Lecture 12

Estimation of Heterogeneous Agent Models

ECO 521: Advanced Macroeconomics I

Benjamin Moll

Princeton University, Fall 2016

Plan

1. Background: bringing heterogeneous agent models to data – the state of the literature
2. Alvarez-Parra, Posch & Wang
3. Other papers estimating heterogeneneous agent models

Bringing HA Models to Data – State of Literature

- 99 percent of papers: calibration
- Remaining 1 percent: some form of estimation, usually GMM
- Usual calibration strategy:
  - take some parameters from literature (e.g. Frisch elasticity of labor supply, say from Chetty et al survey = 0.5-1)
  - calibrate others internally to hit some aggregate moments (e.g. discount rate ρ to match $K/Y = 3$)
  - see e.g. Section 1.5 of these lecture notes:

    `http://www.econ.nyu.edu/user/violante/NYUTeaching/Macrotheory/Spring14/LectureNotes/lecture7_14.pdf`

Calibration vs Estimation

- Big debate in 90s
  - Hansen-Heckman "The Empirical Foundations of Calibration"
  - Browning-Hansen-Heckman "Micro Data and GE Models"
  - Sargent interview `http://www.tomsargent.com/research/SargentinterviewMD.pdf`
- Things to note:
  - calibration and estimation can be similar: (well-done) calibration is basically GMM without standard errors
  - perhaps more relevant distinction: full-information (e.g. MLE) vs limited-information methods (e.g. GMM, calibration)?
  - my impression: main reason for not estimating is computational cost (having s.e.'s better than not having them)
- What may calibration miss?
  - standard errors
  - metric for judging model's goodness of fit
  - metric for comparing different models (model selection)

Parra-Alvarez, Posch & Wang

Parra-Alvarez, Posch and Wang

- Maximum likelihood estimation of Aiyagari-Bewley-Huggett model
    - current version: mainly discuss identification issues
- So far: no data – though will ultimately use SCF

A prototypical heterogeneous agent model

Competitive Stationary Equilibrium

- The optimal behavior of households is characterized by the system of HJB equations:

$$\begin{aligned}
\rho V(a_t, e_l) &= u(c(a_t, e_l)) + V_a(ra_t + we_l - c(a_t, e_l)) + \phi_{hl}(V(a_t, e_h) - V(a_t, e_l)) \\
\rho V(a_t, e_h) &= u(c(a_t, e_h)) + V_a(ra_t + we_h - c(a_t, e_h)) + \phi_{lh}(V(a_t, e_l) - V(a_t, e_h))
\end{aligned}$$

- The optimal behavior of firms is given by:

$$r = \alpha K^{\alpha-1} L^{1-\alpha}, \quad w = (1-\alpha) K^{\alpha} L^{-\alpha}$$

where

$$K = \sum_{e_t \in \{e_l, e_h\}} \int_{\underline{a}}^{\infty} a_t g(a_t, e_t) da_t, \quad L = \sum_{e_t \in \{e_l, e_h\}} \int_{\underline{a}}^{\infty} e_t g(a_t, e_t) da_t$$

which link the dynamic and randomness that occurs at the micro level with the deterministic behavior at the macro level.

A prototypical heterogeneous agent model

Distribution of endowments and wealth

- The subdensities $g\left(a_t, e_t\right)$ correspond to the solution to the (time-invariant) Fokker-Planck equations:

$$
\begin{aligned}
0 &= -\frac{\partial}{\partial a_t}\left[s\left(a_t, e_l\right) g\left(a_t, e_l\right)\right] - \phi_{hl} g\left(a_t, e_l\right) + \phi_{lh} g\left(a_t, e_h\right) \\
0 &= -\frac{\partial}{\partial a_t}\left[s\left(a_t, e_h\right) g\left(a_t, e_h\right)\right] - \phi_{lh} g\left(a_t, e_h\right) + \phi_{hl} g\left(a_t, e_l\right).
\end{aligned}
$$

- The (unconditional) density of wealth is defined as:

$$
g\left(a_t\right) = g\left(a_t, e_l\right) + g\left(a_t, e_h\right)
$$

where the subdensities $g\left(a_t, e_t\right) = g\left(a_t \mid e_t\right) p\left(e_t\right)$ and $p\left(e_t\right)$ is the stationary distribution of a given efficiency level:

$$
p\left(e_t\right) = \frac{1}{\phi_1\left(e_t\right) + \phi_2\left(e_t\right)}\left[\phi_1\left(e_t\right) \mathbf{1}_{\{e_t = e_l\}} + \phi_2\left(e_t\right) \mathbf{1}_{\{e_t = e_h\}}\right].
$$

MLE: A Simple Example to Refresh your Memories

- Suppose we know that the wealth distribution is Pareto with some tail parameter θ

$$g(a) = \theta a^{-\theta-1}, \qquad a \geq 1$$

- We don't know θ but we have an i.i.d sample a_i, $i = 1, .., N$

- Let's use the wealth sample to estimate θ by maximum likelihood

- Follow standard steps of MLE

  1. form likelihood function $\mathcal{L}(\theta | a_1, ..., a_N)$
  2. take logs
  3. find $\hat{\theta}$ that maximizes log-likelihood function, $\log \mathcal{L}(\theta | a_1, ..., a_N)$

MLE: A Simple Example to Refresh your Memories

- Step 1: form likelihood function
  - for each θ, how likely it is to have observed the data that we did in fact observe? Answer:

$$\mathcal{L}(\theta | a_1, \ldots, a_N) = \prod_{i=1}^{N} g(a_i) = \prod_{i=1}^{N} \theta a_i^{-\theta-1}$$

- Step 2: take logs

$$\log \mathcal{L}(\theta | a_1, \ldots, a_N) = \sum_{i=1}^{N} \log\left(\theta a_i^{-\theta-1}\right) = N \log \theta - (\theta+1) \sum_{i=1}^{N} \log a_i$$

- Step 3: maximize log-likelihood function

$$\max_{\theta} \; \log \mathcal{L}(\theta | a_1, \ldots, a_N) = \max_{\theta} \left\{ N \log \theta - (\theta+1) \sum_{i=1}^{N} \log a_i \right\}$$

$$\text{FOC}: \quad \frac{N}{\theta} = \sum_{i=1}^{N} \log a_i \quad \Rightarrow \quad \hat{\theta} = \left(\frac{1}{N} \sum_{i=1}^{N} \log a_i \right)^{-1}$$

MLE: A Simple Example to Refresh your Memories

- ML estimator makes intuitive sense, in particular

$$\text{tail inequality} = \frac{1}{\hat{\theta}} = \frac{1}{N}\sum_{i=1}^{N} \log a_i$$

- Another intuition: $x := \log a \sim \theta e^{-\theta x}$, i.e. exponential distribution

- Mean of exponential distribution is

$$\mathbb{E}[x] = \frac{1}{\theta}$$

- ML estimator of θ is based on sample analogue

$$\frac{1}{\hat{\theta}} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\sum_{i=1}^{N} \log a_i$$

Likelihood Function in PPW

Let $\boldsymbol{a} = [a_1, \ldots, a_N]$ be a sample of N i.i.d observations on individual wealth and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{\mathcal{K}}$ a vector of structural parameters. Recall that the p.d.f of wealth can be computed as:

$$g\left(a_n \mid \boldsymbol{\theta}\right) = g\left(a_n, e_l \mid \boldsymbol{\theta}\right) + g\left(a_n, e_h \mid \boldsymbol{\theta}\right), \quad \forall n = 1, \ldots, N.$$

The log-likelihood function for a given sample is give by:

$$\mathcal{L}_N\left(\boldsymbol{\theta} \mid \boldsymbol{a}\right) = \sum_{n=1}^{N} \log g\left(a_n \mid \boldsymbol{\theta}\right),$$

whereas the maximum likelihood (ML) estimator is defined as:

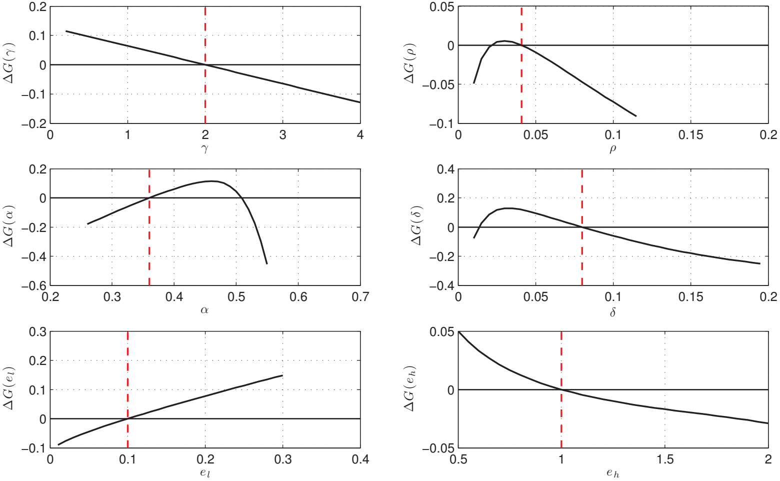
$$\hat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\max} \quad \mathcal{L}_N\left(\boldsymbol{\theta} \mid \boldsymbol{a}\right).$$

Population parameters ($\boldsymbol{\theta}_0$)

$\boldsymbol{\theta}_0 = \{\gamma, \rho, \alpha, \delta, e_h, e_l, \phi_{lh}, \phi_{hl}\}$	
Relative risk aversion, γ	2.0000
Rate of time preference, ρ	0.0410
Capital share in production, α	0.3600
Depreciation rate of capital, δ	0.0800
Endowment of high efficiency, e_h	1.0000
Endowment of low efficiency, e_l	0.1000
Demotion rate, ϕ_{lh}	0.6697
Promotion rate, ϕ_{hl}	4.4644

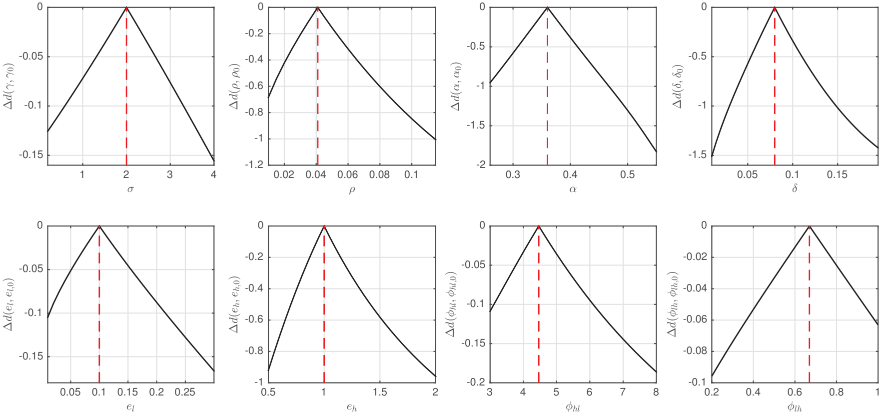
In the model, time is measured in years and parameters should be interpreted accordingly. Demotion and promotion rates computed from Hugget (1993) who reports $p\,(e_h \mid e_l) = 0.5$ and $p\,(e_h \mid e_h) = 0.925$ in a model with six periods per year.

Identification with GMM?



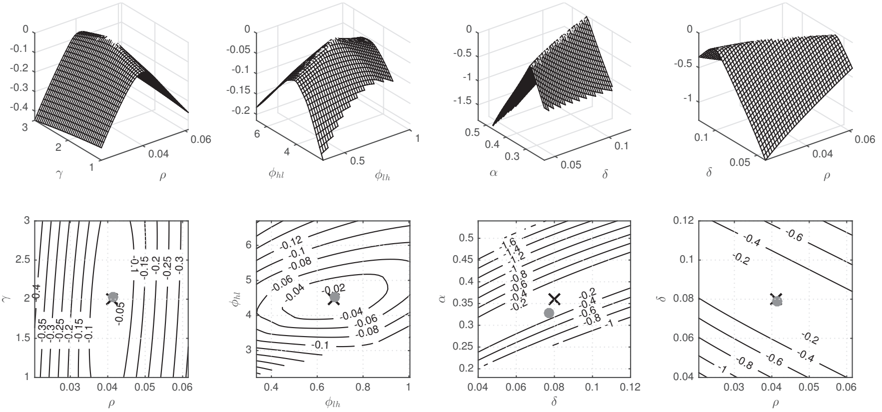
- PPW: (ρ, α, δ) not identified from GMM targeting wealth Gini
- Is this really an argument against GMM?

Population Identification



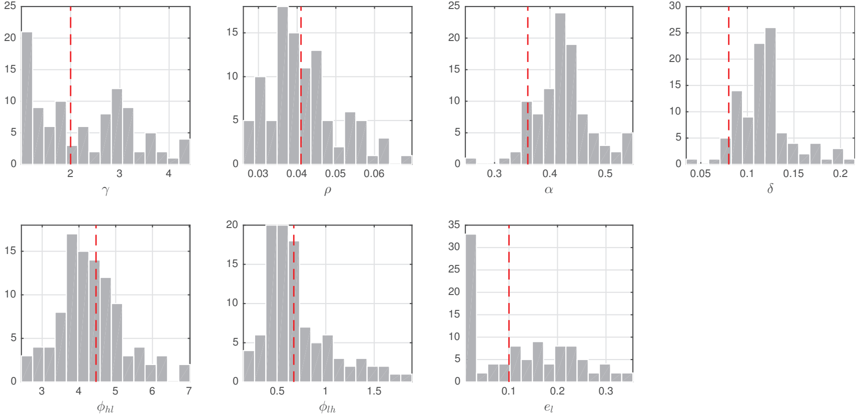
Distance function $d\left(g\left(a \mid \theta\right), g\left(a \mid \theta_0\right)\right)$**.** The graph shows the percentage deviation of the L_1 distance criterion as a function of the parameter space. The population values for the structural parameters, $\boldsymbol{\theta}_0$, are represented by the dotted vertical line.

Population Identification



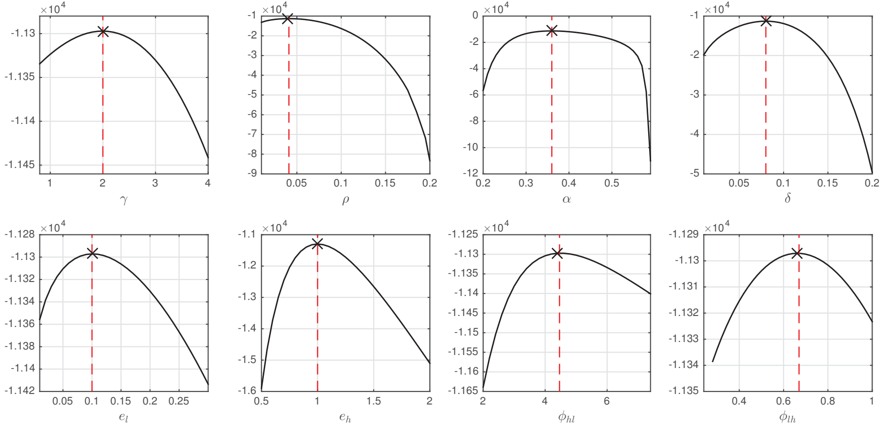
Distance surface. The graph shows the percentage deviation of the L_1 distance function for selected parameters as a function of the parameter space (top) and its respective contour plot (bottom). "$\times$" locates the true parameter values, and ● the maximum of the distance surface, $\tilde{\boldsymbol{\theta}} = \arg\max d\left(\boldsymbol{\theta}, \boldsymbol{\theta}_0\right)$.

Small sample estimates



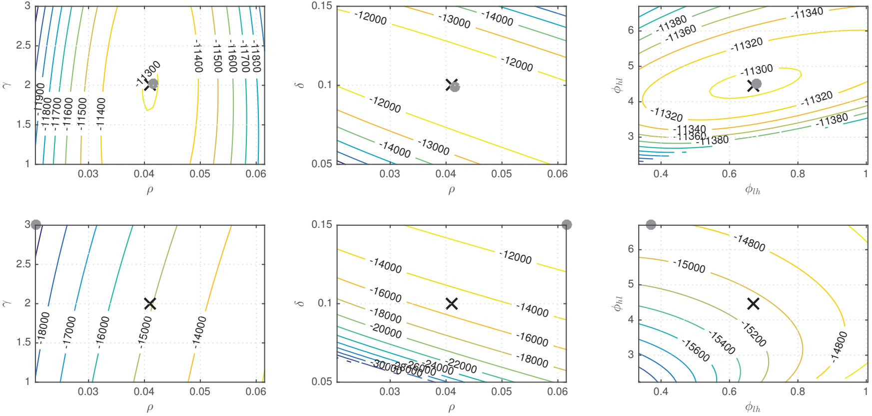
Finite sample distribution of parameter estimates. The graph plots the histogram of estimated parameters across $M = 100$ random samples of size $N = 20000$ generated from the true data generating process. The vertical line denotes the true parameter value.

Sample identification



Individual log-likelihood profile. The graph shows the log-likelihood function $\mathcal{L}(\theta \mid \mathbf{a})$ for each $\theta \in \boldsymbol{\theta}$ along a neighborhood of the its true value for $N = 5000$. The vertical line denotes the true parameter value and '$\times$' marks the maximum value of the log-likelihood profile.

Calibration and estimation



Log-likelihood profile contours for selected parameters. Log-likelihood function $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{a})$ for selected parameters and $N = 5000$. Top contours as before. In bottom contours α and δ are miscalibrated to 0.5 and 0.1 respectively. "$\times$" locates the true parameter values, and ● the maximum of the log-likelihood.

Thoughts on Parra-Alvarez, Posch & Wang

1. PPW include parameters of income process $(\phi_{hl}, \phi_{lh}, e_h, e_l)$ in parameter vector θ to be estimated
   - different from typical strategy: estimate using panel data

2. How about using richer income process? Already know ahead of time that two-state model has a counterfactual income distribution

3. PPW use only marginal distribution of wealth $g(a)$. In practice, typically have more data:
   - how about using joint distribution of income, wealth $g(a, z)$?
   - how about using joint distribution of income, wealth, consumption ("3D inequality" e.g. from PSID)
   - how about using panel data. For MLE: transition densities $f(a_{t+s}, z_{t+s}|a_t, z_t)$ also satisfy Kolmogorov Forward equation

Other papers estimating heterogeneous agent models

Estimation of HA Models without Aggregate Shocks

GMM estimation:

- Abbott, Gallipoli, Meghir and Violante (2016), “Education Policy and Intergenerational Transfers in Equilibrium”
- Benhabib, Bisin and Luo (2016), “Wealth distribution and social mobility in the US: A quantitative approach”
- Luo and Mongey (2016) “Student debt and job choice: Wages vs. job satisfaction”

Maximum likelihood estimation:

- any other papers besides Parra-Alvarez, Posch & Wang?

Comments/open questions:

- above studies mostly use moments from cross-sectional data
- what about panel data?

Estimation of HA Models with Aggregate Shocks

- Goal: use time-series variation of higher-order moments in micro data to identify shocks driving business cycles/effects of policies
  - Winberry (2016), "A Toolbox for Solving and Estimating Heterogeneous Agent Macro Models"
  - Mongey and Williams (2016), "Accounting for Firm Dispersion and Business Cycles"
- Typical approach: discrete time Reiter-type perturbation method
  - use projection method (e.g. Chebyshev collocation) to solve steady state, represent distribution as finite dimensional object
  - IMHO much more complicated and "black-boxy" than continuous-time finite-difference method
  - estimate using Bayesian methodss
  - subset of param's estimated internally, subset fixed externally
- Estimation of model with aggregate shocks is also ultimate goal of tools presented in Lecture 9 (Ahn-Kaplan-Moll-Winberry-Wolf)

Thanks for six fun weeks!