SIMULATION BASED ECONOMETRIC METHODS
Prepared for EC711, Prof. Pierre Perron

Vladimir Yankov
Boston University

April 11, 2007

1 Introduction

This presentation will deal with a class of estimation problems in which the econometric model and the associated inference approaches lead to a criterion function without simple analytical expression. The analytic intractability often arises from the presence of integrals of large dimension in the probability density function and in the moment conditions. The idea is to circumvent this numerical difficulty by an approach based on simulations.
Examples of applicability of the estimation method:

- **Limited Dependent Variables**
- **Aggregation effects**
- **Unobserved heterogeneity**
- **Nonlinear dynamic models with unobservable factors**
- **Specification resulting from the optimization of some expected criterion**

More specific examples are given in the collection of papers on Simulation-based inference edited by Mariano et al. (2000). There are several issues that arise in these estimation procedures. The use of simulations adds and additional source of variability in the model. Therefore, these estimators will have asymptotic properties that will depend not only on the sample size but also on the simulation size. Moreover, the price to pay for the presence of simulations is loss in efficiency. We are going to summarize the main results in the following simulation based estimation methods

- **Method of Simulated Moments (MSM)**
- **Simulated Maximum Likelihood**
- **Indirect Inference and Efficient Method of Moments**

2 Method of simulated moments

The application of the GMM estimation procedure requires that the moments used have analytical expression. When the moment conditions do not have an analytical form, they can be replaced by an approximation based on simulations. The Method of Simulated moments is a simulation based estimation procedure that circumvents the intractability of the moment conditions in a GMM setting.

2.1 Static models

Suppose we take the *static case*. There are i.i.d. observations (stationary and ergodic) on $(y_i, z_i), i = 1..N$ on endogenous $y_i(m \times 1)$ and exogenous variables $z_i(r \times 1)$. In the presentation of the GMM model (Hansen 1992) the true value of the parameter of interest $\theta_0 \in \Theta \subset R^p$ is defined by a set of moment conidtions of the form:

$$E_0[g(y_i, z_i, \theta_0)|z_i] = 0$$

where $g()$ is a given function of size q: $g : R^m \times R^r \times \Theta \rightarrow R^q$ and θ_0 is the unique point in Θ for which the moment conditions hold.
In particular we may have the following functional form:

$$E_0[H(y_i, z_i; \theta_0)|z_i] = m(z_i, \theta_0)$$
$$\text{so that } g(y_i, z_i, \theta_0) = H(y_i, z_i; \theta_0) - m(z_i, \theta_0))$$

The MSM is applied when the moment function g does not have an analytic expression. In this case we might be have a simulator $\tilde{g}(y_i, z_i, u_i; \theta)$ with a known functional form. It will depend on the auxilliary term u_i with a known and fixed distribution conditional on (y, z) and is an unbiased estimator of g:

$$E_u[\tilde{g}(y_i, z_i, u_i; \theta)] = g(y_i, z_i; \theta)$$

where E_u is the expectation with respect to the distribution of u.
Then we can define an MSM estimator as the solution of the optimization problem:

$$\hat{\theta}_{Sn}(\Omega) = argmin_{\theta}\Big[\sum_{i=1}^{n} Z_i \frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y_i, z_i, u_i^s;\theta)\Big]'\Omega\Big[\sum_{i=1}^{n} Z_i \frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y_i, z_i, u_i^s;\theta)\Big] \equiv argmin_{\theta}\Psi_S(\theta) \tag{1}$$

where $u_i^s, i = 1...N,\ s = 1, ..., S$ are independent drawings from the distribution of u and Z_i are instrumental variable functions of z_i.

2.2 Asymptotic properties of MSM

The asymptotic properties of the MSM estimation procedure are derived in McFadden (1989) and in Pakes and Pollard (1989). The idea is to provide conditions under which we can map the problem (1) into the conditions of Hansen (1992) so that we can assure consistency and normality of the simulated estimator. In particular, we might run into problems of having an objective function that is discontinuous over the parameter space Θ - $G_{Sn}(\theta) \equiv \sum_{i=1}^{n} Z_i \frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y_i, z_i, u_i^s;\theta)$. Pakes and Pollard (1989) state conditions under which we achieve consistency. In particular we need $||G_{sn}(\theta) - G(\theta)||$ be uniformly "small" in θ and a uniform law of large numbers. These conditions, however, are not enough in the case of serial correlation.

Consistency and normality (i.i.d.) Under certain regularity conditions that map into the conditions of Hansen (1982) when n goes to infinity and S *is fixed*, the estimator $\hat{\theta}_{Sn}(\Omega)$ is:

1. consistent $\hat{\theta}_{Sn}(\Omega) \rightarrow_p \theta_0$
2. normally distributed $\sqrt{N}(\hat{\theta}_{SN}(\Omega) - \theta_0) \rightarrow^{d}_{n\rightarrow\infty}\ N(0, Q_S(\Omega))$
   where $Q_S(\Omega) = \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} + \frac{1}{S}\Sigma_1^{-1}D'\Omega E_0[ZV_u(\tilde{g}|y, z)Z']\Omega D\Sigma^{-1}$
   $\Sigma_2 = D'\Omega V_0(Zg)\Omega D$
   $D = E_0[Z\frac{\partial g}{\partial\theta'}]$
   $\Sigma_1 = D'\Omega D.$

NOTE: We can achieve consistency with a finite number of simulations S and letting the sample size go to infinity (see McFaden, 1989). However, this type of asymptotics will not hold for the Simulated maximum likelihood estimation where we will need to have N and T go to infinity at a particular rate.

Sketch of the proof:

(1) Consistency: Under a uniform law of large numbers the criterion function (1) will converge almost surely to:

$$[E_z(ZE_y^0E_u\tilde{g}(y,z,u^s;\theta))]'\Omega[E_z(ZE_y^0E_u\tilde{g}(y,z,u^s;\theta))]$$

Since $E_u\tilde{g}(y,z,u^s;\theta) = g(y,z;\theta)$ and $E_y^0g(y,z;\theta_0) = 0$ for the unique $\theta_0 \in \Theta$ (*compact*). The unique minimum of the asymptotic limit of the criterion function is θ_0 and therefore $\hat{\theta}_{Sn} \rightarrow_p \theta_0$.
(2) Asymptotic normality: The first-order conditions of the minimization problem defining the estimator $\hat{\theta}_{SN}(\Omega)$:

$$0 = \frac{1}{n\sqrt{n}}\sum_{i=1}^{N}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{g}'}{\partial\theta}(y_i,z_i,u_i^s,\hat{\theta}_{SN})Z_i'\Omega\sum_{i=1}^{N}Z_i\frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y_i,z_i,u_i^s,\hat{\theta}_{SN})$$

Taking a first-order Taylor approximation of the first-order conidtions around θ_0 and letting N go to infinity we obtain:

$$(1) 0 \approx \frac{1}{N}\sum_{i=1}^{N}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{g}'}{\partial\theta}(y_i,z_i,u_i^s,\theta_0)Z_i'\Omega\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i\frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y_i,z_i,u_i^s,\theta_0)+$$

$$(2) +\frac{1}{N}\sum_{i=1}^{N}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{g}'}{\partial\theta}(y_i,z_i,u_i^s,\theta_0)Z_i'\Omega\frac{1}{N}\sum_{i=1}^{N}Z_i\frac{1}{S}\sum_{s=1}^{S}\frac{\partial \widetilde{g(y_i,z_i},u_i^s,\theta_0)}{\partial\theta'}\sqrt{N}(\hat{\theta}_{SN}-\theta_0)+$$

$$(3) +\frac{1}{N}\sum_{j=1}^{q}\frac{1}{S}\sum_{s=1}^{S}\frac{\partial^2\tilde{g}_j(y_i,z_i,u_i^2;\theta_0)}{\partial\theta\partial\theta'}\Big\{Z_i'\Omega\frac{1}{N}\sum_{i=1}^{N}Z_i\frac{1}{S}\sum_{i=1}^{S}\tilde{g}(y_i,z_i,u_i^s;\theta_0)\Big\}_j\sqrt{N}(\hat{\theta}_{SN}-\theta_0)$$

As $\frac{1}{N}\sum_{i=1}^{N}Z_i\frac{1}{S}\sum_{i=1}^{S}\tilde{g}(y_i,z_i,u_i^s;\theta_0) \rightarrow_p 0$ the last term (3) goes to zero.
In the limit:

$$D = lim_{N\rightarrow\infty}\Big\{\frac{1}{N}\sum_{i=1}^{N}Z_i\frac{1}{S}\sum_{s=1}^{S}\frac{\partial\tilde{g}'}{\partial\theta}(y_i,z_i,u_i^s,\theta_0)\Big\} = E_0[Z\frac{\partial g'}{\partial\theta}(y_i,z_i,u_i^s,\theta_0)]$$

The sample moment condition will obey a Central limit theorem:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i\frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y_i,z_i,u_i^s,\theta_0) \rightarrow_d N(0,\Sigma_g)$$

$$\Sigma_g = V\Big[Z\frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y,z,u^s,\theta_0)\Big] =$$
$$V_0E_u\Big[Z\frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y,z,u^s,\theta_0)\Big] + E_0V_u\Big[Z\frac{1}{S}\sum_{s=1}^{S}\tilde{g}(y,z,u^s,\theta_0)\Big] \Rightarrow$$

$$\Sigma_g = V_0\Big[Zg(y,z,\theta_0)\Big] + \frac{1}{S}E_0\Big[ZV_u\tilde{g}(y,z,u^s,\theta_0)Z'\Big]$$

Therefore $\sqrt{N}(\hat{\theta}_{SN}-\theta_0)$ converges in distribution to $N(0,Q_S(\Omega))$ where:

$$Q_s(\Omega) = (D'\Omega D)^{-1}D'\Omega\Sigma_g\Omega D(D'\Omega D)^{-1} = \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} + \frac{1}{S}\Sigma_1^{-1}D'\Omega E_0(ZV_u\tilde{g}Z')\Omega D\Sigma_1^{-1}$$

Note: The asymptotic variance-covariance matrix is decomposed into two terms the first is the asymptotic covariance matrix of the GMM estimator and the second is decrease in efficiency due to simulations. The efficiency of the SMS estimator also depends on the quality of the simulator - $V_u g(y, z, u^s, \theta_0)$. Therefore:

- If $S \to \infty$ then MSM and GMM are asymptotically equivalent and for finite S - $V_{as}(\hat{\theta}_{SN}) > V_{as}(\hat{\theta}_{GMM})$
- The quality of the simulator - $V_u g(y, z, u^s, \theta_0)$ depends on the number of random generators. For example, if we have a simulator with two random generators $\tilde{g}(y, z, u_1, u_2; \theta)$ and we can integrate out one of them $\hat{g}(y, z, u_1; \theta) = E(g(y, z, u_1, u_2; \theta)|u_1)$. Then we have gain in efficiency:

$$V(\tilde{g}) = V[E(\tilde{g}|u_1)] + E[V(\tilde{g}|u_1)] = V(\hat{g}) + E[V(\tilde{g}|u_1)] > V(\hat{g})$$

- In particular if we have the reduced form specification of $y_i = r(z_i, \epsilon_i, \theta_0)$

  if we use a simulator $\tilde{g}(y_i, z_i, \epsilon; \theta)$ and we reduce the number of randon generators to u a subset of ϵ and use $\hat{g}(y, z, u; \theta) = E_\theta(g(y, z; \theta)|z, u)$

2.2.1 Optimal choice of Ω

The asymptotic variance-covariance matrix of the MSM is:

$$Q_S(\Omega) = (D'\Omega D)^{-1} D'\Omega \Big\{ V_0(Zg) + \tfrac{1}{S} V_0[Z\tilde{g}] \Big\} \Omega D (D'\Omega D)^{-1}$$

From the Gauss-Markov theorem the nonnegative symmetric matrix above is minimized for $\Omega^* = \Big\{ V_0(Zg) + \frac{1}{S} V_0[Z\tilde{g}] \Big\}^{-1}$. Then the optimal asymptotic variance covariance matrix is:

$$Q_S(\Omega^*) = (D'\Omega^* D)^{-1}$$

The way we estimate Ω is similar to the GMM. We start with a consistent estimator of θ - for example $\hat{\theta}_{SN}(I)$. Then we can use the following estimate:

$$\hat{\Omega}^* = \Big\{ \tfrac{1}{N} \textstyle\sum_{i=1}^{N} Z_i \Big[\tfrac{1}{S_2} \sum_{s=1}^{S_2} \tilde{g}(y_i, z_i, u_{i,2}^s; \hat{\theta}_{SN}) \Big] \Big[\tfrac{1}{S_2} \sum_{s=1}^{S_2} \tilde{g}(y_i, z_i, u_{i,2}^s; \hat{\theta}_{SN}) \Big]' Z_i' +$$

$$+ \tfrac{1}{S} \tfrac{1}{N} \textstyle\sum_{i=1}^{N} Z_i \Big[\tilde{g}(y_i, z_i, u_{i,2}^{S_1}; \hat{\theta}_{SN}) - \tfrac{1}{S_2} \sum_{i=1}^{S_2} \tilde{g}(y_i, z_i, u_{i,2}^s; \hat{\theta}_{SN}) \Big] \Big[\tilde{g}(y_i, z_i, u_{i,2}^{S_1}; \hat{\theta}_{SN}) - \tfrac{1}{S_2} \sum_{i=1}^{S_2} \tilde{g}(y_i, z_i, u_{i,2}^s; \hat{\theta}_{SN}) \Big]' Z_i' \Big\}^{-1}$$

where $u_{i,2}^{S_1}$ is the simulated value of u. Here we need a large number of simulated values of $u_{i,2}^{s}$ - S_2. This estimator $\hat{\Omega}^*$ is consistent when N and S_2 tend to infinity. So, in contrast to the asymptotic properties of the MSM estimator of θ where only a small number of simulated values is required for consistency, for the determination of the optimal matrix a larger number of simulated values is required.

2.3 Dynamic models

Lee and Ingram (1991) and Duffie and Singleton (1993) define a SMM in a dynamic models. The discussion above relies on the assumption that there is no serial correlation in the data which in time series applications is usually not satisfied. The presence of dependence in the data produces the following complications which need to be addressed when we use simulations:

- A technical question arises whether we should constrain our optimizaton procedures to a subset of the parameter space Θ where the model generates stationary and ergodic data. The issue is of practical importance as the econometric theory presupposes stationarity and ergodicity. Tauchen (1998) examines the issue and shows that we do not need to worry about imposing dynamic stability conditions on the model itself as dynamic stability is self-enforcing - the objective function attains exceedingly high values for any parameter value in the instability region. However, in the Indirect inference estimators presented below we need to impose such restrictions on the auxiliary model used to generate moment conditions.

- In simulating time series pre-sample values of the series are typically required. In most circumstances, however, the stationary distribution of the simulated process as a function of the parameter space is not known (due to non-linearities, for example). Hence, the initial conditions for the time series will generally not be drawn from their stationary distribution and the simulated process will generally be non-stationary

- Functions of the current value of the simulated state depend on the unknown parameter vector both through the structure of the model and indirectly through the generation of data by simulation. The feedback of the latter dependence on the transition law of the simulated state process implies that the first-moment-continuity condition in Hansen (1982) in establishing the uniform convergence of the sample to the population criterion functions are not directly applicable to the SME. Furthermore, non-stationarity of the simulated series has to be accommodated in establishing the asymptotic normality of SMM.

Lee and Ingram (1991) Econometric model:

- The model (e.g. a stochastic general equilibrium) generates $m \times 1$ stochastic processes $\{y(\theta)\}$ $\theta \in \Theta \subset R^p$
- We observe data set $m \times T$ data set $\{x_t\}$
- Under H_0 : Model is correctly specified $\exists!$ $\theta_0 \in \Theta$ such that $\{y(\theta_0)\}$ and $\{x_t\}$ are drawn from the same distribution
- In practice we would generate $s = 1...S$ of $\{y_s(\theta), s = 1..S\}$ and we would observe $\{x_t, t = 1...T\}$
- Assume $\{x_t, t = 1...T\}$ and $\{y_s(\theta), s = 1..S\}$ are ergodic
- Heuristically the SMM estimator of θ_0 is obtained by equating the sample counterparts of the moments of the simulated process to the sample counterparts of the moments of the observed data process

Suppose we can form the following $q \times 1$ vector of statistics:

$$H_T(x) = \frac{1}{T}\sum_{t=1}^{T} h(x_t)$$

$$H_S(y(\theta)) = \frac{1}{S}\sum_{s=1}^{S} h(y_s(\theta))$$

Then for θ we will have that:

$$H_T(x) \rightarrow^{a.s.} E[h(x_t)] \text{ as } T \rightarrow \infty$$

$$H_S(y(\theta)) \rightarrow^{a.s.} E[h(y(\theta))] \text{ as } S \rightarrow \infty$$

Under the H_0 that the economic model is correctly specified at θ_0:

$$E[h(y(\theta_0))] = E[h(x_t)]$$

- Duffie and Singleton (1993) examine an asset pricing model which leads to the following Markov process for the data x: $x_t = T(x_{t-1}, \epsilon_t, \theta_0)$ and statistics function $h(x_{t-1}, x_{t-2}, ..x_0, \theta_0)$. In special cases for $T(\cdot)$ and $h(\cdot)$ the function mapping $\theta \rightarrow E(h(x_{t-1}, x_{t-2}, ..x_0, \theta_0))$ does not have an analytic form. The econometrician can simulate indepenent new history for driving shocks $\hat{\epsilon}_t$ and recursively define the simulated state process: $y_t^\theta = T(y_{t-1}^\theta, \hat{\epsilon}_t, \theta)$. The statistics functions $h_T(x) = h_T(x_t, x_{t-1}, ..., x_0)$ have a corresponding simulated counterpart $h_S(y^\theta, \theta) = h_S(y_S^\theta, y_{S-1}^\theta, ..., y_0^\theta; \theta)$. The difference from Lee and Ingram is that the statistic functions now are allowed to depend on θ not only directly but indirectly throught the dependence of the entire past history of the simulated process $\{y_s(\theta)\}_{t=0..S}$ on θ. The inference is the same as in Lee and Ingram, however, it requires some additional technical conditions. In particular, for any $\theta \in \Theta$ the process should satisfy geometric ergodicity which is a strong form of stationarity that says that the process converges geometircally to its stationary distribution and the transition function must satisfy some form of continuity. In this case the possible non-stationarity induced by the initial conditions problem can be ignored when studying the asymptotic properties of SMM estimator.

The SMM estimator:

$$\theta_{TS} = argmin_\theta [H_T(x) - H_S(y(\theta))]' W_T [H_T(x) - H_S(y(\theta))]$$

Let us fix an integer $n = \frac{S}{T} > 1$ and define $g_T(\theta)$ and $f_t(\theta)$:

$$g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} f_t(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left[h(x_t) - \frac{1}{n} \sum_{k=1}^{n} h(y_{k,t}(\theta) \right]$$

Then the SMM estimator takes the familiar form:

$$\theta_{TS} = argmin_\theta g_T(\theta)' W_T g_T(\theta)$$

2.3.1 Asymptotic properties

In the formulation of the estimator we have a special case of the generic GMM estimator of Hansen (1982). The consistency and asymptotic normality of the estimator can be verified if we can verify the theorems of Hansen (1982) hold.

Consistency

- $\exists ! \theta_0 \in \Theta : E[f_t(\theta_0)] = 0$ identification
- $\{x_t\}$ and $\{y(\theta)\}$ must be stationary, ergodic and independent of each other
- $h(y_s(\theta))$ must be continuous in the mean:

$$lim_{\delta \to 0} E\Big[sup\{|h(y_s(\theta_0)) - h(y_s(\theta_1))| : \theta_0, \theta_1 \in \Theta, |\theta_0 - \theta_1| < \delta\}\Big] = 0$$

Under the conditions above $\theta_{TS} \to_p \theta_0$ as $T, N \to \infty$.

Asymptotic Normality

- $\frac{\partial h(y_s(\theta))}{\partial \theta}$ must be continuous in the mean in a neighbourhood of θ_0
- $B \equiv E[\frac{\partial h(y_s(\theta))}{\partial \theta}]$ must exist and be finite and full rank
- $W_T \to_p W$
- Let $w_t = f_t(\theta_0)$ and $\nu_i = E[w_t|w_{t-i}, w_{t-i-1}, ...] - E[w_t|w_{t-i-1}, w_{t-i-1-2}, ...],\ i \geq 0$. We assume that:
  - $E[w_t w_t'] < \infty$ exists
  - $E[w_t|w_{t-i}, w_{t-i-1}, ...] \to_p 0$
  - $\sum_{i=0}^{\infty} E[\nu_i \nu_i']^{1/2} < \infty$
- Let $R_x(i) = cov(h(x_t), h(x_{t-i}))$ and $R_y(i) = cov(h(y_t(\theta_0)), h(y_{t-i}(\theta_0))$ then define $\Omega = \sum_{i=-\infty}^{\infty} R_x(i)\Big(= \sum_{i=-\infty}^{\infty} R_y(i)\ under\ the\ H_0\Big)$

Given the assumptions under H_0:

$$\sqrt{T}(H_T(x) - E(h(x_t))) \to_D N(0, \Omega)$$

$$\sqrt{S}(H_S(y(\theta_0)) - E(h(y(\theta_0)))) \to_D N(0, \Omega)$$

Given the independence of $\{x\}$ and $\{y(\theta_0)\}$:

$$V(g_T(\theta)) = cov(H_T(x) - H_S(y(\theta_0)) = (1 + \tfrac{1}{n})\Omega$$

Under the above conditions and the assumption that $N/T \to n$ as $T, S \to \infty$;

$$\sqrt{T}(\theta_{ST} - \theta_0) \to_D N(0, (B'WB)^{-1}B'W(1 + \tfrac{1}{n})\Omega W B(B'WB)^{-1'}) \text{ as } T, S \to \infty$$

Optimal weighing matrix

$$W = [(1 + \frac{1}{n})\Omega]^{-1}$$

The asymptotic distribution:

$$\sqrt{T}(\theta_{ST} - \theta_0) \rightarrow_D N(0, [B'(1 + \frac{1}{n})^{-1}\Omega^{-1}B]^{-1}) \text{ as } T, S \rightarrow \infty$$

Estimator of Ω

Since under $H_0 : R_x(i) = R_y(i)$ we can estimate Ω using the method proposed by Newey and West (1987) and Andrews (1991) to estimate the spectral density at frequency zero:

$$\hat{\Omega} = \sum_{i=-(T-1)}^{T-1} \lambda(i, m)\hat{R}_x(i)$$

where $\lambda(i, m)$ is an appropriately chosen weighting function.

The SMM procedure as in Hansen (1982) provides a goodness-of-fit J statistic:

$$T[H_T(x) - H_S(y(\hat{\theta}_{SN}))]'\hat{\Omega}_T[H_T(x) - H_S(y(\hat{\theta}_{SN}))] \rightarrow_D \chi^2(q - p)$$

Note:

- The randomness of the estimator is derived from two sources: the randomness in the simulation and the randomness in the real data. As $n = \frac{S}{T}$ becomes large, the importance of the randomness in the simulation to the covariance matrix of the estimator declines.
- The underlying assumptions for the derivation of the properties of $\hat{\theta}_{TS}$ is that we have generated a realization from the stationary and ergodic stochastic process $\{y_s(\theta)\}$ and in particular we should select $y_0(\theta)$ from this stationary distribution. This distribution will depend on the unknown parameter vector θ. In a linear model it could be easy to calculate the unconditional distribution of $y_0(\theta)$ however in a non-linear model if y_s depends on y_{s-j} in a non-linear fashion it may be hard to do so. Therefore, a practical suggestion is to use an arbitrary y_0 and generate a realization from the stochastic process $\{y_s(\theta)\}$ of length $2S$ and throw away the first S data points and use the remaining S data points. For S large enough the influence of the starting value should be negligible

2.4 Drawings and Gauss-Newton algorithms

The optimization procedure for the criterion functon usually involves evaluation of the objective function for different values of θ. It is **important** that the drawings of u^s be made at the beginning of the procedure and be kept fixed during the execution of the algorithm. If some new drawings are made at each iteration of the minimization algorithm then we are going to introduce new randomness at each step which will affect the regularity conditions - continuity assumptions on the objective function. The algorithm will have difficulty with sorting out a change in the criterion function due to a variation in the random draw from a change due to a modification in β. Therefore, the asymptotic results will also be violated.

3 Simulated Maximum Likelihood

Here we discuss the simulated analogues of the maximum likelihood estimation method. It involves a class of problems where the pdf $f(y_t|x_t;\theta)$ has an untractable form. In some cases we have at our disposal an unbiased simulator $\tilde{f}(y_t,x_t,u;\theta) : E[\tilde{f}(y_t,x_t,u;\theta)|y_t,x_t) = f(y_t,x_t;\theta)$ where u has a known distribution. Then we can define the Simulated maximum likelihood estimator of θ as:

$$\hat{\theta}_{ST} = argmax_\theta \ \psi_T(\theta) = argmax_\theta \sum_{t=1}^{T} log\Big[\frac{1}{S}\sum_{s=1}^{S}\tilde{f}(y_t,x_t,u_t^s;\theta)\Big] \tag{2}$$

where $\{u_t^s, s=1..S\}$ are independent draws from the distribution of u given x_t and y_t.

3.1 Asymptotic properties

- $T \rightarrow \infty$, S fixed. If S = 1 then the objective function:

  $$\tfrac{1}{T}\sum_{t=1}^{T} log\tilde{f}(y_t,x_t,u_t^s;\theta)$$

  tends asymptotically to:

  $$\psi_\infty(\theta) = E_0 \textstyle\int log\tilde{f}(y_t,x_t,u;\theta)g(u)du$$

  where g is the pdf of u. However, due to the Jensen's inequality the asymptotic objective function will not be equal to the asymptotic objective function of the original problem:

  $$\theta_0 \ = argmax_\theta E_0 log(f(y,x;\theta))$$

  Therefore, when the number of simulations is finite the SML will not be consistent.

- $T, S \rightarrow \infty$

  The objective function in this case:

$$lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} log\Big[lim_{S\to\infty}\frac{1}{S}\sum_{s=1}^{S}\tilde{f}(y_t,x_t,u_t^s;\theta)\Big] =$$

$$= E_0\ log\Big[\int \tilde{f}(y_t,x_t,u;\theta)g(u)du\Big] = E_0 log f(y_t,x_t;\theta)$$

This is the same limit problem as for the usual maximum likelihood estimation and the estimator is consistent. In terms of efficiency the rate of $S\to\infty$ may have an effect on the asymptotic covariance matrix of the SML estimator except if the speed of divergence of S is sufficiently large. The following result is proven in Gourieroux and Monfort (1991):

- If $S,T\to\infty$ and $\sqrt{T}/S\to 0$, then SML estimator is asymptotically equivalent to the ML estimator : $\sqrt{T}(\hat{\theta}_{ST}-\theta_0)\to N(0,I(\theta_0)^{-1})$ where $I(\theta_0)$ is the information matrix.

3.2 Correction for the bias

Gourieroux and Monfort (1991) show that the order of the bias is $\frac{1}{S}$:

$$B = E_0(\hat{\theta}_{ST}-\theta_0) \sim \frac{1}{S}I^{-1}(\theta_0)E_0[a(y,x;\theta)]$$

The square of the bias is:

$$B'I(\theta_0)B = \frac{1}{S^2}E_0[a(y,x;\theta)]'I^{-1}(\theta_0)E_0[a(y,x;\theta)]$$

where $a(y,x;\theta_0) = \frac{E_u\frac{\partial\tilde{f}}{\partial\theta}V_u\tilde{f}}{(E_u\tilde{f})^3} - \frac{cov_u(\frac{\partial\tilde{f}}{\partial\theta},\tilde{f})}{(E_u\tilde{f})^2}$ As expected, the bias depends on the choice of the simulator and may be reduced by a sensible selection of $\tilde{f}$. Moreover, if the underlying ML estimator is precise $I^{-1}(\theta_0)$ this also reduces the bias.
We can use the following first-order correction for the asymptotic bias:

$$ln\tilde{f} \approx lnf + \frac{\tilde{f}-f}{f} - \frac{1}{2}\frac{(\tilde{f}-f)^2}{f^2}$$

$$E_u ln\tilde{f} \approx lnf + \frac{E_u(\tilde{f}-f)}{f} - \frac{1}{2}\frac{E_u(\tilde{f}-f)^2}{f^2} = lnf - \frac{1}{2}\frac{E_u(\tilde{f}-f)^2}{f^2}$$

Therefore, the first-order correction of the SML estimator consists in computing the corrected estimator:

$$\hat{\theta}_{ST} = argmax_\theta\Big\{\sum_{t=1}^{T} log[\frac{1}{S}\sum_{s=1}^{S}\tilde{f}(y_t,x_t,u_t^s;\theta)] - \frac{S}{2}\frac{\sum_{s=1}^{S}\Big[\tilde{f}(y_t,x_t,u_t^s;\theta)-\frac{1}{S}\sum_{s=1}^{S}\tilde{f}(y_t,x_t,u_t^s;\theta)\Big]^2}{(\sum_{s=1}^{S}\tilde{f}(y_t,x_t,u_t^s;\theta))^2}\Big\}$$

4 Indirect Inference

The Indirect inference method was developed by Smith (1993) and Gourieroux et al. (1993). When a model leads to a complicated structural or reduced form and to untractable likelihood functions the indirect inference approach proceeds by replacing the initial model M with an auxilliary (instrumental) one M^a which is analytically tractable.

- We are intersted in estimating θ. Let the initial model M have a likelihood function $\sum_{t=1}^{T} logf(y_t|y_{t-1}, z_t; \theta)$ which is infeasible. However, we have a structural or reduced form coming from our economic model which generates the data $y(\theta)_t = r(z_T, \epsilon_t, \theta)$.

- Introduce an auxilliary parameter β and the we use an auxiliary model to form a criterion function to estimate β

$$\hat{\beta}_T = argmax_\beta \psi_T(y_T, z_T; \beta)$$

- Now we can use the initial model to simulate the endogenous variables for different parameter values of θ - $\{y_t(\theta)^s\}$ for $s = 1...S$ and $t = 1..T$. Then we can recompute $\beta(\theta)$ with the simulated data on y which now depends on θ:

$$\hat{\beta}_{ST}(\theta) = argmax_\beta \sum_{s=1}^{S} \psi_T[y_T(\theta)^s, z_T; \beta]$$

- the final step of the Indirect inference is to choose θ which minimizes the distance between β_T and $\beta_{ST}(\theta)$:

$$\hat{\theta}_{ST}(\Omega) = argmin_\theta[\hat{\beta}_T - \beta_{ST}(\theta)]'\Omega[\hat{\beta}_T - \beta_{ST}(\theta)]$$

  where Ω is a symmetric nonnegative matrix.

- Equivalently we could also base our estimates on the scores of the auxiliary criterion function:

$$\hat{\theta}^*_{ST}(\Sigma) = argmin_\theta \Big(\sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_T(\theta)^s, z_T; \hat{\beta}_T] \Big)' \Sigma \Big(\sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_T(\theta)^s, z_T; \hat{\beta}_T] \Big)$$

  where Σ is a symmetric nonnegative matrix.

4.1 Properties of the Indirect inference estimator

4.1.1 Consistency

Regularity conditions

- $\psi_T(y_T^s(\theta), z_T; \beta) \rightarrow_{a.s.} \psi_\infty(\theta, \beta)$ (deterministic) uniformly in both θ and β as $T \rightarrow \infty$
- $\exists! \beta(\theta) = argmax_\beta \psi_\infty(\theta, \beta)$
- ψ_T and ψ_∞ are both differentiable in β and $\frac{\partial \psi_\infty}{\partial \beta}(\theta, \beta) = lim_T \frac{\partial \psi_T}{\partial \beta}(y_T(\theta)^s, z_T; \beta)$
- For the estimation based on the scores of the criterion functions we need that only solution of the asymptotic first order condition is $\beta(\theta) : \frac{\partial \psi_\infty}{\partial \beta}(\theta, \beta(\theta)) = 0$
- The equation $\beta = \beta(\theta)$ has a unique solution in θ ($\beta(\theta)$ is called **the binding function**).

The proof of consistency then relies on the study of the asymptotic problem:

$$\hat{\beta}_T = argmax_\beta \psi_T(y_T, z_T; \beta) \rightarrow argmax_\beta \psi_\infty(\theta_0, \beta) = \beta(\theta_0)$$

$$\hat{\beta}_{ST}(\theta) = argmax_\beta \sum_{s=1}^{S} \psi_T[y_T(\theta)^s, z_T; \beta] \rightarrow argmax_\beta S\psi_\infty(\theta, \beta) = \beta(\theta)$$

Finally, the minimized distance between $\hat{\beta}_{ST}$ and $\hat{\beta}_T$ is:

$$\hat{\theta}_{ST}(\Omega) = argmin_\theta [\hat{\beta}_T - \beta_{ST}(\theta)]'\Omega[\hat{\beta}_T - \beta_{ST}(\theta)] \rightarrow argmin_\theta [\beta(\theta_0) - \beta(\theta)]'\Omega[\beta(\theta_0) - \beta(\theta)] = \theta_0$$

The derivation of the asymptotic normality of the Indirect inference estimator could be found in Gourieroux et. al (1993). Here I just state the result:

> Under a set of regularity conditions the indirect inference estimator $\hat{\theta}_{ST}(\Omega)$ is consistent and asymptotically normal for S fixed and $T \rightarrow \infty$:
>
> $$\sqrt{T}(\hat{\theta}_{ST}(\Omega) - \theta_0) \rightarrow_d N(0, W(S, \Omega))$$
>
> where
>
> $$W(S, \Omega) = (1 + \tfrac{1}{S})[\tfrac{\partial b'}{\partial \theta}(\theta_0)\Omega\tfrac{\partial b}{\partial \theta'}(\theta_0)]^{-1}\tfrac{\partial b'}{\partial \theta}(\theta_0)\Omega\Omega^{*-1}\Omega\tfrac{\partial b}{\partial \theta'}(\theta_0)[\tfrac{\partial b'}{\partial \theta}(\theta_0)\Omega\tfrac{\partial b}{\partial \theta'}(\theta_0)]^{-1}$$

where $\Omega^* = J_0 \bar{I}_0^{-1} J_0$ and

$$J_0 = plim_T - \frac{\partial^2 \psi_T}{\partial\beta\partial\beta'}[y_T, z_T; \beta(\theta_0)]$$

$$\bar{I}_0 = lim_T E_0 \Big\{ \sqrt{T} \frac{\partial \psi_T}{\partial \beta}(y_T, z_T, \beta(\theta_0)) - E_0[\sqrt{T} \frac{\partial \psi_T}{\partial \beta}(y_T, z_T, \beta(\theta_0)) | z_T] \Big\}$$

A similar argument based on the asymptotic score of the auxiliary model gives the consistency of θ^*_{ST}. Moreover, it can be shown that the two estimators are asymptotically equivalent:

$$\sqrt{T}(\hat{\theta}^*_{ST} - \hat{\theta}_{ST}(J_0 \Sigma J_0)) = op(1)$$

Optimal choice of Ω
The optimal matrix $\Omega = \Omega^* = J_0 \bar{I}_0^{-1} J_0$ and the asymptotic variance-covariance matrix of the indirect inference estimator is simplified:

$$W(S, \Omega) = (1 + \frac{1}{S})[\frac{\partial b'}{\partial \theta}(\theta_0) J_0 \bar{I}_0^{-1} J_0 \frac{\partial b}{\partial \theta'}(\theta_0)]^{-1}$$

We can take derivatives of first order condition $\frac{\partial \psi_\infty}{\partial \beta}[\theta, \beta(\theta)] = 0$ with respect to θ to express:

$$\frac{\partial b'}{\partial \theta}(\theta_0) = J_0^{-1} \frac{\partial^2 \psi_\infty}{\partial \beta \partial \theta'}$$

Therefore, the variance-covariance matrix is:

$$W(S, \Omega) = (1 + \frac{1}{S})[\frac{\partial^2 \psi_\infty}{\partial \beta \partial \theta'} \bar{I}_0^{-1} \frac{\partial^2 \psi_\infty}{\partial \beta \partial \theta}]^{-1}$$

Therefore, a consistent estimator is obtained if we replace ψ_∞ with ψ_T and $\beta(\theta_0)$ with $\hat{\beta}$. For $\bar{I}_0$ we can use the approach described in Gourieroux et al (1993) appendix 2:

$$\bar{I}_0 = lim_T V_0 \Big\{ \sqrt{T} \sum_{t=1}^{T} \frac{\partial \psi_T}{\partial \beta}(y_T, z_T, \beta(\theta_0)) \Big\}$$

Therefore we can use the Newey and West (1987) and select an optimal truncation according to Andrews(1991):

$$\hat{I}_0 = \hat{\Gamma}_0 + \sum_{i=1}^{T} \lambda(i, m)[\hat{\Gamma}(i) + \hat{\Gamma}(i)']$$

where:

$$\hat{\Gamma}(k) = \frac{1}{T} \sum_{t=k+1}^{T} \frac{\partial \psi_{t-k}}{\partial \beta}(\hat{\beta}_T) \frac{\partial \psi_t}{\partial \beta'}(\hat{\beta}_T)$$

5 Efficient Method of Moments

The auxiliary criterion function can be a tractable likelihood function (M^a) in which case we have the PML estimator:

$$\hat{\beta}_{ST} = argmax_{\beta} \sum_{s=1}^{S} \sum_{t=1}^{T} log f^a(y_t^s(\theta)|y_{t-1}^s(\theta), z_T; \beta)$$

As a PML estimator β_{ST} will minimize the Kullback-Leibler information criterion $KLIC = E_{\theta} log \frac{f(y|x;\theta}{f^a(y|x;\beta)}$ which gives the "closenes" between f and f^a.
Gallant and Tauchen (1996) propose that we use the score of the auxiliary model evaluated at the pseudo-maximum likelihood $\hat{\beta}_T$ - we can choose θ such that the score is as close as possible to zero:

$$\hat{\theta}_{ST} = argmin_{\theta} \Big[\sum_{s=1}^{S} \sum_{t=1}^{T} \frac{\partial log f^a}{\partial \beta} (y_t^s(\theta)|y_{t-1}^s(\theta), z_t; \hat{\beta}_T) \Big]' \Sigma \Big[\sum_{s=1}^{S} \sum_{t=1}^{T} \frac{\partial log f^a}{\partial \beta} (y_t^s(\theta)|y_{t-1}^s(\theta), z_t; \hat{\beta}_T) \Big]$$

where Σ is a non-negative symmetric matrix. We have already examined this type of estimator based on the score of the auxiliary model.

References

[1] Andrews, D. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", Econometrica, Vol. 59, May 1991, pp. 817-857

[2] Andrews D. "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation", Econometrica Vol 67, No 3 (May 1999), pp. 543-564

[3] Dhaene, G., Gourieroux, C. and Scaillet, O. "Instrumental Models and Indirect Inference", Econometrica, Vol. 66 May 1998, 673-688

[4] Duffie, D. and Singleton, K. "Simulated Moments Estimation of Markov Models of Asset Prices" Econometrica, Vol. 61, 1993, pp. 929-952

[5] Gouriéroux, C. and Monfort, A. "Simulation-based Econometric Methods", 1996 Oxford University Press

[6] Gourieroux, C., Monfront, A. and Renault, E. "Indirect Inference", Journal of Applied Econometrics Vol. 8, Supplement Special Issue on Econometric Inference Using Simulation Techniques 1993 pp. S85-S118

[7] Gourieroux C. and Monfort, A. "Simulation Based Econometrics in Models with Heterogeneity" 1991 Anales d'Economie et de Statistique, 20/21 69-107

[8] Hansen, L. P. "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica Vol. 50. No 4, (Jul., 1982), pp. 1029-1054

[9] Heaton, J. "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications", Econometrica, Vol. 63, No 3 (May, 1995), pp. 681-717

[10] Lee, B., and B. Ingram "Simulation Estimation of Time Series Models", Journal of Econometrics, 47, 1991, pp.197-205

[11] Mariano, R, Schuermann, T., and Weeks, M "Simulation-based Inference in Econometrics", Cambridge University Press 2000

[12] McFadden, D. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration", 1989 Econometrica pp.995-1026

[13] McFadden, D. and Ruud, P. "Estimation by Simulation", The Review of Economics and Statistics Vol. 76, No 4 1994, pp. 591-608

[14] Pakes, A. and Pollard, D. "Simulation and the Asymptotics of Optimization Estimators" Econometrica, Vol. 57 1989, pp.1027-1057

[15] Perron, Pierre Lecture notes on Advanced Econometrics I, EC 708, 2006, Boston University

[16] Perron, Pierre Lecture notes on Time Series Econometrics, EC 712, 2006, Boston University

[17] Smith, A.A. "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions", 1993 Journal of Applied Econometrics, Vol.8

[18] Tauchen, G. "The Objective Function of Simulation Estimators near the Boundary of the Unstable Region of the Parameter Space", The Review of Economics and Statistics, Vol. 80, No. 3, (Aug. 1998), pp. 389-398