat{\theta} = \arg\min_{\theta}(\hat{\mathbf{p}} - \mathbf{p}(\theta))'\mathbf{W}_N(\hat{\mathbf{p}} - \mathbf{p}(\theta))$$

Estimation Approach 2: GMM

The Residual

- Define:

$$\xi_m = D_m - p(X_m, A_m, A'_m; \phi, \beta)$$

- Key property: $\mathbb{E}[\xi_m | X_m, A_m, A'_m] = 0$ at true parameters

Moment Conditions

- This implies:

$$\mathbb{E}\left[(D_m - p(X_m, A_m, A'_m; \phi, \beta)) \cdot \mathbf{z}_m\right] = 0$$

- Natural instrument choices—functions of the state:

$$\mathbf{z}_m = [1,\ X_m,\ A_m,\ A'_m,\ X_m \cdot A_m]'$$

The GMM Estimator

- Sample moment:

$$g_M(\phi) = \frac{1}{M}\sum_m \left(D_m - p(X_m, A_m, A'_m; \phi, \beta)\right)\mathbf{z}_m$$

- GMM estimator:

$$\hat{\phi} = \arg\min_{\phi} g_M(\phi)'\mathbf{W}_M g_M(\phi)$$

Summary

Model	Estimation Method	Key Features
Generalized Roy	Two-step MLE	Selection correction; two-stage inference
Search	MLE	Measurement error integration; derived parameters
Labor Supply	GMM	Instrumental variables; weighting matrix choice
Savings	Minimum Distance	Variance matching; over-identification
Entry-Exit	MD or GMM	Equilibrium constraints; choice probability matching

Common Themes

- All are **extremum estimators**: optimize an objective function
- All require theory for:
  - **Consistency**: $\hat{\theta} \to \theta_0$
  - **Asymptotic distribution**: $\sqrt{N}(\hat{\theta} - \theta_0) \to ?$
- Several involve a **weighting matrix W**—optimal choice matters
- Some have **multi-step** or **derived parameter** complications