

# Computational Methods in Economics

## Simulations Methods

March 22, 2020

## General idea

*Consider a model in which the true value,  $\theta_0$ , of a parameter vector is implicitly defined as the unique solution to an equation  $G(\theta) = 0$  for a suitable vector-value function,  $G$ . A natural way to estimate  $\theta_0$  is to construct a sequence  $\{G_n\}$  of random functions that converges to  $G$  in some sense, then find the  $\hat{\theta}_n$  that makes  $G_n(\hat{\theta}_n)$  as close to zero as possible.*

# Simulated Maximum Likelihood

- ▶ Models where the density involves an integral with no closed form solution
- ▶ replace the integral with a Monte Carlo integral
- ▶ Some examples
  - ▶ Random Parameters
  - ▶ Multinomial Probit

## SML: Illustration

Suppose the conditional density  $f(y|x, \theta)$  for an observation is intractable

$$f(y_i|x_i, \theta) = \int h(y_i|x_i, \theta, \epsilon) dg(\epsilon) \quad (1)$$

Use monte-carlo methods to evaluate the density:

$$\hat{f}(y_i|x_i, \theta) = \frac{1}{S} \sum_{s=1}^S h(y_i|x_i, \theta, \epsilon_{is}) \quad (2)$$

where  $\epsilon_{iS} = \{\epsilon_{i1}, \dots, \}$  are draws from  $g(\cdot)$

# Variants

Two types of estimation:

- ▶ Frequency simulator: construct the latent variable, and generate the outcome variable
- ▶ Sampling methods: construct the empirical probabilities from draws

## Example 1: Multivariate Models

Consider  $J(> 2)$  binary choices such that, for example the trivariate probit,

$$y_1 = \begin{cases} 1 & \text{if } x_1' \beta + \epsilon_1 > 0 \\ 0 & \text{if } x_1' \beta + \epsilon_1 < 0 \end{cases} \quad (3)$$

$$y_2 = \begin{cases} 1 & \text{if } x_2' \gamma + \epsilon_2 > 0 \\ 0 & \text{if } x_2' \gamma + \epsilon_2 < 0 \end{cases} \quad (4)$$

$$y_3 = \begin{cases} 1 & \text{if } x_3' \delta + \epsilon_3 > 0 \\ 0 & \text{if } x_3' \delta + \epsilon_3 < 0 \end{cases} \quad (5)$$

$$(\epsilon_1, \epsilon_2, \epsilon_3) \sim \mathbb{N}_3 \left( 0, \Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix} \right) \quad (6)$$

## Probabilities and Likelihood

For example,

$$p_{111} = Pr(y_1 = 1, y_2 = 1, y_3 = 1) \quad (7)$$

$$= Pr(x'_1\beta + \epsilon_1 > 0, x'_2\gamma + \epsilon_2 > 0, x'_3\delta + \epsilon_3 > 0) \quad (8)$$

$$= Pr(x'_1\beta > -\epsilon_1, x'_2\gamma > -\epsilon_2, x'_3\delta > -\epsilon_3) \quad (9)$$

$$= Pr(\epsilon_1 < x'_1\beta, \epsilon_2 < x'_2\gamma, \epsilon_3 < x'_3\delta) \quad (10)$$

$$= \int_{-\infty}^{x'_1\beta} \int_{-\infty}^{x'_2\gamma} \int_{-\infty}^{x'_3\delta} \phi_3(z_0, z_1, z_2, \Sigma) dz_0 dz_1 dz_2 \quad (11)$$

The log-likelihood is given by:

$$\log \mathcal{L} = \sum_i^n \sum_j \sum_k \sum_l y_{ijkl} \log(p_{ijkl}) \quad (12)$$

# Application

- ▶ Given  $\beta_0$  and  $\Sigma_0$
- ▶ Simulate  $S$  multivariate distributions using  $\Sigma_0$
- ▶ Construct  $y_{1s}^*$ ,  $y_{2s}^*$  and  $y_{3s}^*$
- ▶ Evaluate the empirical probabilities, as the frequency 
$$\frac{\sum_s I(y_s^* > 0)}{S}$$

In practice, in multivariate models, the frequency simulator is not efficient.



# Evaluation

- ▶ Propose an algorithm to estimate the binary choice model (probit) using a frequency simulator.
- ▶ Redo part 7 of assignment 4 using the proposed simulator.
- ▶ Assignment 5 is available, and due on Tuesday 31st of March.

Simulated Maximum Likelihood

Method of Moments

Simulated Method of Moments

## Moment estimators

Let  $X_1, \dots, X_n$  be a sample from a distribution  $P_\theta$  that depends on a parameter  $\theta$ , ranging over some set  $\Theta$ . The method of moments consists of estimating  $\theta$  by the solution of a system of equations

$$\frac{1}{n} \sum_{i=1}^n f_j(X_i) = E_\theta f_j(X) \quad j = 1, \dots, k$$

for given functions  $f_1, \dots, f_k$ . The parameter is chosen such that the sample moments match the theoretical moments.

# Method of Moments in Linear Models

- ▶ Orthogonality condition in Linear Models

$$E(x(y' - x)) = 0 \quad (13)$$

- ▶ Moment Condition

$$\frac{1}{N} \sum_i x_i (y_i - x_i' \beta) \quad (14)$$

- ▶ Moment Estimator

$$\hat{\beta}_{\text{MM}} = \left( \sum_i x_i x_i' \right)^{-1} \left( \sum_i x_i y_i \right) \quad (15)$$

# Nonlinear Model

- ▶ Consider

$$Y_i = g(X_i, b_0) + u_i$$

- ▶ Orthogonality Condition

$$E[X'(y - g(X, b_0))] = 0$$

- ▶ Moments condition

$$E_0 h(Y, X, a_0) = 0$$

- ▶ The function  $h$  is  $H$ -dimensional and the parameter  $a$  is of size  $K$ .

# Formal Idea

## Definition

The basic idea of generalized method of moments is to choose a value for  $a$  such that the sample mean is closest to zero.

$$\frac{1}{n} \sum_{i=1}^n h(Y_i, X_i, a)$$

# Formal Definition

## Definition

Let  $\mathbb{S}_n$  be an  $(H \times H)$  symmetric positive definite matrix that may depend on the observations. The generalized method of moments (GMM) estimator associated with  $\mathbb{S}_n$  is a solution  $\tilde{a}_n(\mathbb{S}_n)$  to the problem

$$\min_a \left[ \sum_{i=1}^n h(Y_i, X_i, a) \right]' \mathbb{S}_n \left[ \sum_{i=1}^n h(Y_i, X_i, a) \right]$$

# Assumptions

- H1 The variables  $(Y_i, X_i)$  are independent and identically distributed.
- H2 The expectation  $E_0 h(Y, X, a)$  exists and is zero when  $a$  is equal to the true value  $a_0$  of the parameter of interest.
- H3 The matrix  $S_n$  converges almost surely to a nonrandom matrix  $S_0$
- H4 The parameter  $a_0$  is identified from the equality constraints, i.e.  $E_0 h(Y, X, a)' S_0 E_0 h(Y, X, a) = 0$
- H5 The parameter value  $a_0$  is known to belong to a compact set  $\mathcal{A}$
- H6 The quantity  $(1/n) \sum_{i=1}^n h(Y_i, X_i, a)$  converges almost surely and uniformly in  $a$  to  $E_0 h(Y, X, a)$
- H7 The function  $h(Y, X, a)$  is continuous in  $a$
- H8 The matrix  $\left[ E_0 \frac{h(Y, X, a)}{\partial a} \right]' S_0 \left[ E_0 \frac{h(Y, X, a)}{\partial a'} \right]$  is nonsingular, which implies  $H \geq K$ .



# Asymptotic Normality

Under the assumptions, we have

$$\sqrt{n}(\tilde{a}_n(S_n) - a_0) \sim \mathbb{N}(0, \Sigma(S_0))$$

where

$$\begin{aligned} \Sigma(S_0) = & \left( \left[ E_0 \frac{h(Y, X, a)}{\partial a} \right]' S_0 \left[ E_0 \frac{h(Y, X, a)}{\partial a'} \right] \right)^{-1} \\ & \left( \left[ E_0 \frac{h(Y, X, a)}{\partial a} \right]' S_0 V_0(h(Y, X, a_0)) S_0 \left[ E_0 \frac{h(Y, X, a)}{\partial a'} \right] \right)^{-1} \\ & \left( \left[ E_0 \frac{h(Y, X, a)}{\partial a} \right]' S_0 \left[ E_0 \frac{h(Y, X, a)}{\partial a'} \right] \right)^{-1} \end{aligned}$$

# Optimal GMM

- ▶  $\mathbb{S}_0$  is not known.
- ▶ Two-step procedure
  - ▶ Estimate

$$\min_a \left[ \sum_{i=1}^n h(Y_i, X_i, a) \right]' I \left[ \sum_{i=1}^n h(Y_i, X_i, a) \right]$$

where  $I$  is the identity matrix, and recover  $\hat{a}$ .

- ▶ Matrix of variance/covariance

$$\hat{\mathbb{S}} = \frac{1}{N} \sum_{i=1}^n h(Y_i, X_i, \hat{a}) h(Y_i, X_i, \hat{a})'$$

# Relationship to Simulations

When models are highly nonlinear, moments conditions are not that simple.

- ▶ Indirect Inference
- ▶ Simulated method of moments

# SMM

The basic idea of simulated estimation methods is to adjust the parameter of interest  $\theta$  in order to get similar properties for the observed endogenous variables  $y_t$  and their simulated counterparts ( $y_t^s(\theta)$ ).

- ▶ What does “similar properties” mean? This can be qualitative and quantitative.
- ▶ GMM requires a closed form for the specification of the moments. Under SMM, the moments are replaced by an approximation based on simulations. Such an approximation is referred to as **simulator**. Simulated counterparts have to be easier to obtain.

# SMM

The GMM estimator is given by:

$$\zeta = \underset{\zeta}{\operatorname{argmin}} \quad \Delta(\zeta)\Sigma(\zeta)\Delta(\zeta)'$$

Where  $\Delta(\zeta)$  is the distance between the simulated and empirical moments for the specific set of moments  $\zeta$ .

# Inference

- ▶ In practice, most applications will set the matrix of variance/covariance to a diagonal matrix where each diagonal is equal to 1 divided by the variance of each moment. As a consequence, each moment will be weighted by the inverse of its variance.
- ▶ There is no closed form solution for most moment. Bootstrap is used to evaluate the variance.
- ▶ The variance of estimator has this closed form solution

$$V(\zeta) = \frac{1}{N}(\widehat{G}'\widetilde{\Sigma(\zeta)}^{-1}\widehat{G})^{-1}$$

where  $G$  is the vector of gradient of the optimum.

## SMM - Finding moments

- ▶ Finding moments is the hardest part of this estimation process.
- ▶ Which moments matter? Depends on the question. Moment selection is usually based on the previous literature. Not the best practice.
- ▶ There is an old literature on efficient moment selection, but too costly to implement when model is complicated.

## Indirect inference

- ▶ Indirect inference is a solution to the problem of finding moments. Sometimes researchers do not care about moments per se, but they do care about some correlations.
- ▶ For example, I estimate a model of schooling investment in general equilibrium, and what I care about is the observed correlation between ability, education, experiences and wages in the data.
- ▶ This idea could be captured using expected wages conditional on education, ability and experiences. However, such an undertaking will require too many moments.
- ▶ The idea of indirect inference in this setting is to run a regression of wages on ability, education and experience in the data and in the simulated data, and minimizes the difference between the two sets of coefficients.



# Indirect inference

- ▶ Interest in a set of parameters  $\theta$  such that  $G(\theta) = 0$
- ▶  $G()$  is unknown or extremely difficult to evaluate
- ▶ Find an auxiliary model that fits the statistical properties of the data
- ▶ Estimate the parameters that match best the description of the data

## Details

- ▶ From the data, find  $\zeta^d = \arg \max_{\zeta} L(\{X\}, \zeta)$
- ▶ For a candidate parameter  $\theta_c$ , evaluate the auxiliary model  $\hat{\zeta}^s(\theta_c) = \arg \max_{\zeta} L(\{X\}, \zeta)$
- ▶ There is a mapping  $\tilde{\beta}(\theta)$  that rationalizes the parameters. It can be recovered by simulating and averaging  $\hat{\beta}^s(\theta)$

$$\tilde{\beta}(\theta) \approx \frac{1}{H} \sum_h^H \hat{\beta}_h^s(\theta)$$

- ▶ Minimize the distance  $(\hat{\beta}^d - \tilde{\beta}(\theta))^T W (\hat{\beta}^d - \tilde{\beta}(\theta))$