12. Simulation-Based Methods

Original version of slides: January 1999

These slides were prepared in 1999. They cover material similar to Chapter 11 of our subsequent book *Microeconometrics: Methods and Applications*, Cambridge University Press, 2005.

- Overview
- Method of Moments
- Maximum Likelihood
- Probit Example
- Complications
- Method of Simulated Moments
- Simulated Maximum Likelihood
- Simulators
- References

1 Overview

Problem:

Estimation may require evaluation of a moment (an integral in the continuous case) for which there is **no analytical solution**.

$$\mathsf{E}[h(u)] = \int h(u) f(u) du, \qquad (1)$$

where

h and u are scalars
u has density $f(\cdot)$.

Solutions:

- Numerical integration
- Monte-carlo integration:

$$\widehat{\mathsf{E}}[h(u)] = \frac{1}{S} \sum_{s=1}^{S} h(u^s),$$

where u^s random draws from $f(\cdot)$.

- More generally use other simulators $\widehat{\mathsf{E}}[h(u)]$.

2 Method of Moments

Suppose theory leads to nonlinear model

$$y_i = h(\mathbf{x}_i, u_i, \boldsymbol{\theta})$$

where

$h(\cdot)$ is a nonlinear function,
$\mathbf{x}_i$ is a regressor vector
u_i is an error term with density $f(\cdot)$
$\boldsymbol{\theta}$ is a $k \times 1$ vector of parameters.

Then the conditional mean is

$$\begin{aligned} \mathsf{E}[y_i|\mathbf{x}_i, \boldsymbol{\theta}] &= \mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})] \\ &= \int h(\mathbf{x}_i, u_i, \boldsymbol{\theta}) f(u_i) du_i. \end{aligned}$$

Nonlinear least squares (NLS) minimizes

$$\sum_{i=1}^{n}(y_i - \mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})])^2.$$

NLS first-order conditions:

$$\sum_{i=1}^{n}\frac{\partial \mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}(y_i - \mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]) = \mathbf{0}.$$

Estimation difficult as no analytical solution

for $\mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]$.

Method of Moments (NLS) is weighted LS estimator of the form

$$\sum_{i=1}^{n} \mathbf{w}_i(y_i - \mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]) = \mathbf{0}, \qquad (2)$$

where
$\mathbf{w}_i = \mathbf{w}(\mathbf{x}_i)$ is specified $q \times 1$ weight.

Advantage over NLS: $\mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]$ appears only once in the first-order conditions and in a linear manner.

[ASIDE: Why called a *method of moments estimator?*.
In the i.i.d. case $\mathsf{E}[y_i - \mu] = 0$
suggests use sample analog $\sum_{i=1}^{n}(y_i - \mu) = 0$.
Leads to $\widehat{\mu} = \bar{y}$.

For $\mathsf{E}[y_i - \mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]] = 0$
use $\sum_{i=1}^{n} \mathbf{w}_i(y_i - \mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]) = \mathbf{0}$,
where $\mathbf{w}_i$ is $q \times 1$ so q equations in the q unknowns $\boldsymbol{\theta}$.

3 Maximum Likelihood

The density $h(y_i, u_i|\mathbf{x}_i, \boldsymbol{\theta})$ may depend on an unobservable u_i.

Integrating out u_i yields density

$$\begin{aligned} f(y_i|\mathbf{x}_i, \boldsymbol{\theta}) &= \mathsf{E}[h(y_i, u_i|\mathbf{x}_i, \boldsymbol{\theta})] \\ &= \int h(y_i, u_i|\mathbf{x}_i, \boldsymbol{\theta}) f(u_i) du_i. \end{aligned}$$

The joint density is $\mathsf{L} = \prod_{i=1}^{n} \mathsf{E}[h(y_i, u_i|\mathbf{x}_i, \boldsymbol{\theta})]$

Maximum likelihood (ML) estimator maximizes the log-likelihood function

$$\mathcal{L} = \ln \mathsf{L} = \sum_{i=1}^{n} \ln \mathsf{E}[h(y_i, u_i|\mathbf{x}_i, \boldsymbol{\theta})].$$

ML first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{1}{\mathsf{E}[h(y_i, u_i|\mathbf{x}_i, \boldsymbol{\theta})]} \frac{\partial \ln \mathsf{E}[h(y_i, u_i|\mathbf{x}_i, \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} = \mathbf{0}.$$

Again problems arise if there is no analytical expression for $\mathsf{E}[h(y_i, u_i|\mathbf{x}_i, \boldsymbol{\theta})]$.

4 Binary Probit

For binary data, y takes values of 0 and 1 only,
with $\Pr[y_i = 1] = p_i$ and $\Pr[y_i = 0] = 1 - p_i$.

The density is $f(y_i) = y_i^{p_i}(1 - y_i)^{1-p_i}$
with $\mathsf{E}[y_i] = p_i$ and $\mathsf{V}[y_i] = p_i(1 - p_i)$.

Probit model lets

$$p_i = \Pr[y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + u_i > 0]$$

where $u_i \sim \mathsf{N}[0, 1]$.

Equivalently

$$p_i = \Pr[u_i \leq \mathbf{x}_i'\boldsymbol{\beta}].$$

Thus

$$\begin{aligned} \mathsf{E}[y_i|\mathbf{x}_i, \boldsymbol{\beta}] &= \mathsf{E}[1(u_i \leq \mathbf{x}_i'\boldsymbol{\beta})] \\ &= \int_{-\infty}^{\infty} 1(u_i \leq \mathbf{x}_i'\boldsymbol{\beta})\phi(u_i)du_i, \end{aligned}$$

where

$1(A)$ equals 1 if event A occurs and equals 0 otherwise

$\phi(\cdot)$ is the standard normal density.

Here

$$h(u_i, \mathbf{x}_i, \boldsymbol{\theta}) = 1(u_i \leq \mathbf{x}_i' \boldsymbol{\beta}).$$

The integral is more commonly rewritten as

$$\int_{-\infty}^{\mathbf{x}_i' \boldsymbol{\beta}} \phi(u_i) du_i = \Phi(\mathbf{x}_i' \boldsymbol{\beta}),$$

where $\Phi(\cdot)$ is the standard normal c.d.f.

There is no analytical solution.
There are good numerical approximations.
We nonetheless consider simulation methods to approximate it.

Reason: In the $m-$choice *multinomial probit model*, the univariate integral becomes an $(m-1)-$variate integral.

Use a simulation version of the method of moments estimator. Estimate $\boldsymbol{\beta}$ as the solution to

$$\sum_{i=1}^{n} \mathbf{w}_i(y_i - \widehat{\mathsf{E}}[1(u_i \leq \mathbf{x}_i'\boldsymbol{\beta})]) = \mathbf{0},$$

where

$$\widehat{\mathsf{E}}[1(u_i \leq \mathbf{x}_i'\boldsymbol{\beta})] = \frac{1}{S}\sum_{s=1}^{S} 1(u_i^s \leq \mathbf{x}_i'\boldsymbol{\beta}),$$

and $u_i^1, ..., u_i^S$ are S independent draws from the standard normal.

Added complication in this, and other discrete choice examples, is that the simulator $\widehat{\mathsf{E}}[1(u_i \leq \mathbf{x}_i'\boldsymbol{\beta}]$ is not continuous.
So small changes in $\boldsymbol{\beta}$ may lead to no change in $\widehat{\mathsf{E}}[1(u_i \leq \mathbf{x}_i'\boldsymbol{\beta})]$.
This leads to the use of alternative simulators.

5 Complications

- As $S \to \infty$, $\widehat{\mathsf{E}}[h(u)] \to \mathsf{E}[h(u)]$.
  One can use $\widehat{\mathsf{E}}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]$ without any other changes.
- In practice computationally too burdensome to let S be large. The calculation needs to be performed for all n observations and at each round of the iterative procedure used to solve for $\widehat{\boldsymbol{\theta}}$.

This leads to attempts to instead use a low value of S.
Problems that then arise are that

- $\widehat{\boldsymbol{\theta}}$ is less precise (larger variance) due to the introduction of noise due to imperfect approximation of $\mathsf{E}[h(u)]$;
- $\widehat{\boldsymbol{\theta}}$ is possibly biased and inconsistent due to bias introduced by imperfect approximation of $\mathsf{E}[h(u)]$.

The solution to the first problem is to introduce better methods to estimate $\mathsf{E}[h(u)]$, using alternative *simulators*. The class of estimators is then called *simulation-based estimators*. The particular simulator (**??**) is called the *frequency simulator*.

One solution to the second problem is to account for the bias.
A second better solution is to chose an estimator where unbiased estimation of $\mathsf{E}[h(u)]$ leads to unbiased estimation of $\boldsymbol{\theta}$. This is clearly the case for the method of moments estimator (2), since $\mathsf{E}[h(u)]$ appears linearly in the first-order conditions. It is not the case for the ML estimator (**??**) since the logarithm of $\mathsf{E}[h(u)]$ is taken.

6 Method of Simulated Moments (MSM)

The *method of simulated moments* (MSM) estimator solves

$$\sum_{i=1}^{n} \mathbf{w}_i(y_i - \widehat{\mathsf{E}}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]) = \mathbf{0}, \qquad (3)$$

where

$\mathbf{w}_i = \mathbf{w}(\mathbf{x}_i)$ is a $q \times 1$ weighting function

$\widehat{\mathsf{E}}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]$ is an unbiased simulator for $\mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]$.

Because $\mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]$ appears in a linear manner, an unbiased simulator for $\mathsf{E}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]$ leads to an unbiased simulator of the estimating equations.

For an unbiased simulator, suppressing dependence on $\mathbf{x}_i$ and $\boldsymbol{\theta}$,

$$\widehat{\mathsf{E}}[h(u_i)] = \mathsf{E}[h(u_i)] + \xi_i, \qquad (4)$$

where ξ_i are i.i.d. with mean 0. So

$$\sum_{i=1}^{n} \mathbf{w}_i \left(y_i - \mathsf{E}[h(u_i, \widehat{\boldsymbol{\theta}})] - \xi_i \right) = \mathbf{0}. \qquad (5)$$

By an exact first-order Taylor series expansion

$$\sum_{i=1}^{n} \mathbf{w}_i \left(y_i - \mathsf{E}[h(u_i, \boldsymbol{\theta}) \right.$$

$$+ \sum_{i=1}^{n} \mathbf{w}_i \left. \frac{\partial \mathsf{E}[h(u_i, \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}^*} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \; = \; \mathbf{0},$$

and hence

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = - \left(\sum_{i=1}^{n} \mathbf{w}_i \left. \frac{\partial \mathsf{E}[h_i]}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}^*} \right)^{-1} \sum_{i=1}^{n} \mathbf{w}_i (y_i - \mathsf{E}[h_i] - \xi_i),$$

where $h_i = \mathsf{E}[h(u_i, \boldsymbol{\theta})]$.

Now

$$\begin{aligned} \mathsf{V}\left[y_i - \mathsf{E}[h_i] - \xi_i\right] \; &= \; \mathsf{V}\left[y_i - \widehat{\mathsf{E}}[h_i]\right] \\ &= \; \mathsf{V}\left[y_i\right] + \mathsf{V}\left[\widehat{\mathsf{E}}[h_i]\right], \end{aligned}$$

as the simulator is independent of y_i.

So $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ asymptotically normal with variance

$$\left(\sum_{i=1}^{n} \mathbf{w}_i \frac{\partial \mathsf{E}[h_i]}{\partial \boldsymbol{\theta}'}\right)^{-1} \times \left(\sum_{i=1}^{n} \mathbf{w}_i \left(\mathsf{V}\left[y_i\right] + \mathsf{V}\left[\widehat{\mathsf{E}}[h_i]\right]\right) \mathbf{w}_i'\right) \times \left(\sum_{i=1}^{n} \mathbf{w}_i \frac{\partial \mathsf{E}[h_i]}{\partial \boldsymbol{\theta}'}\right)^{-1}$$

Now suppose the simulator is the frequency simulator (**??**), so that $\widehat{\mathsf{E}}[h_i(\mathbf{x}_i, u_i, \boldsymbol{\theta})] = \frac{1}{S}\sum_{s=1}^{S} h(\mathbf{x}_i, u_i^s, \boldsymbol{\theta})$. Then

$$\begin{aligned} \mathsf{V}\left[\widehat{\mathsf{E}}[h(\mathbf{x}_i, u_i, \boldsymbol{\theta})]\right] &= \mathsf{V}\left[\frac{1}{S}\sum_{s=1}^{S} h(\mathbf{x}_i, u_i^s, \boldsymbol{\theta})\right] \\ &= \frac{1}{S}\mathsf{V}\left[y_i\right], \end{aligned}$$

since $y_i = h(\mathbf{x}_i, u_i, \boldsymbol{\theta})$. It follows that

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{a}{\sim} \mathsf{N}\left[\mathbf{0}, \left(1 + \frac{1}{S}\right) \left(\sum_{i=1}^{n} \mathbf{w}_i \frac{\partial \mathsf{E}[h_i]}{\partial \boldsymbol{\theta}'}\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{w}_i \mathsf{V}\left[y_i\right.\right.\right.$$

We have obtained the remarkable result that the effect of simulation using the frequency siulator is to inflate the variance of $\widehat{\boldsymbol{\theta}}$ by $\left(1 + \frac{1}{S}\right)$!

7 Simulated Maximum Likelihood (SML)

An alternative estimator is the simulated maximum likelihood estimator.

This is biased.

There is bias correction.

8 Simulators

8.1 Simulation of Random Variables

There are several methods to draw random variables. The inversion method ...

8.2 Direct Simulator

...

8.3 Importance Sampling

....

8.4 Other Samplers

GHK simulator

Decomposition Simulator

Antithetic Acceleration

8.5 Gibbs Sampler

We summarize one of these, the *Gibbs sampler*, in a more general estimation context. Suppose we wish to obtain a random draw from the joint density $f(u_1, u_2)$. This joint density is difficult or impossible to draw from, but the conditional densities $f(u_1|u_2)$ and $f(u_2|u_1)$ are known and tractable. Throughout we suppress dependence of these densities on dependent variable $\mathbf{y}$ and regressors $\mathbf{X}$. The Gibbs sampler alternates between draws from these conditional densities. Denote the s^{th} round estimates of the parameters by (u_{1s}, u_{2s}). Then one simulates u_1 from the conditional density $f(u_1|u_{2s})$, then $u_{2,s+1}$ from the conditional density $f(u_2|u_{1,s+1})$, then $u_{1,s+2}$ from the conditional density $f(u_1|u_{2,s+1})$, and so on. If this process continues for long enough we eventually obtain a draw of (u_1, u_2) from $f(u_1, u_2)$, u_1 from the marginal density $f(u_1)$ and u_2 from the marginal density $f(u_2)$, even though the original starting values will not be draws from $f(u_1, u_2)$. To obtain m draws on (u_1, u_2) one can use m consecutive draws from one Gibbs sampling run, once the run has continued long enough. These values will be correlated, making posterior standard error calculation difficult, and it is better to perform m separate Gibbs sampling runs. In more complicated applications, we may wish to draw, for example, from $f(u_1, u_2, u_3)$, by repeatedly sampling from the three conditional densities. Then analysis is much simpler if the model is specified so that conditional densities are of simpler forms, such as $f(u_1|u_2, u_3) = f(u_1|u_2)$.

9 References

- Gourieroux, C. and A. Monfort (1996), *Simulation-Based Econometric Methods*, Oxford: Oxford University Press.
- Hajivassiliou, V.A. and P.A. Ruud (1994), "Classical Estimation Methods for LDV Models using Simulation, in R.F. Engle and D.L. McFadden eds., *Handbook of Econometrics: Vol.4*, Amsterdam: North-

Holland, 2384-2441.

- McFadden, D. (1989), “A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration”, *Econometrica*, Vol.57, 995-1026.
- Pakes, A.S. and Pollard, D. (1989), “Simulation and the Asymptotics of Optimization Estimators”, *Econometrica*, Vol.57,
- *Review of Economics and Statistics*, November 1994 issue has several articles.
- Stern, S. (1997), “Simulation-Based Estimation”, *Journal of Economic Literature*, Vol.*35, 2006-2039.*