Method of Simulated Moments

Thierry Kamionka

2021-2022

Definition (Stationarity)

Let us consider a stochastic process $\{Y_i\}$. $\{Y_i\}$ is stationary if for any integer r and for any $(k_1, k_2, \ldots, k_r)$, the distribution of $(Y_i, Y_{i+k_1}, Y_{i+k_2}, \ldots, Y_{i+k_r})$ is the same as the distribution of $(Y_j, Y_{j+k_1}, Y_{j+k_2}, \ldots, Y_{j+k_r})$.

Example : An i.i.d. sequence is a stationary stochastic process with no serial dependence: the joint density of $(Y_j, Y_{j+k_1}, Y_{j+k_2}, \ldots, Y_{j+k_r})$ is the product of the marginal densities.

Definition (Weak Stationarity)

Let us consider a stochastic process $\{Y_i\}$. $\{Y_i\}$ is weakly stationary if

$$\mu_i = E[Y_i] = \mu$$

for all i.

The autocovariance function exists, is finite and

$$\gamma(i, i-j) = cov(Y_i, Y_{i-j}) = E[(Y_i - \mu_i)(Y_{i-j} - \mu_{i-j})]$$

only depends on j : $\gamma(i, i-j) = \gamma(j)$.

Definition (Ergodicity)

A **stationary stochastic** process $\{Y_i\}$ is ergodic if for any bounded function $f : \boldsymbol{R}^k \longrightarrow \boldsymbol{R}$ and any bounded function $g : \boldsymbol{R}^{k'} \longrightarrow \boldsymbol{R}$

$$\lim_{n\longrightarrow\infty} \mid E[f(Y_i,\ldots,Y_{i+k})\, g(Y_{i+n},\ldots,Y_{i+n+k'})] \mid$$

$$= \mid E[f(Y_i,\ldots,Y_{i+k})] \mid \mid E[f(Y_{i+n},\ldots,Y_{i+n+k'})] \mid$$

The stationary process $\{Y_i\}$ is ergodic if it is **asymptotically independent.**

Theorem (e.g. Hayashi, 2000)

Let $\{Y_i\}$ be a stationary and ergodic stochastic process, then

$$\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow[a.s.]{} \mu$$

where $\mu = E[Y_i]$.

It is a generalization of the Law of Large Numbers (Kolmogorov). Hence, serial dependence is allowed but cancels asymptotically.

For any function f mesurable, $\{f(Y_i)\}$ is stationary and ergodic. Then, any moment (if it exists and is finite) of Y_i can be estimated consistently using the corresponding **sample average**.

The **method of moment** can also be seen as a semi parametric method. In this case, Y_i is a vector of stationary and ergodic random variables with probability distribution function P_0. The parameter θ_0 is defined by

$$E_{P_0}[h(Y_i, \theta_0)] = 0$$

where θ_0 is the true value of the parameter.

The distribution P_0 is estimated using the empirical distribution and θ can be estimated solving

$$\frac{1}{n}\sum_{i=1}^{n} h(y_i, \theta) = 0$$

with respect to θ or minimizing a norm of this quantity in the overidentified case.

Alternatively, for a **parametric model** index by θ, the method of moments can be considered as an estimation method.

Let us assume that the sample consists in $\{Y_1, \ldots, Y_n\}$, where Y_i is a vector of **stationary and ergodic** random variables.

The variables are distributed according to P_θ where θ is a vector of parameters. The vector Y_i may include a vector of exogenous variables X_i.

The vector of parameters θ is defined by the moment conditions

$$E_\theta[h(Y_i, \theta)] = 0$$

where E_θ is the expectation with respect to the distribution P_θ.

P_θ is estimated by the empirical distribution. The moments conditions are assumed to identify θ

$$E_\theta[h(Y_i, \theta)] = 0 \Longrightarrow \theta = \theta_0$$

Let $h_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} h(y_i;\theta)$. The GMM estimator is a solution of

$$\min_{\theta\in\Theta} h_n(\theta)'\ W_n\ h_n(\theta)$$

where W_n is a positive definite matrix.

If the function $h_n(y_i;\theta)$ is difficult to calculate - for instance, it is a multiple integral - it can be **approximated using simulations.**

Let us assume that there exists a latent variable, namely ξ^*, such that

$$h(y_i;\theta) = E_{\xi^*}[\tilde{h}(Y_i,\xi^*;\theta) \mid y_i]$$

where E_{ξ^*} is the expectation with the conditional distribution of ξ^*.

For some applications, it can provide some gain of efficiency to draw ξ^* conditionally on $(y_1,\ldots,y_n)$ (see, McFadden and Ruud, 1994). Here, we assume ξ^* is drawn from its **marginal distribution.**

Let us consider a **multinomial logit (MNL) model** with **random coefficient** α (see McFadden and Train, 2000). These coefficients are assumed to be distributed according to the pdf $g(\alpha;\theta)$:

$$P_C(j \mid x, \theta) = \int L_C(j \mid x, \alpha)\, g(\alpha;\theta)\, d\,\alpha$$

where

$$L_C(j \mid x, \alpha) = \frac{\exp(x_j \alpha)}{\sum_{k \in C} \exp(x_k \alpha)}$$

with $C = \{1, \ldots, J\}$ is the set of alternatives the individual can chose. Let $x = (x_1, \ldots, x_J)$ where x_j is a $1 \times K$ vector of characteristics of the individual and of attributes of alternative j.

The mixing pdf $g(\alpha;\theta)$ can belong to a continuous parametric family such as multivariate normal or may have a finite support (in this case, this MNL model it is a latent class model).

Let us consider $\alpha = \beta + \Lambda\xi^*$ where β is a $K \times 1$ vector "mean" coefficients, Λ is $K \times M$ matrix of **factor loadings** (with exclusion restrictions for identification) and ξ^* is a $M \times 1$ vector of **factor levels** that are i.i.d. with a density $f(\xi^*)$.

Let $vec(A)$ denote the function that stacks the columns of A into a vector. Let $\gamma = vec(\Lambda')$ and $\theta' = (\beta', \gamma')$ denote the vector of parameters.

Let us consider

$$x_C(\xi^*) = \sum_{k \in C} x_k \, L_C(k \mid x, \beta + \Lambda\xi^*),$$

and

$$x_{jC}(\xi^*) = x_j - x_C(\xi^*).$$

where x_j is a $1 \times K$ vector.

Let $E_{\xi^*|j}$ denote an expectation with respect to the density of ξ^* given that the event j is chosen. This density is

$$\frac{L_C(j \mid x, \beta + \Lambda\xi^*)\ f(\xi^*)}{\int L_C(j \mid x, \beta + \Lambda\xi^*)\ f(\xi^*)\ d\,\xi^*}$$

We have

$$P_C(j \mid x; \theta) = E_{\xi^*}[L_C(j \mid x, \beta + \Lambda\xi^*)],$$

The derivative with respect to β of $\log(P_C(j \mid x; \theta)) = \log(\int L_C(j \mid x, \beta + \Lambda\xi^*) f(\xi^*) d\xi^*)$ is

$$\frac{\int \frac{\partial}{\partial\beta} L_C(j \mid x, \beta + \Lambda\xi^*) f(\xi^*) d\xi^*}{\int L_C(j \mid x, \beta + \Lambda\xi^*) f(\xi^*) d\xi^*}$$

$$= \frac{\int (x_j' - \sum_{k \in C} x_k' L_C(k \mid x, \beta + \Lambda\xi^*))\ L_C(j \mid x, \beta + \Lambda\xi^*) f(\xi^*) d\xi^*}{\int L_C(j \mid x, \beta + \Lambda\xi^*) f(\xi^*) d\xi^*}$$

This derivative is

$$\frac{\int x_{jC}(\xi^*)' \, L_C(j \mid x, \beta + \Lambda\xi^*) f(\xi^*) d\xi^*}{\int L_C(j \mid x, \beta + \Lambda\xi^*) f(\xi^*) d\xi^*}$$

Consequently,

$$\frac{\partial}{\partial\beta} \log(P_C(j \mid x; \theta)) = E_{\xi^*|j}[x_{j\ell C}(\xi^*)']$$

The derivative with respect to $\gamma = vec(\Lambda')$ of $\log(P_C(j \mid x; \theta)) = \log(\int L_C(j \mid x, \beta + \Lambda\xi^*) f(\xi^*) d\xi^*)$ can be obtained staking the elements:

$$\frac{\int \frac{\partial}{\partial \gamma_\ell} L_C(j \mid x, \beta + \Lambda \xi^*) f(\xi^*) d\xi^*}{\int L_C(j \mid x, \beta + \Lambda \xi^*) f(\xi^*) d\xi^*}$$

$$= \frac{\int \xi^* (x_{j\ell} - \sum_{k \in C} x_{k\ell} L_C(k \mid x, \beta + \Lambda \xi^*))\, L_C(j \mid x, \beta + \Lambda \xi^*) f(\xi^*) d\xi^*}{\int L_C(j \mid x, \beta + \Lambda \xi^*) f(\xi^*) d\xi^*}$$

$$= E_{\xi^* \mid j}[\xi^* x_{j\ell C}(\xi^*)]$$

where γ_ℓ are the elements of the line ℓ of Λ ($\ell = 1, \ldots, K$).

Thus,

$$\frac{\partial}{\partial \gamma_\ell} \log(P_C(j \mid x; \theta)) \in \boldsymbol{R}^M$$

Consequently,

$$\frac{\partial}{\partial \gamma} \log(P_C(j \mid x; \theta)) = vec(E_{\xi^* \mid j}[\xi^* x_{jC}(\xi^*)])$$

where $E_{\xi^* \mid j}[\xi^* x_{jC}(\xi^*)]$ is a $M \times K$ matrix.

It is possible to simulate $P_C(j \mid x, \theta)$ and its derivatives in order to estimate the vector of parameters θ.

Let us consider ξ_r^* for $r = 1, \ldots, H$, i.i.d. random draws obtained from the pdf $f(\xi^*)$.

Let us consider the following estimator of $P_C(j \mid x, \theta)$

$$P_{C,H}(j \mid x, \theta) = \frac{1}{H} \sum_{r=1}^{H} L_C(j \mid x; \beta + \Lambda \xi_r^*) \equiv E_H[\, L_C(j \mid x; \beta + \Lambda \xi^*)\,]. \tag{1}$$

$P_{C,H}(j \mid x, \theta)$ is an **unbiased estimator** of $P_C(j \mid x, \theta)$. It is continuous with respect to θ and twice derivable. The derivative of $\log(P_C(j \mid x, \theta))$ are functions of elements

$$E_{\xi^*|j}[\, b(\xi^*)\,] \equiv \frac{E_{\xi^*}[\, b(\xi^*) L_C(j \mid x; \beta + \Lambda \xi^*)\,]}{P_C(j \mid x, \theta)}$$

where, for instance, $b(\xi^*) = \xi^* x_{j\ell C}(\xi^*)$ or $b(\xi^*) = x_{jC}(\xi^*)'$.

The expectation $E_{\xi^*|j}[\, b(\xi^*)\,]$ can be simulated by

$$E^H_{\xi^*|j}[\, b(\xi^*)\,] \equiv \frac{E_H[\, b(\xi^*) L_C(j \mid x; \beta + \Lambda\xi^*)\,]}{P_{C,H}(j \mid x, \theta)}$$

It is a continuously differentiable function of θ at the first and second order.

Let us remark that the simulator $\log(P_{C,H}(j \mid, \theta))$ of $\log(P_C(j \mid, \theta))$ **is not unbiased** due to the non linear transformation $\log(.)$.

All the simulators considered here are consistent when $H \longrightarrow +\infty$.

The second derivatives can be obtained by numerical differentiation of the first derivatives.

When one optimize the objective function with respect to θ, the drawings ξ^*_h should remain fixed. This can be achieved by storing these drawings in memory of by drawing them again using the same seed.

MNL Models - MSM

Remark: SMLE can be obtained maximizing $\sum_{i=1}^{n} \log(P_{C,H}(y_i \mid x;\theta))$, where $P_{C,H}(j \mid x;\theta)$ is given by (1) and $y_i \in C$ is the alternative chosen by individual i.

Let d_j denote a binary variable that is equal to 1 if alternative j is chosen and is equal to 0 otherwise. Let $d = (d_1, \ldots, d_J)$.

The method of moments estimator of θ is such that the generalized residual $d_j - E_{\xi^*}[\, L_C(j \mid x;\beta + \Lambda\,\xi^*)\,]$ is orthogonal to any instrument vector $\omega_j(x;\theta)$ with the same dimension than θ.

The moment is (one observation)

$$h(d, x;\theta) = \sum_{j\in C} (d_j - E_{\xi^*}[\, L_C(j \mid x;\beta + \Lambda\,\xi^*)\,])\,\omega_j(x;\theta)$$

Let us denote $s_j(x;\theta) = \frac{\partial}{\partial\theta}\log(P_C(j \mid x,\theta))$.

Let us consider the function $h(d, x; \theta)$ - for a given individual - under the **restriction** that $\omega_j(x;\theta) = s_j(x;\theta)$

$$h(d, x; \theta) = \sum_{j \in C} (d_j - E_{\xi^*}[\, L_C(j \mid x; \beta + \Lambda\, \xi^*)\,])\ \frac{\partial}{\partial \theta} \log(P_C(j \mid x; \theta))$$

$$= \sum_{j \in C} (d_j - P_C(j \mid x; \theta))\ \frac{\partial}{\partial \theta} \log(P_C(j \mid x; \theta))$$

$$= \sum_{j \in C} \left(d_j\ \frac{\partial}{\partial \theta} \log(P_C(j \mid x; \theta)) - \frac{\partial}{\partial \theta} P_C(j \mid x; \theta) \right)$$

$$= \sum_{j \in C} d_j\ \frac{\partial}{\partial \theta} \log(P_C(j \mid x; \theta)) - \sum_{j \in C} \frac{\partial}{\partial \theta} P_C(j \mid x; \theta)$$

$$= \sum_{j \in C} d_j\ \frac{\partial}{\partial \theta} \log(P_C(j \mid x; \theta)) - \frac{\partial}{\partial \theta} \sum_{j \in C} P_C(j \mid x; \theta)$$

Consequently,

$$h(d, x; \theta) = \sum_{j \in C} d_j \frac{\partial}{\partial \theta} \log(P_C(j \mid x; \theta))$$

Let us consider the **contribution** of a typical individual to the log likelihood

$$\log(L(\theta)) = \sum_{j=1}^{J} d_j \, \log(P_C(j \mid x; \theta))$$

The score function for this individual is

$$\frac{\partial}{\partial \theta} \log(L(\theta)) = \sum_{j=1}^{J} d_j \frac{\partial}{\partial \theta} \log(P_C(j \mid x; \theta))$$

If $\omega_j(x; \theta) = s_j(x; \theta)$, **then** $h(d, x, \theta)$ is the score of an observation. Then, the method of **moments estimator (MME)** is the **maximum likelihood estimator (MLE)**.

Any vector of instrument $\omega_j(x;\theta)$ such that the covariance with $s_j(x;\theta)$ is of maximum rank can be used (McFadden and Train, 2000) and allows to obtain an **MM estimator** that is **consistent** and **asymptotically normal** (but not in general efficient).

In order to obtain the MSM estimator, the conditional probability $P_C(j \mid x;\theta)$ is replaced by the simulator $P_{C,H}(j \mid x;\theta)$ (see (1)) and we use independent and identically distributed drawings of ξ^*. Let $h^H(d,x;\theta)$ denote the simulator of $h(d,x;\theta)$ we obtain.

The **Method of Simulated Moments (MSM)** estimator of θ is a root of

$$\frac{1}{n}\sum_{i=1}^{n} h^H(d^i, x_i;\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J}(d_j^i - P_{CH}(j \mid x_i;\theta))\,\omega_j(x_i;\theta)$$

where $d^i = (d_1^i, \ldots, d_J^i)$ and $d_j^i = 1$ if individual choses alternative j and $d_j^i = 0$ otherwise.

Let us assume that $\hat{\theta}_{n,H}$ is a root of $\frac{1}{n}\sum_{i=1}^{n} h^H(d^i, x_i; \theta)$.

Under regularity conditions, McFadden (1989, 1996) shows that the MSM estimator is **consistent and asymptotically normal**.

It is not necessary to obtain such a result that H (number of draws) increases with n as long as the **simulators of the generalized residual are independent**.

The Method of Simulated Moments (MSM) estimator is obtained here by replacing the conditional probabilities $P_c(j \mid x; \theta)$ by an unbiased simulator and by using independent simulators to obtain the instruments $\omega_j(x_i; \theta)$.

Iteration: The optimization can be done by estimating θ using crude instruments, as they do not depend of θ, in order to obtain a consistent estimator $\hat{\theta}$. Then one can simulate the suggested instruments for this particular value of θ. And, then, one may obtain a new value of the estimator of θ.

The asymptotic covariance matrix of the MSM of θ (namely, $\hat{\theta}_{n,H}$) is consistently estimated by

$$\Psi_{n,H}(\hat{\theta}_{n,H})^{-1}\ \Sigma_{n,H}(\hat{\theta}_{n,H})\ \Psi_{n,H}(\hat{\theta}_{n,H})^{-1}$$

where

$$\Psi_{n,H}(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in C}(\frac{\partial}{\partial\theta}P_{C,H}(j \mid x_i;\theta))\omega_j(x_i;\theta)'$$

$$\Sigma_{n,H}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j\in C}\omega_j(x_i;\theta)\,P_{C,H}(j \mid x_i;\theta))\omega_j(x_i;\theta)'\right.$$

$$\left. - [\sum_{j\in C}\omega_j(x_i;\theta)\,P_{C,H}(j \mid x_i;\theta)]\,[\sum_{j\in C}\omega_j(x_i,\theta)\,P_{C,H}(j \mid x_i;\theta)]'\right)$$

The SMLE is asymptotically efficient (if $H/\sqrt{n} \longrightarrow \infty$ as $n \longrightarrow \infty$). But it is difficult computationally to obtain (H should be large).

The more $\omega_j(x;\theta)$ is correlated with $s_j(x;\theta) = \frac{\partial}{\partial\theta}\log(P_C(j \mid x;\theta))$, the more efficient is the MSM estimator. Consequently, a relatively large number of draws H should be obtained in order that the simulated score $s_j^H(x;\theta)$ approximate $s_j(x;\theta)$ accurately, where

$$s_j^H(x;\theta) = \frac{\partial}{\partial\theta}\log(P_{C,H}(j \mid x;\theta)).$$

$\square$

The moment conditions are the following ones

$$E_{\theta_0}[\,h(Y_i,\theta)\,] = 0.$$

We assume there exists a latent variable ξ^* such that

$$h(y_i;\theta) = E_{\xi^*}[\tilde{h}(y_i,\xi^*;\theta)]$$

The distribution of the latent variable is assumed to be such that it is easy to simulate. Let $\tilde{h}_{i,H}(\theta)$ denote the simulator of the function $h(y_i,\theta)$. We have

$$\tilde{h}_{i,H}(\theta) = \frac{1}{H}\sum_{r=1}^{H} \tilde{h}(y_i,\xi^*_{i,r};\theta)$$

where $\xi^*_{i,1},\ldots,\xi^*_{i,H}$ are **independent identically distributed** according to the distribution of ξ^*.

The **MSM estimator** is a solution to

$$\min_{\theta\in\Theta}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{h}_{i,H}(\theta)\right)' \; W \; \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{h}_{i,H}(\theta)\right)$$

The draws $\xi^*_{i,r}$ are such that we can use the same values whatever the value of θ. In practice, the definition of the function $\tilde{h}$ we use is such that the distribution of the latent variable ξ^* do not depend on θ.

Example: Mixed MNL models

Let us assume that $Y = (d, x)$ and

$$\tilde{h}(y, \xi^*; \theta) = \sum_{j\in C}(d_j - L_C(j \mid x, \beta + \Lambda\xi^*))\,\omega_j(x;\theta)$$

On can verify that

$$h(y;\theta) = E_{\xi^*}[\tilde{h}(y,\xi^*;\theta)] = \sum_{j\in C}(d_j - E_{\xi^*}[\,L_C(j \mid x, \beta+\Lambda\xi^*)\,])\,\omega_j(x;\theta) \quad \square$$

Static case - Importance Sampling

Let ξ^* denote a latent variable. Let us assume we need to evaluate an expectation, namely

$$E_{\xi^*}(b(\xi^*))$$

using simulations. Let F denote the cdf of the distribution of ξ^*. We can obtain draws from F by inverting the cdf for instance. Example: Mixed MNL models

$$b(\xi^*) \equiv \tilde{h}(y, \xi^*; \theta) = \sum_{j \in C} (d_j - L_C(j \mid x; \beta + \Lambda \xi^*)) \, \omega_j(x; \theta) \qquad \square$$

Then the expectation can be estimated using

$$\hat{E}_{\xi^*}(b(\xi^*)) = \frac{1}{H} \sum_{r=1}^{H} b(\xi_r^*)$$

where $\xi_1^*, \ldots, \xi_H^*$ are i.i.d. draw from the distribution F.

If it is difficult to simulate in the distribution corresponding to F, we may use **importance sampling** . Let us remark that

$$E_{\xi^*}(b(\xi^*)) = \int \frac{b(\xi^*)f(\xi^*)}{g(\xi^*)} g(\xi^*)\, d\,\xi^*$$

where g is a pdf with the same support than the pdf f.

The importance sampling simulator is

$$\hat{E}_{\xi^*}(b(\xi^*)) = \frac{1}{H} \sum_{r=1}^{H} \frac{b(\xi_r^*)f(\xi_r^*)}{g(\xi_r^*)}$$

where $\xi_1^*, \ldots, \xi_H^*$ are i.i.d. **draws from the distribution with pdf** g.

Dynamic case - Path simulation

To illustrate the context, let us consider the **Markov model of asset prices** proposed by Duffie and Singleton (1993).

The production of the consumption commodity is fixed by the equation

$$F(k_t, z_t) = z_t \, k_t^{\phi}, \text{ where } 0 < \phi < 1,$$

where F is a production function, k_t is a level of capital stock at time t, z_t is a technology shock and ϕ is a parameter.

For each period t, the firm optimizes its profits by choosing the level of capital to rent from consumer

$$d_t = \max_{k_t} \{ z_t \, k_t^{\phi} - r_t^k k_t \}$$

where r_t^k is the **rental rate** and d_t represents the **dividends** that the firm pays to owners of the shares.

Let p_t denote the **shares market value of the firm** (capitalization).

The **budget constraint of the consumer** is

$$1 \times c_t + p_t(s_{t+1} - s_t) + (k_{t+1} - \mu k_t) = d_t s_t + r_t^k k_t \qquad (2)$$

where c_t denote consumption at time t, s_t is the **share of the capital of the firm** owned by the representative consumer (asset shares) and $(1 - \mu)$ is the **depreciation rate** on the capital stock.

(similar to an economy who produces wheat for consumption and as input)

The **consumer maximizes** the expectation of the utility stream with respect to the capital he rents to the firm and consumption. Using the budget constraint, one can deduce the share holdings.

The **consumer maximizes** under (2)

$$\max_{\{c_t,k_t\}} E\left[\sum_{t=1}^{\infty} \delta^t \frac{(c_t-1)^{1-\alpha}}{1-\alpha} u_t\right]$$

where α represents the **coefficient of the relative risk aversion** (CRRA), δ is a **subjective discount factor** $(0 < \delta < 1)$ and u_t is a taste shock.

Let us assume that $X_t' = (z_t, u_t)$ is a Markov process, such that

$$X_t = \kappa(X_{t-1}, \epsilon_t, \rho_0)$$

where $\epsilon_t \in \boldsymbol{R}^2$, is an **i.i.d. stochastic process**, κ is a transition function and ρ_0 is a vector of parameters.

For instance, Michner (1984), assumes that $u_t = 1$ and $\ln(z_{t+1}) = \zeta_z + \rho \ln(z_t) + \epsilon_{t+1}$, where $\{\epsilon_t\}$ is an i.i.d. normal.

In order to estimate the vector of parameters $\beta = (\phi, \alpha, \rho_0, \mu, \delta)'$, **the model is solved** in order to determine the equilibrium transition function

$$Y_{t+1} = \lambda(Y_t, \epsilon_{t+1}, \beta)$$

where $Y_t = (X_t', k_t)'$ is the augmented state process.

For any value of the vector of parameters β, it is possible to obtain a simulated state process $\{\hat{Y}_t\}$ using the transition function

$$\hat{Y}_{t+1} = \lambda(\hat{Y}_t, \hat{\epsilon}_{t+1}, \beta)$$

where $\{\hat{\epsilon}_t\}$ is an **i.i.d. sequence** of $\{\epsilon_t\}$.

From $\{\hat{Y}_t\}$, a history $\{\hat{Y}_t\}_{t=1}^{\mathcal{T}}$ of $\mathcal{T}$ equilibrium states can be generated.

Let $f_t^* = s(Y_t, Y_{t-1}, \ldots, Y_{t-\ell+1})$, where s is an 'observation function' made of a finite history of Y_t.

Example : For instance, a component of $s(Y_t, Y_{t-1}, \ldots, Y_{t-\ell+1})$ can be $k_t \times k_{t-1}$, an other can be c_t.

Let $\hat{f}_t = s(\hat{Y}_t, \hat{Y}_{t-1}, \ldots, \hat{Y}_{t-\ell+1})$, is a function of the history of $\hat{Y}_t$.

Intuition : The MSM estimator is a value of β that matches the sample moments of the actual observation process $\{f_t^*\}$ and the sample moments of the simulated observation process $\{\hat{f}_t\}$.

Remark : Under a different set of assumptions, if it is difficult to simulate a sequence $\{\hat{\epsilon}_t\}$, one can consider to use an importance function (Kamionka, 1998).

Example: Michner (1984)

The **taste shock** is such that $u_t = 1$ for all t.

Let us assume we have a 100% **depreciation rate** ($\mu = 0$).

The utility is logarithmic ($\alpha = 1$).

The equilibrium asset-pricing function and the evolution of the capital stock are

$$p_t = \frac{\delta}{(1-\delta)} \, (1-\phi) \, z_t \, k_t^{\phi},$$

$$d_t = (1-\phi) \, z_t \, k_t^{\phi},$$

$$k_{t+1} = \delta \, \phi \, z_t \, k_t^{\phi}.$$

Let Y_i denote a vector of ergodic stationary random variables.

These random variables are distributed according to the distribution P_θ, where θ is a vector of parameters.

The vector Y_i can include, depending on the case, a vector of exogenous variables, namely X_i. Let E_{P_θ} denote the expectation with respect to the distribution P_θ.

Identification : $E_{P_{\theta_0}}[\,h(Y_i;\theta)\,] = 0 \Longrightarrow \theta = \theta_0$ (θ_0 is the **solution of the asymptotic optimization program**).

Let us assume that

$$h(Y_i,\theta) = E_{\xi^*}[\,\tilde{h}(Y_i,\xi^*;\theta) \mid Y_i\,],$$

where $\tilde{h}()$ is a known function and ξ^* is a random variable such that its distribution is known and does not depend on θ. ξ^* is independent of $\{Y_i,\ i = 1,\dots,n\}$.

Assumption: $\{\tilde{h}(Y_i,\xi^*;\theta)\}$ is **stationary and ergodic.**

We generate $\{\xi^*_{i,r},\ i = 1, \ldots, n\}_{r=1,\ldots,H}$ i.i.d. using the same distribution as $\{\xi^*_i,\ i = 1, \ldots, n\}$.

Consequently, $\{\tilde{h}(Y_i, \xi^*_{ir}; \theta)\}_{r=1,\ldots,H}$ are i.i.d. conditional on Y_i.

Let us remind that the **MSM** estimator is a solution of

$$\min_{\theta\in\Theta} \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{h}_{i,H}(\theta)\right)' W_n \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{h}_{i,H}(\theta)\right)$$

where $\tilde{h}_{i,H}(\theta) = \frac{1}{H}\sum_{r=1}^{H} \tilde{h}(Y_i, \xi^*_{ir}; \theta)$.

Let us assume that $h_i = h_i(Y_i, \theta_0)$.

$$\hat{\theta}_n = \underset{\theta\in\Theta}{argmin}\, C_n(\theta)$$

where $C_n(\theta) = \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{h}_{i,H}(\theta)\right)' W_n \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{h}_{i,H}(\theta)\right)$.

$$\Sigma_0 = \sum_{k=-\infty}^{\infty} Cov_{P_{\theta_0}}(h_i, h_{i-k})$$

$$= \sum_{k=-\infty}^{\infty} E_{P_{\theta_0}}(h_i - E_{P_{\theta_0}}(h_i))\,(h_{i-k} - E_{P_{\theta_0}}(h_{i-k}))',$$

Assumption: Σ_0 is non singular and $W_n \longrightarrow W = \Sigma_0^{-1}$ almost surely.

Let us assume that H is fixed and regularity conditions are satisfied (see Duffie and Singleton, 1993).

Theorem (Consistency)

Under these regularity conditions, the MSM estimator $\hat{\theta}_n$ converges to θ_0 in probability as $n \longrightarrow \infty$.

Let us assume that θ_0 and the sequence of MSM estimators $\hat{\theta}_n$ are interior to the parameters space Θ.

Let us assume that the matrix

$$D_0 = E_{P_{\theta_0}}\left[\frac{\partial h_i}{\partial \theta'} \right] = E_{P_{\theta_0}}\left[\frac{\partial}{\partial \theta'} E_{\xi^*}[\tilde{h}(Y_i, \xi^*; \theta_0) \mid Y_i] \right]$$

exists, is finite and has a full rank.

Let us assume $\tilde{h}(Y_i, \xi^*; \theta)$ is continuously differentiable with respect to θ.

The objective function $C_n(\theta)$ converges almost surely to the asymptotic objective function $C(\theta)$. **Let us assume** that $C(\theta) > C(\theta_0)$,for all $\theta \in \Theta$ and $\theta \neq \theta_0$.

Remark : If the model is correct at θ_0 then $E[h(y; \theta_0)] = E[\tilde{h}(y, \xi^*; \theta_0)]$.

Let us assume some additional regularity conditions (see Duffie and Singleton, 1993).

Theorem (Asymptotic normality, Under optimal weighting)

Under these regularity conditions and for H fixed, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{d} N(0, (1 + \frac{1}{H})(D_0'\Sigma_0^{-1}D_0)^{-1})$$

The covariance matrix is similar to the expression obtained by McFadden (1989) and Pakes and Pollard (1989).

When H is large, the asymptotic covariance matrix is close to $(D_0'\Sigma_0^{-1}D_0)^{-1}$.

Other set of assumptions : Lee and Ingram (1991)

Let us assume some additional regularity conditions. **Let us assume** that the weighting matrix W_n **converges** in probability to W. Let us remark that the total number of simulations is $n \times H$ and tends to infinity as n tends to infinity.

Theorem (Asymptotic normality)

Under these regularity conditions and for H fixed, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{d} N(0, (D_0' W D_0)^{-1} D_0' W (1 + \frac{1}{H}) \Sigma_0 W D_0 (D_0' W D_0)^{-1})$$

Optimal choice for $W = [(1 + \frac{1}{H})\Sigma_0]^{-1}$.

In order to obtain a **consistent estimator** of Σ_0, see Newey and West, 1987.

Let us assume the objective function we minimize is such that

$$\min_{\theta \in \Theta} h_{n,H}(\theta)' \; W_n \; h_{n,H}(\theta)$$

where

$$h_{n,H}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \tilde{h}_{i,H}(\theta)$$

and

$$\tilde{h}_{i,H}(\theta) = \frac{1}{H} \sum_{r=1}^{H} \tilde{h}(Y_i, \xi^*_{i,r}; \theta)$$

Let us denote $\hat{h}_i = \tilde{h}_{i,H}(\hat{\theta}_n)$.

Let m denote the **number of nonzero autocorrelation** of $\tilde{h}_{i,H}(\theta_0)$.

Let us denote

$$\hat{\Omega}_k = \frac{1}{n} \sum_{i=1+k}^{n} \hat{h}_i \hat{h}'_{i-k}$$

An estimator of Σ_0 is

$$\tilde{\Sigma}_0 = \hat{\Omega}_0 + \sum_{k=1}^{m} (\hat{\Omega}_k + \hat{\Omega}'_k).$$

Remark : In the particular case of an i.i.d. sample, $\tilde{\Sigma}_0 = \hat{\Omega}_0$.

However, the estimator of the asymptotic variance-covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ obtained using $\tilde{\Sigma}_0$ need not be positive semi-definite. Alternatively, let us consider

$$\hat{\Sigma}_0 = \hat{\Omega}_0 + \sum_{k=1}^{m} w(k,m)\ (\hat{\Omega}_k + \hat{\Omega}'_k),$$

where $w(k,m) = 1 - (k/(1+m))$.

Let us assume we observe (y_i, x_i), $i = 1, \ldots, n$. x_i is a vector of exogenous variables. Let us consider the following objective function

$$Q_n(\underline{y}_n, \underline{x}_n; \lambda)$$

where $\underline{y}_n = (y_1, \ldots, y_n)$ and $\underline{x}_n = (x_1, \ldots, x_n)$.

Q_n is the log of a **quasi-likelihood function**.

Let $\hat{\lambda}$ is a solution of

$$\max_{\lambda} Q_n(\underline{y}_n, \underline{x}_n; \lambda)$$

Consequently, $\hat{\lambda}$ can be considered as a Quasi Maximum Likelihood estimator (QLME).

Let us denote $y_i(\theta_0) \equiv y_i$, for $i = 1, \ldots, n$, in order to insist on the fact that the realizations were obtained from the true model (indexed by θ_0).

The **Efficient Method of Moment estimator** (EMM estimator) is such that the expectation with respect to the true distribution of the vector of **pseudo score** should be equal to zero (Gallant and Tauchen, 1996)

$$\min_{\theta \in \Theta} \| \, E_{P_\theta}[\, \frac{\partial}{\partial \lambda} Q_n(\underline{y}_n(\theta), \underline{x}_n; \hat{\lambda}) \,] \, \|$$

where $y_i(\theta)$ is distributed according to P_θ (**true family of distribution** for a hypothetical value of θ).

We generate, for each value of θ, H samples of size n, $\underline{y}_n^r = (y_1^r, \dots, y_n^r)$, according to the distribution P_θ, conditionally on $\underline{x}_n = (x_1, \dots, x_n)$.

In practice, the expectation in the objective function in replaced by an average obtained using the simulated data $\underline{y}_n^r$, for $r = 1, \dots, H$.

The EMM estimator is a solution of

$$\min_{\theta \in \Theta} \left(\frac{\partial}{\partial \lambda'} \frac{1}{H} \sum_{r=1}^{H} Q_n(\underline{y}_n^r(\theta), \underline{x}_n; \hat{\lambda}) \right) W_n \left(\frac{\partial}{\partial \lambda} \frac{1}{H} \sum_{r=1}^{H} Q_n(\underline{y}_n^r(\theta), \underline{x}_n; \hat{\lambda}) \right)$$

Remarks : The EMM estimator is a **method of moments estimator** as it is based on moment conditions (expectation of pseudo score is zero). The information of the original sample in "included" in the value of $\hat{\lambda}$ and in the vector $\underline{x}_n$.

Example: Let us assume that $x_i = 1$ and y_i, $i = 1, \ldots, n$, are i.i.d. and distributed as an exponential random variable with parameters $\theta > 0$.

$$f(y_i \mid x_i; \theta) = \theta \exp(-\theta\, y_i)$$

f is the pdf of the **true distribution**.

We consider a normal distribution with mean λ and variance 1 for the **auxiliary model**

$$g(y_i \mid x_i; \lambda) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(y_i - \lambda)^2)$$

The objective function is the log-quasi likelihood

$$Q_n = -\frac{1}{2}\sum_{i=1}^{n}(y_i - \lambda)^2 - n \ln(\sqrt{2\pi})$$

The QMLE is $\hat{\lambda} = \sum_{i=1}^{n} y_i/n = \bar{y}_n$. $\hat{\lambda}$ is obtained by maximizing Q_n with respect to λ.

Simulations: Let us **assume** that $z_{i,r}$ are i.i.d. uniform on (0,1) for $i = 1, \ldots, n$ and $r = 1, \ldots, H$.

Then $y_i^r(\theta) = -\frac{1}{\theta}\ln(1 - z_{i,r})$ for $i = 1, \ldots, n$ and $r = 1, \ldots, H$.

The EMM estimator ($W = 1$) is solution of the program

$$\hat{\theta}_{EMM} = \underset{\theta \in \boldsymbol{R}^+}{argmin}\, (\frac{1}{H}\sum_{r=1}^{H}\sum_{i=1}^{n}(y_i^r(\theta) - \hat{\lambda}))^2$$

$$\hat{\theta}_{EMM} = \underset{\theta \in \boldsymbol{R}^+}{argmin}\, (\frac{1}{H}\sum_{r=1}^{H}\sum_{i=1}^{n}(-\frac{1}{\theta}\ln(1 - z_{i,r}) - \bar{y}_n))^2$$

$$\hat{\theta}_{EMM} = -\frac{\sum_{r=1}^{H}\sum_{i=1}^{n}\ln(1 - z_{i,r})}{H\sum_{i=1}^{n} y_i}$$

Let us remark that

$$\hat{\theta}_{EMM} = \frac{\dfrac{1}{nH}\displaystyle\sum_{r=1}^{H}\sum_{i=1}^{n} - \ln(1 - z_{i,r})}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i}$$

and

$$\frac{1}{nH}\sum_{r=1}^{H}\sum_{i=1}^{n}[-\ln(1 - z_{i,r})] \underset{n,H\to\infty}{\xrightarrow{a.s.}} E(D)$$

where $D \sim$ exponential variable with mean 1.

We know, moreover, that

$$\frac{1}{n}\sum_{i=1}^{n} y_i \underset{n\to\infty}{\xrightarrow{a.s.}} E(Y_i \mid x_i) = \frac{1}{\theta_0}$$

Then (Slutsky)

$$\hat{\theta}_{EMM} \underset{n,H\to\infty}{\overset{a.s.}{\longrightarrow}} \theta_0.$$

Remarks: Here $\hat{\lambda}$ is such that $\hat{\lambda} = \bar{y}_n$ tends almost surely to $1/\theta_0$. This is closely related to the fact that the quasi likelihood belongs to an exponential linear family. The **binding function** is $\lambda(\theta) = 1/\theta$ (see Indirect inference). $\hat{\lambda}$ tends almost surely to $\lambda(\theta_0)$ as n tends to ∞.

□

Let us **assume** that the objective function tends asymptotically to a limit (the log-quasi likelihood is multiplied by 1/n)

$$\frac{1}{H}\sum_{r=1}^{H} Q_n(\underline{y}_n^r(\theta), \underline{x}_n, \lambda) \underset{n\to\infty}{\longrightarrow} Q_\infty(\lambda, \theta)$$

Let us **denote**

$$\lambda(\theta) = \underset{\lambda}{argmax}\, Q_\infty(\lambda, \theta)$$

where $\lambda(\theta)$ is called the **binding function.**

Let us denote

$$I_0 = \lim_{n\to\infty} V_{P_{\theta_0}}[\sqrt{n}\frac{\partial}{\partial\lambda}Q_n(\underline{y}_n, \underline{x}_n; \lambda_0) - E_{P_{\theta_0}}[\sqrt{n}\frac{\partial}{\partial\lambda}Q_n(\underline{y}_n, \underline{x}_n; \lambda_0) \mid \underline{x}_n]\,]$$

$$J_0 = \underset{n\to\infty}{plim} -\frac{\partial^2}{\partial\lambda\partial\lambda'}Q_n(\underline{y}_n, \underline{x}_n; \lambda_0)$$

The EMM estimator and the Indirect Inference estimator are asymptotically equivalent (Gourieroux, Monfort, Renault, 1993)

Let us assume W_n is a positive definite matrix and converges to a positive definite matrix W.

Theorem (Gourieroux, Monfort, Renault, 1993)

Under some regularity conditions, the EMM estimator is consistent and asymptotically normal (H is fixed and n tends to ∞*)*

$$\sqrt{n}(\hat{\theta}_{EMM} - \theta_0) \xrightarrow[n\to\infty]{d} N(0, V)$$

where

$$V = \left(1+\frac{1}{H}\right)\left(\frac{\partial}{\partial\theta}\lambda(\theta_0)'\; W\; \frac{\partial}{\partial\theta'}\lambda(\theta_0)\right)^{-1}$$

$$\left(\frac{\partial}{\partial\theta}\lambda(\theta_0)'\; W\; J_0^{-1}\; I_0\; J_0^{-1}\; W\; \frac{\partial}{\partial\theta'}\lambda(\theta_0)\right)$$

$$\left(\frac{\partial}{\partial\theta}\lambda(\theta_0)'\; W\; \frac{\partial}{\partial\theta'}\lambda(\theta_0)\right)^{-1}$$

Theorem

Under some regularity conditions, for the optimal matrix $W^* = J_0 I_0^{-1} J_0$, the EMM estimator is asymptotically normal

$$\sqrt{n}(\hat{\theta}_{EMM} - \theta_0) \xrightarrow[n \to \infty]{d} N(0, V^*)$$

where

$$V^* = \left(1 + \frac{1}{H}\right) \left(\frac{\partial}{\partial \theta} \lambda(\theta_0)' \, J_0 \, I_0^{-1} \, J_0 \, \frac{\partial}{\partial \theta'} \lambda(\theta_0)\right)^{-1}$$

The expression of the asymptotic covariance matrix V^* includes the derivative of the binding function $\lambda(\theta)$ calculated for the true value of θ.

It is possible possible to show (Gourieroux and Monfort, 1996) that

$$V^* = \left(1+\frac{1}{H}\right)\left(\frac{\partial^2}{\partial\theta\partial\lambda'}Q_\infty(\lambda_0,\theta_0)\; I_0^{-1}\frac{\partial^2}{\partial\lambda\partial\theta'}Q_\infty(\lambda_0,\theta_0)\right)^{-1}$$

where $\lambda_0 = \lambda(\theta_0)$.

Carrasco, M., Florens, J.-P., 2002, Simulation-based method of moments and efficiency, Journal of Business and Economic Statistics, 20(4), 482-492.

Duffie, D., Singleton, K., 1993, Simulated moments estimation of Markov models of asset prices, Econometrica, 61(4), 929-952.

Gallant, A.R., Tauchen, G., 1996, Which moments to match?, Econometric Theory, 12, 657-681.

Gourieroux, Ch., Monfort, A., Renault, E., 1993, Indirect Inference, Journal of Applied Econometrics, S85-S118.

Gourieroux, Ch., Monfort, A., 1996, Simulation Based Econometrics Methods, Core Lecture Series, Louvain-La-Neuve.

Hayashi, F., 2000, Econometrics, Princeton University Press.

Kamionka, T., 1998, Simulated maximum likelihood estimation in transition models, Econometrics Journal, 1, C129-C153.

Lee, B.-S., Ingram, B.F., 1991, Simulation estimation of times-series models, 47, 197-205.

McFadden, D., 1989, A method of simulated moments for estimation of discrete choice models without numerical integration, Econometrica, 57, 995-1026.

McFadden, D., 1996, Lectures on simulation-assisted statistical inference, EC-squared Conference, Florence, Italy December 12-14.

McFadden, D., Train, K., 2000, Mixed MNL models for discrete response, Journal of Applied Econometrics, 15, 447-470.

McFadden, D., Ruud, P., 1994, Estimation by simulation, Review of Economics and Statistics, 76, 591-608.

Michener, R., 1984, Permanent income in general equilibrium, Journal of Monetary Economics, 13, 297-305.

Newey, K., West, K.D., 1987, A simple, Positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica, 55(3), 703-708.

Pakes, A., Pollard, D., 1989, Simulation and asymptotics of optimization estimators, Econometrica, 57, 1027-1057.