Simulation Based Method of Moments and Efficiency

Marine Carrasco
The University of Rochester

Jean-Pierre Florens
Université de Toulouse

June 14, 2002

Abstract

The method of moments is based on a relation $E^{\theta_0}\left(h\left(X_t,\theta\right)\right)=0$ from which an estimator of θ is deduced. In many econometric models, the moment restrictions can not be evaluated numerically due, for instance, to the presence of a latent variable. Monte Carlo simulations method make possible the evaluation of the GMM criterion. This is the basis for the Simulated Method of Moments. Another approach consists in defining an auxiliary model and find the value of the parameters that minimizes a criterion based either on the pseudo score (Efficient Method of Moment) or on the difference between the pseudo-true value and the quasi-maximum likelihood estimator (Indirect Inference). If the auxiliary model is sufficiently rich to encompass the true model then these two methods deliver an estimator that is asymptotically as efficient as the maximum likelihood estimator.

KEY WORDS: Asymptotic efficiency; Efficient Method of Moments; Indirect Inference; Monte Carlo integration.

1. Introduction

Method of moments estimation can be used in different statistical contexts. First, this method may be seen as a semi-parametric approach. In that case, the moment conditions define the parameters of interest and provide an estimation procedure. Assume that the observed data $x_t, t=1,...,T$ are stationary and ergodic with common probability distribution F_0 and that the parameter of interest θ_0 is defined by the condition:

$$E^{F_0}\left(h(X,\theta_0)\right)=0 \tag{1.1}$$

This condition formalizes the structural economic model and θ_0 is then defined implicitly as a function of F_0. In (1.1), F_0 is estimated by the empirical distribution and θ_0 is estimated as the solution of $\frac{1}{T}\sum_{t=1}^{T}h(x_t,\theta)=0$ or, in the overidentified case, as the argument of the minimization of a suitable norm of $\frac{1}{T}\sum_{t=1}^{T}h(x_t,\theta)$ (see

Hansen (1982)). To implement this method, a model need not be fully specified. Second, the method of moments may be also considered as an estimation procedure in a pure parametric context. Assume that θ_0 is the true value of the parameter θ and that maximum likelihood is not computationally feasible. However, economic theory provides moment conditions for which $E^{\theta_0}(h(X,\theta)) = 0$ has a unique solution $\theta = \theta_0$. Implementation of the generalized method of moments (GMM) proceeds as above: the expectation with respect to the true distribution is replaced by an empirical mean and a norm of this mean is minimized in θ.

The first objective of our paper is to study moment estimation in cases where the implementation of this method requires simulations of random numbers. It should be noted that our discussion is confined to parametric simulations *i.e.* simulations generated by a sampling distribution or by importance sampling where parameters are fixed. These simulations are implemented as part of the GMM estimation (and the properties of the estimators are derived from asymptotic theory). "Nonparametric bootstrap" simulations where data are drawn from the empirical distribution, possibly weighted or transformed, are not considered in this paper, this type of simulation is used to improve the approximation of the distribution of the estimators. A consequence of this restriction is that our presentation is necessarily restricted to parametric models.

Two types of applications of simulation methods are examined in this article. The first case is an application of Monte Carlo numerical integration to GMM estimation where the evaluations of the function h requires an integration. If this integration cannot be done by analytic computations, it should be replaced by a numerical evaluation, in particular by a random approximation based on simulations. The objective of the theory is then to show under which conditions this numerical approximation does not affect the consistency of the estimator and how the asymptotic variance is modified. The second type of application of simulation methods concerns the case where the moment condition involves an auxiliary parameter λ. Let us consider a moment equation

$$E^{\theta_0}(h(X,\lambda)) = 0 \tag{1.2}$$

such that $\lambda_0 = \tilde{\lambda}(\theta_0)$ is the unique solution. The estimator $\hat{\lambda}$ of λ deduced from (1.2) converges to in pseudo true value $\tilde{\lambda}(\theta_0)$ also called binding function. If $\tilde{\lambda}$ is a one to one mapping, it may be inverted in order to estimate θ_0 by $\tilde{\lambda}^{-1}(\hat{\lambda})$. In case of overidentification, θ may be estimated by minimizing a norm of $\hat{\lambda} - \tilde{\lambda}(\theta)$. This procedure is called Indirect Inference. If $\tilde{\lambda}(\theta)$ is unknown, it can be approximated by simulations. Another approach consists in minimizing $\left\| E^{\theta}(h(X,\lambda)) \right\|$ with respect to θ where the expectation is approximated by an average over simulated data. This method is called the efficient method of moments (EMM).

However, the EMM addresses a different question which is actually the second objective of this paper. If θ, a vector of parameters, can be estimated using a GMM procedure (direct or indirect), this estimator is necessarily not more efficient than the maximum likelihood estimator (MLE). A natural question is then to search under which conditions GMM estimation reaches the efficiency bound. A necessary and sufficient condition is that the score function is an element of the closed linear space spanned by the elements of the vector of functions h. This suggests using an auxiliary model that has an increasing number of terms.

This paper is organized as follows. Section 2 presents various examples of applications. In the third section, we present the method of simulated moments. In the fourth section, we review the methods based on a misspecified moment, namely the indirect inference and the efficient method of moments. The fifth section discusses specification tests and small-sample properties of the estimators presented in this paper. Section 6 concludes.

2. Examples

2.1. Mixed multinomial logit model

In a mixed multinomial logit (MMNL) model, the agent makes a decision $c \in C = \{1, 2, ..., K\}$ with a conditional probability

$$P_C(c|x, \theta) = \int L_C(c; x, \alpha) G(d\alpha; \theta)$$

where

$$L_C(c; x, \alpha) = \frac{e^{x_c \alpha}}{\sum_{d \in C} e^{x_d \alpha}}.$$

This probability is conditional on an observed vector of characteristics $x = (x_1, ..., x_K)$. α is a vector of random parameters with distribution G. McFadden and Train (2000) apply this model to the demand for vehicles. Assume $\alpha = \beta + \Lambda\xi$ where ξ has a distribution that does not depend on the parameters of interest, P_C can be rewritten as

$$P_C(c|x, \theta) = E_\xi L_C(c; x, \beta + \Lambda\xi).$$

Let d_c denote an indicator that is one when c is chosen and zero otherwise. Let $d = (d_1, ..., d_K)$. Unconditional moments can be built using any instrument $\omega(x, \theta)$:

$$h(d, x; \theta) = \sum_{c \in C} \{d_c - E_\xi L_C(c; x, \beta + \Lambda\xi)\} \omega(x, \theta). \tag{2.1}$$

As $E_\xi L_C(c; x, \beta + \Lambda\xi)$ involves a multiple integral that may be hard to solve analytically, it is estimated by simulations. This is the basis for the Method of Simulated Moments (MSM).

2.2. Dynamic programming model

A dynamic asset pricing model has motivated the extension of the MSM to time series by Duffie and Singleton (1993). We discuss here the application of Heaton (1995). He investigates a representative consumer model in which the consumer is assumed to have time nonseparable preferences. The preferences of the consumer are time additive over a good called services $s(t)$:

$$U(s) = E\left\{\sum_{t=0}^{\infty} \beta^t \frac{s(t)^{1-\gamma} - 1}{1-\gamma}\right\}, \ \gamma > 0$$

where $s(t)$ is a linear function of current and past consumption:

$$s(t) = \sum_{j=0}^{\infty} a(j)\, c(t-j).$$

The marginal utility of consumption at t is given by

$$muc(t) = E\left\{\sum_{\tau=0}^{\infty} \beta^\tau a(\tau)\, s(t+\tau)^{-\gamma} \,|\mathcal{F}(t)\right\} \tag{2.2}$$

The presence of $a(\tau)$ in (2.2) makes it difficult to estimate the model by GMM. Hence, Heaton employs MSM. First, he investigates a simple model for the growth rates of the consumption and dividend series. Denote

$$Y_t = \begin{pmatrix} \ln(c_t/c_{t-1}) \\ \ln(d_t/d_{t-1}) \end{pmatrix}.$$

The moments are based on

$$h(Y_t;\theta) = \begin{pmatrix} Y_t - EY_t \\ Y_{*t}Y'_{*t} - EY_{*t}Y'_{*t} \\ Y_{*t}Y'_{*t-4} - EY_{*t}Y'_{*t-4} \\ Y_{*t}Y'_{*t-8} - EY_{*t}Y'_{*t-8} \\ Y_{*t}Y'_{*t-12} - EY_{*t}Y'_{*t-12} \end{pmatrix}$$

where $Y_{*t} = Y_t - EY_t$. The usual GMM consists in matching the empirical moments (based on monthly observations) with the theoretical ones. As the moments of Y_t are unknown, they will be evaluated by averaging over simulated data. To do so, Heaton needs to specify completely a model for Y_t, he chooses a basic Gaussian AR(12) model for weekly data and takes the average to obtain monthly data.

In a second experiment, he specifies the law of motion for the prices of stocks and bonds and generates dividends, consumptions, accumulated durable goods and habit stock by solving the dynamic problem numerically. The solution of some simple macroeconomic models can be approximated using a loglinearization of the equations, however here a more sophisticated method is needed, such as for instance the projection method advocated by Judd (1998). To estimate the preference parameters, Heaton investigates two sets of moments based on the expectations, variances and covariances of stock and bond returns. He finds evidence in favor of a preference structure where consumption is substitutable in the short run and habits are persistent in the long-run.

2.3. Diffusion process

Financial data are frequently assumed to follow a stochastic differential equation

$$dy_t = \mu\left(y_t;\theta\right)dt + \sigma\left(y_t;\theta\right)dW_t. \tag{2.3}$$

But in practice, data are observed at discrete interval and the likelihood of y_{t+1} conditional on y_t does not have a simple expression. The model can be approximated using a Euler scheme (see e.g. Gouriéroux and Monfort, 1996):

$$y^{(n)}_{\frac{k+1}{n}} = y^{(n)}_{\frac{k}{n}} + \frac{1}{n}\mu\left(y^{(n)}_{\frac{k}{n}};\theta\right) + \frac{1}{\sqrt{n}}\sigma\left(y^{(n)}_{\frac{k}{n}};\theta\right)\varepsilon_k \tag{2.4}$$

where ε_k is iid standard normal. As discussed in Broze, Scaillet, and Zakoian (1998), $y^{(n)}_{k/n}$ is only an approximation of the true process, there might be a discretization bias due to the discretization step $1/n$. However, they show that this discretization bias vanishes when the step is sufficiently small. It will be assumed that we can draw directly from the conditional distribution of y_{t+1} given y_t. So y_{t+1} can be represented as the solution of

$$y_{t+1} = H\left(y_t, \xi_t;\theta\right).$$

2.4. Stochastic volatility model

Some studies suggest that the short-term interest rate dynamics is better modeled by a stochastic volatility model. Andersen and Lund (1997) consider the following continuous-time model

$$\begin{aligned} dr_t &= \kappa_1\left(\mu - r_t\right)dt + \sigma_t r_t^{\gamma} dW_{1t},\ \gamma > 0 && (2.5)\\ d\ln\sigma_t^2 &= \kappa_2\left(\alpha - \ln\sigma_t^2\right)dt + \beta dW_{2t}. && (2.6)\end{aligned}$$

The short term interest rate r_t is observable while the stochastic volatility σ_t is unobservable. The joint process $\{r_t, \sigma_t\}$ is Markovian and can be simulated using a Euler discretization of (2.5) and (2.6). Then one isolates the sequence of $\{r_t\}$ from the joint sequence. References to applications of Indirect Inference and Efficient Method of moments to diffusions and stochastic volatility models will be given at the end of the sections relative to these methods.

3. Method of simulated moments

The method of simulated moments (MSM) has been first introduced by McFadden (1989) and Pakes and Pollard (1989) for the estimation of discrete response models in an iid environment. Later it has been extended to time-series by Lee and Ingram (1991) and Duffie and Singleton (1993).

3.1. General setting

Observations $y_1, ..., y_T$ are realizations obtained from a parametric model indexed by θ. Consider a vector of moment conditions

$$E^{\theta_0} h\left(Y_t, \theta\right) = 0$$

that identifies the parameter θ, that is $E^{\theta_0} h\left(Y_t, \theta\right) = 0 \Rightarrow \theta = \theta_0$. The GMM estimator is the solution of

$$\min_{\theta} h_T\left(\theta\right)' W h_T\left(\theta\right)$$

for some positive definite weighting matrix W and $h_T\left(\theta\right) = \sum_{t=1}^{T} h\left(Y_t; \theta\right)/T$. When the function $h\left(Y_t; \theta\right)$ is difficult to calculate, in particular when it involves a multiple integration, it may be evaluated using simulations. Assume there exists a latent variable ξ_t with a distribution that is easy to simulate and a function $\tilde{h}\left(Y_t, \xi_t, \theta\right)$ so that

$$E_{\xi}\left[\tilde{h}\left(Y_t, \xi_t, \theta\right) | Y_t\right] = h\left(Y_t; \theta\right)$$

where E_{ξ} denotes the expectation with respect to the distribution of ξ_t. In some cases, a gain in efficiency can be obtained by drawing ξ_t conditionally on the observables $\{y_1, ..., y_T\}$. This is the conditional simulation method discussed e.g. by McFadden and Ruud (1994). In the sequel, we will assume that ξ_t is drawn from its marginal distribution. The MSM consists in drawing independently $\xi_t^1, ..., \xi_t^R$ in the distribution of ξ_t and replacing $h\left(Y_t, \theta\right)$ by its simulator

$$\tilde{h}_{t,R}\left(\theta\right) \equiv \frac{1}{R} \sum_{r=1}^{R} \tilde{h}\left(Y_t, \xi_t^r; \theta\right).$$

Hence the MSM is the solution to

$$\hat{\theta}_{MSM} = \arg\min_{\theta} \left(\frac{1}{T} \sum_{r=1}^{T} \tilde{h}_{t,R}\left(\theta\right)\right)' W \left(\frac{1}{T} \sum_{r=1}^{T} \tilde{h}_{t,R}\left(\theta\right)\right) \tag{3.1}$$

It is important to parametrize $\tilde{h}$ in such a way that the distribution of ξ_t does not depend on θ. Hence the same "common random numbers" $\left\{\xi_t^1, ..., \xi_t^R\right\}$, $t = 1, ..., T$ are used for all values of θ in the optimization procedure. This is a necessary condition for the smoothness of the objective function with respect to θ.

Example 1. The Mixed Multinomial Logit model (continued)

Denote $Y = (d, x)$ and

$$\tilde{h}\left(Y, \xi, \theta\right) = \sum_{c \in C} \left\{d_c - L_C\left(c; x, \beta + \Lambda\xi\right)\right\} \omega\left(x, \theta\right).$$

It is immediate to see that $E_{\xi}\left[\tilde{h}\left(Y, \xi; \theta\right) | Y\right] = h\left(Y; \theta\right)$ where $h\left(Y; \theta\right)$ is defined in (2.1).

Example 2. Dynamic model

Assume that Y_t is the solution of

$$\begin{aligned} Y_t &= H\left(Y_{t-1}, \varepsilon_t; \theta\right) \\ Y_0 &= \varepsilon_0 \end{aligned}$$

where ε_t has a known distribution independent of θ. The diffusion process discussed in Subsection 2.3 satisfies such a relation. Solving recursively, Y_t is the solution of

$$\begin{aligned} Y_t &= \tilde{H}\left(Y_0, \varepsilon_t, \varepsilon_{t-1}, ..., \varepsilon_1; \theta\right) \\ &= \tilde{H}\left(\xi_t, \theta\right) \end{aligned}$$

where $\xi_t = \left(\varepsilon_t, \varepsilon_{t-1}, ..., \varepsilon_1, \varepsilon_0\right)$. Assume that θ is identified from a moment condition of the type

$$\begin{aligned} h\left(Y_t, \theta\right) &= a\left(Y_t\right) - E\left(a\left(Y_t\right)\right) \\ &= a\left(Y_t\right) - E_\xi\left[a\left(\tilde{H}\left(\xi_t, \theta\right)\right)\right]. \end{aligned}$$

Then $\tilde{h}$ is chosen so that

$$\begin{aligned} \tilde{h}\left(Y_t, \xi_t, \theta\right) &= a\left(Y_t\right) - a\left(\tilde{H}\left(\xi_t, \theta\right)\right) \\ &\equiv a(Y_t) - \tilde{a}(\xi_t, \theta). \end{aligned}$$

3.2. Simulators

Static case

In static models, like discrete choice models, the goal is to calculate $E_\xi\left[a\left(\xi\right)\right]$ via simulations. Random numbers can be drawn from a distribution G either by inverting the distribution G or as often done by using more sophisticated methods like an acceptance-rejection method (see Devroye, 1986). Once a sequence of iid random numbers $\xi^r, ..., \xi^R$ has been drawn from the distribution of ξ, $E_\xi\left[a\left(\xi\right)\right]$ can be estimated by

$$\hat{E}_\xi\left[a\left(\xi\right)\right] = \frac{1}{R}\sum_{r=1}^{R} a\left(\xi^r\right).$$

When it is difficult to simulate from G or when a is not smooth, other methods should be used like importance sampling. $E_\xi\left[a\left(\xi\right)\right]$ can be rewritten as

$$E_\xi\left[a\left(\xi\right)\right] = \int \frac{a\left(\xi\right) g\left(\xi\right)}{g^*\left(\xi\right)} g^*\left(\xi\right) d\xi$$

where g^* is a pdf with the same support as g. The importance sampling simulator for $E_\xi\left[a\left(\xi\right)\right]$ is

$$\hat{E}_\xi\left[a\left(\xi\right)\right] = \frac{1}{R}\sum_{r=1}^{R} \frac{a\left(\xi^r\right) g\left(\xi^r\right)}{g^*\left(\xi^r\right)}$$

where $\xi^r, ..., \xi^R$ are independent draws from g^*. For a thorough discussion of simulators, see e.g. Geweke (1996), Hajivassiliou and Ruud (1997) and Stern (1997) and on importance sampling see e.g. Geweke (1989).

Dynamic case

In the following, we consider time-series models. There are two simulation strategies, one consists in drawing Y_t conditionally to Y_{t-1}, the other consists in drawing a full sequence of $\{Y_1, Y_2, ..., Y_T\}$.

Conditional simulation

Assume that Y_t admits a representation

$$Y_t = H(Y_{t-1}, X_t, \varepsilon_t, \theta) \tag{3.2}$$

where $\{\varepsilon_t\}$ are iid with known distribution g and X_t is a vector of observable exogenous variables. Then for each t, one can draw random elements $\varepsilon_t^r, r = 1, ..., R$ that are iid of distribution g. Then

$$Y_t^r = H(Y_{t-1}, X_t, \varepsilon_t^r, \theta)$$

has the same conditional distribution as Y_t. Conditional simulation may present some advantage in terms of efficiency of the estimator (see Section 6).

Path simulation

In the path simulation scheme, $(Y_1^r, ..., Y_T^r)$, $r = 1, ..., R$ are drawn from the joint distribution of $(Y_1, ..., Y_T)$ possibly conditionally on the exogenous variables $(X_1, ..., X_T)$. To illustrate this simulation scheme, consider a model

$$U_t = H(U_{t-1}, \varepsilon_t, \theta) \tag{3.3}$$

where again ε_t are iid with pdf g. Assume that U_t admits a stationary distribution. $U_t = (Y_t, \varsigma_t)$ where Y_t is observable and ς_t is an unobservable latent variable. ς_t may be the volatility in a stochastic volatility model or a technology shock in a rational expectation model (see Duffie and Singleton 1993). Replacing successively in (3.3), it can been shown that

$$\begin{aligned} U_t &= \tilde{H}(U_0, \varepsilon_1, ..., \varepsilon_t, \theta) \\ &= \tilde{H}(U_0, \underline{\varepsilon}_t, \theta) \end{aligned}$$

where $\underline{\varepsilon}_t = (\varepsilon_1, ..., \varepsilon_t)$. Using iid draws of ε_t in g, $\varepsilon_t^r, r = 1, ..., R$, one can construct $\underline{\varepsilon}_t^r = (\varepsilon_0^r, \varepsilon_1^r, ..., \varepsilon_t^r)$. Conditionally on U_0, we have iid draws from the distribution of $(U_1, ..., U_T)$:

$$U_t^r = \tilde{H}(U_0, \underline{\varepsilon}_t^r, \theta), \ t = 1, ..., T.$$

For an arbitrary U_0, U_t^r is not necessarily stationary. As discussed in Duffie and Singleton (1993), one needs to insure that the effect of U_0 dies out quickly by assuming that U_t is geometrically ergodic (this imposes conditions on H). Then, one isolates the component Y_t^r of U_t^r.

Note that in the unconditional simulation scheme, one can choose to draw a single series $\{U_s, s = 1, ..., S\}$ distributed as U_t. If S is fixed, the estimator will not be consistent in general. If $S/T \to R$, then the asymptotic properties of this estimator are the same as those of the estimator based on $\{U_t^r, t = 1, ..., T, r = 1, ..., R\}$.

3.3. Asymptotic properties of MSM estimator

Assume that the observations are given by Y_t a vector of stationary and ergodic random variables. They are distributed according to a distribution P^{θ} where θ is the parameter of interest. The vector Y_t may include an exogenous variable Z_t for θ. E^{θ} denotes the expectation with respect to P^{θ}.

Identification condition: $E^{\theta_0} h(Y, \theta) = 0 \Rightarrow \theta = \theta_0$.

We have a function $\tilde{h}$ that depends on Y and some random variable ξ independent of $\{Y_t, t = 1, ..., T\}$ and which distribution is known and does not depend on θ such that

$$E_{\xi}\left[\tilde{h}\left(Y_t, \xi_t, \theta\right) | Y_t\right] = h\left(Y_t, \theta\right). \tag{3.4}$$

Moreover, assume that $\left\{\tilde{h}\left(Y_t, \xi_t, \theta\right)\right\}$ is stationary and ergodic. One can generate $\{\xi_1^r, ..., \xi_T^r\}_{r=1,...,R}$ independently and identically distributed as $\{\xi_1, ..., \xi_T\}$. It follows that $\left\{\tilde{h}\left(Y_t, \xi_t^r, \theta\right)\right\}_{r=1,...,R}$ are also iid conditional on Y_t and satisfy (3.4). The MSM estimator satisfies (3.1) for some positive definite matrix W.

In the static case, $\left\{\tilde{h}\left(Y_t, \xi_t^r, \theta\right)\right\}_{t=1,...,T}$ are uncorrelated conditional on $\{Y_t, t = 1, ..., T\}$, however this is not usually the case in time-series. Below, we consider the general case where $\left\{\tilde{h}\left(Y_t, \xi_t^r, \theta\right)\right\}_{t=1,...,T}$ are autocorrelated conditional on the observed path $\{Y_1, ..., Y_T\}$. Denote $h_t = h(Y_t, \theta_0)$, $\tilde{h}_{r,t} = \tilde{h}\left(Y_t, \xi_t^r, \theta_0\right)$, $\underline{Y} = (Y_1, ..., Y_T)$ and

$$\begin{aligned}
D &= E^{\theta_0}\left[\frac{\partial h_t}{\partial \theta'}\right], \\
\Sigma_0 &= \lim_{T\to\infty} \frac{1}{T} \sum_{k=-T}^{T} E^{\theta_0} cov\left(\tilde{h}_{r,t}, \tilde{h}_{r,t-k} | \underline{X}\right), \\
I_0 &= \lim_{T\to\infty} \frac{1}{T} \sum_{k=-T}^{T} cov^{\theta_0}\left(h_t, h_{t-k}\right)
\end{aligned}$$

The following proposition is a slight generalization of Gouriéroux and Monfort (1991):

Proposition 3.1. *Assume R fixed and that certain regularity assumptions hold. Then the MSM estimator defined in (3.1) is consistent and satisfies*

$$\sqrt{T}\left(\hat{\theta}_{MSM} - \theta_0\right) \underset{T\to\infty}{\overset{d}{\to}} \mathcal{N}\left(0, \left(D'WD\right)^{-1} D'W\left(I_0 + \frac{1}{R}\Sigma_0\right) WD\left(D'WD\right)^{-1}\right).$$

The optimal weighting matrix is given by $W = \left(I_0 + \frac{1}{R}\Sigma_0\right)^{-1}$. In that case the asymptotic variance simplifies to

$$\left(D'\left(I_0 + \frac{1}{R}\Sigma_0\right)^{-1} D\right)^{-1}.$$

Special cases:

- If $\left\{\tilde{h}\left(Y_t, \xi_t^r, \theta_0\right)\right\}_{t=1,...,T}$ are uncorrelated conditional on $\{Y_1, Y_2, ..., Y_T\}$ then Σ_0 simplifies to
$$\Sigma_0 = E^{\theta_0} Var\left[\tilde{h}(Y_t, \xi_t^r, \theta_0)|\underline{Y}\right].$$

- **Model MSM1.** Now consider a static setting that is typical of discrete response models. Assume that the observations $y = (x, z)$ satisfy
$$x = H\left(z, \xi, \theta\right)$$
and the moment conditions are based on
$$\begin{aligned} h\left(y, \theta\right) &= \omega\left(z\right)\left(x - E^{\theta}\left[x|z\right]\right), \\ \tilde{h}\left(y, \xi, \theta\right) &= \omega\left(z\right)\left(x - H\left(z, \xi, \theta\right)\right). \end{aligned}$$
Pakes and Pollard (1989) and McFadden (1989) show that, for the optimal weighting matrix $W = I_0^{-1}$, the asymptotic variance simplifies to
$$\left(1 + \frac{1}{R}\right)\left(D' I_0^{-1} D\right)^{-1}.$$
Estimation of multinomial probit models is particularly challenging because the criterion to minimize is discontinuous in θ. Therefore, Pakes, Pollard, and McFadden established the asymptotic properties of the MSM estimator using the theory of empirical processes.

- **Model MSM2.** Consider the special case of Example 2 where
$$h(Y_t; \theta) = a(Y_t) - E^{\theta}\left[a(Y_t)\right]$$
$$\tilde{h}\left(Y_t, \xi_t^r; \theta\right) = a(Y_t) - \tilde{a}(\xi_t^r, \theta)$$
In this case $\Omega = \Sigma$ so that for the optimal weighting matrix $W = \Sigma^{-1}$, Duffie and Singleton (1992) show that
$$\sqrt{T}\left(\hat{\theta}_{MSM} - \theta_0\right) \to \mathcal{N}\left(0, \left(1 + \frac{1}{R}\right)\left(D' I_0^{-1} D\right)^{-1}\right).$$

- $\left(D' I_0^{-1} D\right)^{-1}$ corresponds to the asymptotic variance of the GMM estimator. Therefore, in the two previous cases, the asymptotic efficiency of the MSM estimator with respect to the GMM estimator of $R/(1+R)$. For $R = 1$, this relative efficiency is 50% while for $R = 9$, it is 90%. When R goes to infinity, MSM is as efficient as GMM. However, just as GMM, it is usually inefficient relatively to the MLE due to the arbitrary choice of moment conditions.

A list of applications of MSM can be found in the book by Gouriéroux and Monfort (1996) and in the surveys by Gouriéroux and Monfort (1993), Liesenfeld and Breitung (1999), and Stern (1997). Recent applications of MSM to microeconomics include Laffont, Ossard and Vuong (1995) (to auctions), McFadden and Train (1995) (to multinomial logit models of transportation), and Sieg (2000) (estimation of a bargaining game with asymmetric information). MSM has found many applications in macroeconomics including the estimation of a representative agent asset pricing model by Heaton (1995), a multinomial probit model of household portfolio choice by Asea and Turnowsky (1998), real business cycle models by Collard, Fève and Perraudin (2000), Collard, Fève, Langot, and Perraudin (2001), Hairault, Langot, and Portier (1997), and Jonsson and Klein (1996), a exchange rate target zone model by Iannizzoto and Taylor (1999) and Galindo (2000). Finally, it has been recently applied to the estimation of option pricing models with jumps by Amaro de Matos (2001).

4. Methods using an auxiliary model

MSM consists in matching empirical moments and theoretical moments that are estimated by simulations. The Indirect Inference and Efficient Method of Moments also rely on GMM by using the parameter estimates or the scores of an auxiliary model to define a set of moment conditions.

4.1. Indirect Inference

Gouriéroux, Monfort and Renault (1993) and Smith (1993) introduce an estimation method based on the inversion of a binding function that they named Indirect Inference. For a detailed review of this method, see Gouriéroux and Monfort (1996). Consider observations $x_1, ..., x_T$ generated by P^θ. Assume that $x_t = (y_t, z_t)$ where z_t is exogenous for the parameter of interest θ in the true model that is the distribution of z_t does not depend on θ. Assume that an estimator $\hat{\lambda}$ is obtained by maximization of a criterion

$$\hat{\lambda} \equiv \arg\max_{\theta} Q_T \left(\underline{y}_T, \underline{z}_T; \lambda \right) \tag{4.1}$$

where $\underline{y}_T = (y_1, ..., y_T)$, $\underline{z}_T = (z_1, ..., z_T)$. The criterion is usually a quasi-likelihood, although Q_T does not have to be a density and all results stated here hold as long as $\partial Q_T / \partial \lambda$ satisfies a central limit theorem, that is $\sqrt{T} \partial Q_T / \partial \lambda \xrightarrow{d} \mathcal{N}(0, I_0)$. Under the usual assumptions, it follows that

$$\sqrt{T} \left(\hat{\lambda} - \tilde{\lambda}(\theta_0) \right) \xrightarrow{d} \mathcal{N}(0, \Omega(\theta_0))$$

where $\Omega(\theta_0) = J_0^{-1} I_0 J_0^{-1}$ with

$$I_0 = \lim_{T \to \infty} V^{\theta_0} \left[\sqrt{T} \frac{\partial}{\partial \lambda} Q_T \left(\underline{y}_T, \underline{z}_T; \lambda_0 \right) \right],$$

$$J_0 = \lim_{T\to\infty} \frac{\partial^2}{\partial\lambda\partial\lambda'} Q_T\left(\underline{y}_T, \underline{z}_T; \lambda_0\right).$$

If $\tilde{\lambda}(\theta)$ were known and $\theta \to \tilde{\lambda}(\theta)$ is a one-to-one mapping, θ could be estimated consistently from

$$\hat{\theta} = \arg\min_{\theta} \left(\hat{\lambda} - \tilde{\lambda}(\theta)\right)' W \left(\hat{\lambda} - \tilde{\lambda}(\theta)\right).$$

where W is a symmetric positive definite matrix.

Example. Assume that the DGP is a moving average

$$\begin{cases} y_t = \varepsilon_t - \theta\varepsilon_{t-1} \\ \varepsilon_t \sim iid(0,1) \end{cases}$$

with $\theta \in [-1, 1]$. Assume one selects as auxiliary model an AR(1) so that

$$Q_T\left(\underline{y}_T; \lambda\right) = \frac{1}{T}\sum_{t=1}^{T} (y_t - \lambda y_{t-1})^2.$$

In this case, the binding function $\tilde{\lambda}(\theta)$ has a closed form expression $\tilde{\lambda}(\theta) = -\theta/\left(1+\theta^2\right)$ and is invertible, see Ghysels, Khalaf, and Vodounou (2001).

However, in general, $\tilde{\lambda}(\theta)$ is unknown and needs to be estimated. Assume that it is relatively easy to simulate in P^{θ}. For each θ, one can generate R samples of size T, $\underline{y}_T^r(\theta) = (y_1^r(\theta), ..., y_T^r(\theta))$, with the same distribution as $\{y_1, ..., y_T\}$ conditionally on the observed path of the exogenous variable $\underline{z}_T$ to obtain an estimator $\tilde{\lambda}_T^r(\theta)$, $r = 1, ..., R$. Denote $\tilde{\lambda}_{TR}(\theta) = \frac{1}{R}\sum_{r=1}^{R} \tilde{\lambda}_T^r(\theta)$. Alternatively, one can generate a single series of size TR, $\underline{y}_{TR}(\theta)$, conditional on $(\underline{z}_T, \underline{z}_T, ..., \underline{z}_T)$ to obtain an estimator $\tilde{\lambda}_{TR}(\theta)$. The result is the same in either case although the second approach saves computations. Then the estimator is the solution of

$$\widehat{\theta}_{IND} = \arg\min_{\theta} \left(\tilde{\lambda}_{TR}(\theta) - \hat{\lambda}\right)' W \left(\tilde{\lambda}_{TR}(\theta) - \hat{\lambda}\right). \tag{4.2}$$

Denote $\Lambda(\theta) = \partial\tilde{\lambda}(\theta)/\partial\theta'$, $\Lambda_0 = \Lambda(\theta_0)$, $\Omega_0 = \Omega(\theta_0)$, $\lambda_0 = \tilde{\lambda}(\theta_0)$ and

$$K_0 = \lim_{T\to\infty} VE^{\theta_0}\left[\sqrt{T}\frac{\partial}{\partial\lambda} Q_T\left(\underline{y}_T, \underline{z}_T; \lambda_0\right) | \underline{z}_T\right].$$

Gouriéroux, Monfort, and Renault (1993) prove the following result:

Proposition 4.1. *For the optimal weighting matrix* $W_0 = J_0(I_0 - K_0)^{-1} J_0$, *we have*

$$\sqrt{T}\left(\hat{\theta}_{IND} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \left(1 + \frac{1}{R}\right)\left[\Lambda_0' J_0 (I_0 - K_0)^{-1} J_0 \Lambda_0\right]^{-1}\right).$$

Because of the presence of common exogenous variables in $\tilde{\lambda}_{TR}(\theta)$ and $\hat{\lambda}$, these two estimators are not independent from each other. As a result, the variance of $\widehat{\theta}_{IND}$ contains the extra term $-K_0$. If both z_t and y_t were simulated, the term K_0 would vanish and the asymptotic variance would be equal to

$$\left(1+\frac{1}{R}\right)\left[\Lambda_0' J_0 I_0^{-1} J_0 \Lambda_0\right]^{-1}.$$

Denote

$$D = \lim_{T\to\infty} \frac{\partial^2}{\partial\theta' \partial\lambda} Q_T\left(\underline{y}_T, \underline{z}_T; \lambda_0\right).$$

The asymptotic variance of $\sqrt{T}\left(\hat{\theta}_{IND}-\theta_0\right)$ can be rewritten as

$$\left(1+\frac{1}{R}\right)\left[D'\left(I_0-K_0\right)^{-1} D\right]^{-1}.$$

We get a result analogous to MSM, where $\partial Q_T/\partial\lambda$ plays the role of h_T.

We consider now a slight generalization of the setting proposed by Gouriéroux et al. (1993). The estimator $\hat{\lambda}$ does not have to be a QMLE. It may also be obtained from the optimization of a GMM criterion or M-criterion. Assume that

$$\sqrt{T}\left(\hat{\lambda}-\tilde{\lambda}\left(\theta_0\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega\left(\theta_0\right)\right) \tag{4.3}$$

and that the dimension, q, of $\tilde{\lambda}(\theta)$ is not smaller than that, p, of θ. Moreover assume, for identification, that the mapping $\theta \to \tilde{\lambda}(\theta)$ is one to one.

To simplify, we consider an unconditional simulation. For given θ, one generates either a single series $x_1, ..., x_S$ drawn in the distribution P^θor R iid series of size T, $\{x_1^r, ..., x_T^r\}, r = 1, ..., R$, both independently of the sample of observations $\{x_1, ..., x_T\} \equiv \underline{x}_T$. Assuming $S/T \to R$, both approaches give the same result. We adopt the second for sake of uniformity in our exposition. From the rth simulated series, one obtains an estimator $\tilde{\lambda}^r(\theta)$ independent of $\hat{\lambda}$ that satisfies

$$\sqrt{T}\left(\tilde{\lambda}_T^r(\theta)-\tilde{\lambda}(\theta)\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega(\theta)\right).$$

Assume that

$$E\tilde{\lambda}_T^r(\theta) = \tilde{\lambda}(\theta) \text{ for all } \theta,\ r = 1, ..., R.$$

Let

$$\tilde{\lambda}_{TR}(\theta) = \frac{1}{R}\sum_{r=1}^{R} \tilde{\lambda}_T^r(\theta).$$

Then the Simulated Indirect Inference (IND) estimator is given by (4.2). A simple Taylor expansion of the first order condition of (4.2) around λ_0 gives the following result.

Proposition 4.2. *Under certain regularity assumptions, we have consistency and*

$$\sqrt{T}\left(\hat{\theta}_{IND}-\theta_0\right)\to N\left(0,\left(1+\frac{1}{R}\right)\left[\Lambda_0' W\Lambda_0\right]^{-1}\Lambda_0' W\Omega_0 W\Lambda_0\left[\Lambda_0' W\Lambda_0\right]^{-1}\right).$$

For the optimal weighting matrix $W_0=\Omega_0^{-1}$, *we have*

$$\sqrt{T}\left(\hat{\theta}_{IND}-\theta_0\right)\xrightarrow{d}\mathcal{N}\left(0,\left(1+\frac{1}{R}\right)\left[\Lambda_0'\Omega_0^{-1}\Lambda_0\right]^{-1}\right).$$

Example. Estimation of a diffusion (continued)

Broze, Scaillet, and Zakoian (1998) estimate a diffusion process (2.3) by IND. They simulate the data using a Euler approximation as described in (2.4) with a discretization step h_1 (denoted previously $1/n$). They use as auxiliary model another Euler approximation using a coarser grid $h_2 \gg h_1$. They show that if the discretization step h_1 is fixed, the resulting IND estimator is biased because the data are not drawn from the true Data Generating Process (DGP). However if h_1 goes to zero sufficiently fast with the sample size T, the estimator is unbiased. They perform a Monte Carlo experiment on an Ornstein Uhlenbeck process with a small sample size $T=50$ and they find that an optimal choice of h_1 is $1/100$, while an optimal choice of R is 5.

In what precedes, it is assumed that the DGP is completely specified and it is possible to draw simulations from the DGP. Dridi and Renault (1999) extend the IND to a semi-parametric context and give conditions for the consistency of the estimators when the model from which data are drawn is (partially) misspecified. Indirect Inference has been applied to the estimation of diffusion processes by Broze, Scaillet, and Zakoian (1998), to the estimation of stochastic volatility models by Monfardini (1998), Calzolari, De Iori, and Firorentini (1998), and Pastorello, Renault, and Touzi (2000), to models of unemployment by Magnac, Robin, and Visser (1995), An and Liu (2000), and Topa (2001), and to rational expectation dynamic models by Collard, Fève, and Perraudin (2000).

4.2. Efficient Method of Moments

Next, we present the method proposed by Gallant and Tauchen (1996). For a survey, see Tauchen (1997) and Gallant and Tauchen (2001, 2002). The setting is close to that of the previous subsection. The observations are again given by (y_t, z_t), $t=1,...,T$ where z_t is exogenous. Again $\hat{\lambda}$ is obtained from (4.1). The idea behind the Efficient Method of Moments (EMM) is that, for the true θ, the expectation of the pseudo score should be zero, this suggests an estimator based on

$$\min_{\theta}\left\|E^{\theta}\left[\frac{\partial Q_T}{\partial\lambda}\left(\underline{y}_T(\theta),\underline{z}_T;\hat{\lambda}\right)\right]\right\|.$$

The difference between EMM and IND is analogous to that between Lagrange Multiplier and Wald tests in the sense that EMM is based on the score while IND is based on the estimator. In practice, the expectation is replaced by an average over simulated data $\left\{\underline{y}_T^r(\theta),\ r=1,...,R\right\}$ conditionally on the observed exogenous variables $\underline{z}_T$. The EMM estimator is given by

$$\widehat{\theta}_{EMM} = \arg\min_{\theta} \frac{\partial Q_{TR}\left(\theta,\hat{\lambda}\right)}{\partial \lambda'} W \frac{\partial Q_{TR}\left(\theta,\hat{\lambda}\right)}{\partial \lambda}$$

where W is symmetric and positive definite matrix and $Q_{TR}\left(\theta,\hat{\lambda}\right) \equiv \frac{1}{R}\sum_{r=1}^{R} Q_T\left(\underline{y}_T^r(\theta),\underline{z}_T;\hat{\lambda}\right)$. Gouriéroux, Monfort, and Renault (1993, Section 2.5) show that IND and EMM are asymptotically equivalent so that the following result holds.

Proposition 4.3. *Under certain regularity assumptions and for the optimal weighting matrix, $W_0 = (I_0 - K_0)^{-1}$, we obtain*

$$\sqrt{T}\left(\hat{\theta}_{EMM} - \theta_0\right) \overset{d}{\to} \mathcal{N}\left(0, \left(1+\frac{1}{R}\right)\left[D'\left(I_0 - K_0\right)^{-1} D\right]^{-1}\right)$$

where I_0, K_0, and D are defined in Subsection 4.1.

In absence of exogenous variables (which is often the case in finance), one may generate a single series $y^1,...,y^S$ with $S = TR$ if the process is stationary (Case 2 of Gallant and Tauchen) or R series $y^1,...,y^T$ if the process is not stationary (Case 3 of Gallant and Tauchen). For $W_0 = I_0^{-1}$, the EMM estimator satisfies

$$\sqrt{T}\left(\hat{\theta}_{EMM} - \theta_0\right) \overset{d}{\to} \mathcal{N}\left(0, \left(1+\frac{1}{R}\right)\left[D'I_0^{-1}D\right]^{-1}\right).$$

We now discuss the differences between EMM and IND. Contrary to the EMM, the IND has the advantage not to require the specification of a score since the first-step estimator $\hat{\lambda}$ may be obtained from any criterion. On the other hand, IND requires the computation of one or several estimators based on the simulated sample(s), and this for each value of θ till convergence obtains, which often is burdensome. The choice of the auxiliary model is driven by efficiency considerations. Gouriéroux et al. suggest to use a model close to the true one, for instance use a Euler discretization for a diffusion model. The idea is that if the auxiliary model is sufficiently close to the true one, then the loss of efficiency should be small. Gallant and Tauchen propose to use a quasi likelihood for which the dimension (m) of λ increases with the sample size. They use a flexible form so that the pseudo score will encompass the true score for m sufficiently large, resulting in an estimator that is asymptotically efficient. They advocate the use of a SemiNonParametric (SNP) specification of the auxiliary model. It is based on a truncated Hermite series expansion. In the case without exogenous variables, the quasi-likelihood is

$$Q_T = \frac{1}{T}\sum_{t=1}^{T} \ln f_M\left(y_t|y_{t-1},...,y_{t-L};\lambda\right) \qquad (4.4)$$

with

$$f_M\left(y_t|y_{t-1},...,y_{t-L};\lambda\right)=\frac{\left[P_M\left(u_t\right)\right]^2\phi\left(u_t\right)}{\int\left[P_M\left(u\right)\right]^2\phi\left(u\right)du}\frac{1}{R_t} \tag{4.5}$$

where ϕ is the standard normal distribution and

$$\begin{aligned}
u_t &= \frac{y_t-\mu_t}{R_t}, \\
\mu_t &= \lambda_{10}+\sum_{l=1}^{L_\mu}\lambda_{1l}y_{t-l}, \\
R_t &= \lambda_{20}+\sum_{l=1}^{L_r}\lambda_{2l}\left|y_{t-l}\right|, \\
P_M\left(u\right) &= \sum_{l\leq M_u}a_{l,t}u^l, \\
a_{l,t} &= \sum_{|\eta|\leq M_y}\lambda_{3l\eta}\prod_{j=1}^{L_p}y_{t-j}^{\eta_j}
\end{aligned}$$

with $\eta=\eta_1+...+\eta_{L_p}$, $|\eta|=\sum_{j=1}^{L_p}\left|\eta_j\right|$, $M=\left(M_u,M_y\right)$, $L=\max\left(L_\mu,L_r,L_p\right)$. $\lambda=\left(\lambda_{1l},l=1,...,L_\mu,\lambda_{2l},l=1,...,L_r,\lambda_{3l\eta},l=1,...,M_y\right)$. To achieve identification, one imposes $\lambda_{300}=1$. The orders $\left(L_\mu,L_r,L_p\right)$ and M are chosen using the BIC criterion. The SNP specification permits to approach efficiency when M and L increase.

We discuss the efficiency of EMM based on a SNP auxiliary model. Most of the arguments are taken from Gallant and Tauchen (2001, 2002). Gallant and Long (1997) give a necessary and sufficient condition for the EMM estimator to be efficient. We reformulate it for arbitrary moment conditions. Assume that the DGP is characterized by the pdf

$$p\left(y_t|x_t;\theta_0\right)$$

where $x_t=\left(y_{t-1},...,y_{t-d}\right)$, d is possibly infinite. Consider a set of moment conditions $h_m=\left\{h\left(1\right),h\left(2\right),...,h\left(m\right)\right\}$. Assume that the $\{h(i)\}$ are uncorrelated so that the asymptotic variance of the GMM estimator $\sqrt{T}\left(\hat{\theta}_{GMM}-\theta_0\right)$ takes the form

$$\Sigma_m=\left\{E^{\theta_0}\left[\frac{\partial h_m}{\partial\theta'}\right]E^{\theta_0}\left(h_mh_m'\right)^{-1}E^{\theta_0}\left[\frac{\partial h_m}{\partial\theta}\right]\right\}^{-1}. \tag{4.6}$$

Let H denote the space formed by the linear combinations $\sum_{i=1}^{\infty}\omega_ih\left(i\right)$ for arbitrary constants ω_i and $\overline{H}$ be its closure in terms of a L^2 norm or Sobolev norm. The following proposition is a restatement of efficiency results that can be found in Gallant and Long (1997) and Tauchen (1997). See Hansen (1985) for related results and Carrasco and Florens (2002) for a proof in the context of a continuum of moment conditions in an iid environment.

Proposition 4.4. *Σ_m defined in (4.6) converges to the Cramer Rao efficiency bound as m goes to infinity, if and only if the scores of p are in $\overline{H}$.*

As the moment conditions formed by the pseudo-score based on the SNP auxiliary model have the property to span the true scores (in terms of convergence in a Sobolev norm) when T and R go to infinity, it follows that EMM is asymptotically efficient. However, Gallant and Tauchen (1996) do not give the rate of convergence of M and L for which the estimator is asymptotically efficient. They warn against using too large M and L that would ruin the $\sqrt{T}-$speed of convergence of the EMM estimator. Note that as EMM and IND are asymptotically equivalent, IND based on an SNP auxiliary model will share the same asymptotic efficiency property as EMM.

The SNP is not the only way to achieve efficiency. Gallant and Tauchen (1999) discuss the relative efficiency of MSM and EMM in an iid setting. Consider MSM based on powers of y :

$$h_{MSM} = \begin{pmatrix} y - Ey \\ y^2 - Ey^2 \\ \vdots \\ y^m - Ey^m \end{pmatrix}.$$

Let us denote h_{EMM} the pseudo score derived from (4.4) and (4.5) where m denotes the dimension of λ.

$$h_{EMM} = \frac{\partial \ln f_M}{\partial \lambda} \left(y_t | y_{t-1}, ..., y_{t-L}; \lambda \left(\theta_0 \right) \right).$$

Denote $h_{MSM}(i) = y^i - Ey^i$ and $h_{EMM}(i) = \partial \ln f_M / \partial \lambda_i$ where $i = 1, 2, ...$ Gallant and Tauchen (1999) give another version of Proposition 4.4, that is, if the set of moments $\{h(i),\ i = 1, ..., \infty\}$ spans the space $L^2(p) = \left\{ g | \int g(y)^2 p(y; \theta_0)\, dy < \infty \right\}$, the asymptotic variance of the estimator based on $\{h(1), h(2), ..., h(m)\}$ converges to the Cramer Rao efficiency bound as m goes to infinity. This condition implies the necessary and sufficient condition of Proposition 4.4 because if .$\{h(i),\ i = 1, ..., \infty\}$ spans the whole space $L^2(p)$, the score necessarily belong to $\overline{H}$. Sometimes, a finite number of moments is sufficient to encompass the score, for example the score of the normal density is in the span of $\{h_{MSM}(i),\ i = 1, 2\}$ and $\{h_{EMM}(i),\ i = 1, 2\}$. However, in general, an infinity of moments is needed. The polynomials span $L^2(p)$ if $p(.; \theta)$ has a moment generating function (Gallant, 1980). Therefore both MSM and EMM are asymptotically efficient (as T and m go to infinity). Therefore these estimators based on m moment conditions should have a good efficiency relative to MLE if m is reasonably large. In some special cases, a measure of the relative efficiency (for fixed m) can be derived analytically. Gallant and Tauchen (1999) compute this measure when the DGP is a mixture of normal distributions and find that EMM is more efficient than MSM. This is not surprising as EMM uses an expansion on Hermite polynomials which is well-known to fit the normal distribution particularly well. An argument in favor of EMM over MSM is that the number of terms in the SNP can be selected using a BIC criterion while there is no natural way of choosing m in MSM. Finally, note that power functions are not the only moments that have the property to encompass the score. Moments based on the characteristic function (c.f.) have this property, see Carrasco and Florens (2002) and Carrasco, Chernov, Florens, and Ghysels (2001). Applications of the c.f. will be discussed in Section 6.

EMM has been applied to the estimation of stochastic volatility (SV) models by Andersen, Chung, and Sorensen (1999), Chernov, Ghysels (2000), Andersen and Lund (1997), Gallant, Hsieh, and Tauchen (1997), and Solibakke (2001), to SV models with jumps by Andersen, Benzoni and Lund (2001), to a multifactor SV model by Gallant, Hsu, and Tauchen (1999), to a multifactor SV model with jumps by Chernov, Gallant, Ghysels, and Tauchen (2001), to interest rate diffusions by Dai and Singleton (2001) and Gallant and Tauchen (1998), to regime switching models with long memory by Chow and Liu (1999) and Liu (2000), to a target-zone model for exchange rate by Chung (2001) and Chung and Tauchen (2001), and to labor by Nagypal (2000) (although she does not use SNP).

5. Specification tests and small sample properties

5.1. Specification tests and diagnostics

The three estimation methods we consider consists in minimizing a quadratic form. We can apply the J-test proposed by Hansen (1982) to test whether the overidentifying restrictions are equal to zero. Denote $J_{R,T}$ the criterion. For MSM, assume that the model is of the type MSM1 or MSM2, we have

$$J_{R,T} = T\frac{R}{R+1}\left(\frac{1}{T}\sum_{r=1}^{T}\tilde{h}_{t,R}\left(\hat{\theta}_{MSM}\right)\right)'\hat{I}^{-1}\left(\frac{1}{T}\sum_{r=1}^{T}\tilde{h}_{t,R}\left(\hat{\theta}_{MSM}\right)\right).$$

For IND,

$$J_{R,T} = T\frac{R}{R+1}\left(\tilde{\lambda}_{TR}\left(\hat{\theta}_{IND}\right)-\hat{\lambda}\right)'\hat{J}\left(\hat{I}-\hat{K}\right)^{-1}\hat{J}\left(\tilde{\lambda}_{TR}\left(\hat{\theta}_{IND}\right)-\hat{\lambda}\right).$$

For EMM,

$$J_{R,T} = T\frac{R}{R+1}\left\{\frac{\partial Q_{TR}}{\partial\lambda'}\left(\hat{\theta}_{EMM},\hat{\lambda}\right)\left(\hat{I}-\hat{K}\right)^{-1}\frac{\partial Q_{TR}}{\partial\lambda}\left(\hat{\theta}_{EMM},\hat{\lambda}\right)\right\}$$

where $\hat{I}$, $\hat{J}$, and $\hat{K}$ denote consistent estimators of I_0, J_0, and K_0. If the model is correctly specified, these statistics $J_{R,T}$ converge to a $\chi^2(m-q)$ where m is the number of moments in the MSM and the dimension of λ in IND and EMM, and q is the dimension of θ. The coefficient $R/(1+R)$ converges to one when R goes to infinity. The asymptotic distribution of $J_{R,T}$ is proved by Gouriéroux et al. (1993) in the context of IND. The same result holds for EMM as the IND and EMM criterions based on the same auxiliary model are equivalent (Gallant and Tauchen, 1996). When the model from which the data are simulated is not correctly specified, the test $J_{R,T}$ for IND becomes an encompassing test (Dhaene, Gouriéroux and Scaillet, 1998). It tests whether the model P^{θ} encompasses Q, in other words, whether P^{θ} can explain the behavior of Q.

If this J-test rejects then one should perform diagnostic tests based on the t-statistic of the individual elements of the vector of moments. These tests will signal which of these elements fails to explain the model. Consider the EMM estimator without exogenous variables. Gallant, Hsieh, and Tauchen (1997) propose using

$$T_T = \left\{ diag \left[I_0 - D_{\theta} \left(D_{\theta}' \hat{I}^{-1} D_{\theta} \right)^{-1} D_{\theta}' \right] \right\} \sqrt{T} \frac{\partial}{\partial \lambda} Q_{TR} \left(\hat{\theta}; \hat{\lambda} \right)$$

where $D_{\theta} = \frac{\partial^2}{\partial \lambda \partial \theta'} Q_{TR} \left(\hat{\theta}_{EMM}, \hat{\lambda} \right)$. They show that T_T is asymptotically $\mathcal{N}(0, I)$. A similar test could be applied in the context of MSM and IND.

We consider now a test proposed by Eichenbaum, Hansen, and Singleton (1988) to test whether a subset of orthogonality conditions hold. Assume that $\theta = (\theta_1', \theta_2')'$, $\theta_1 \in R^{q_1}$, $\theta_2 \in R^{q_2}$ and the moment conditions can be partionned as $h(\theta) = \left(h_1(\theta_1)', h_2(\theta)' \right)'$, $h_1 \in R^{m_1}$ and $h_2 \in R^{m_2}$. Suppose that $E^{\theta_0}(h_1(\theta_{10})) = 0$ and that a researcher wants to test $H_0 : E^{\theta_0}(h_2(\theta_0)) = 0$. The test is based on the difference between the $J-$statistic associated with h and that associated with h_1 :

$$C_T = T h_T \left(\hat{\theta} \right)' \hat{I}^{-1} h_T \left(\hat{\theta} \right) - T h_{1T} \left(\hat{\theta}_1 \right)' \left(\hat{I}_{11} \right)^{-1} h_T \left(\hat{\theta}_1 \right)$$

where $\hat{I}$ is an estimator of the long-run variance-covariance matrix of h and $\hat{I}^{11}$ is the upper-left $m_1 \times m_1-$block of $\hat{I}$. Eichenbaum et al. show that, under the null hypothesis, C_T converges to a chi-square with $m_2 - q_2$ degrees of freedom. Heaton (1995) applies this test to compare different models, for instance a pure habit persistence model against a complete model including also local substitution.

5.2. Small sample properties

In this subsection, we discuss the results of Monte Carlo studies. Michaelides and Ng (2000) compare the performance of Method of Simulated Moments, IND and EMM in small samples $T = 100$ or 200. They use the same auxiliary model for IND and EMM. MSM is based on the expectation, variance and covariance of y_t. They investigate various numbers of simulations $R = 5, 10, 25$ and find $R = 10$ adequate. The small sample performance of IND is best for the estimation of an MA(1), while EMM beats IND and MSM for the estimation of a rational expectations model of speculative storage. In terms of computational time, SMS is the fastest and IND is the slowest. Ghysels, Khalaf, and Vodounou (2001) investigate the performance of IND, EMM, and other methods for estimating an AR(1) using an AR(p) auxiliary model. They find that IND performs particularly well when the MA(1) coefficient is close to one. Andersen, Chung and Sorensen (1999) investigate the robustness of EMM to various auxiliary models. The DGP they consider is a discrete-time stochastic volatility model. They show that EMM is sensitive to the choice of the auxiliary model in small samples but performs well in large samples. Interestingly, they do not recommend introducing terms in the SNP in small samples ($T \leq 1500$), while a small number of terms ($M_u = 2$, $M_y = 0$) is beneficial in large samples ($T = 4000$). Zhou (2002) perform a Monte Carlo study on MLE, Quasi-MLE (QMLE), GMM, and EMM for a continuous-time square-root process. He finds the following ranking by decreasing efficiency: MLE, QMLE, EMM, and GMM. All these studies focused on specific models and their conclusions may not carry to another model. However, one can say that the GMM (and therefore MSM) estimator based on ad hoc moment conditions generally is not very efficient but is easy to implement. On the other hand, EMM and IND are more computationally intensive but deliver nearly efficient estimators if they use an auxiliary model that approximates well the DGP.

6. New developments and conclusion

In this subsection, we discuss recent papers that propose new simulated methods that are either easier to apply or more efficient than the existing methods.

Singleton (2001) investigates using the conditional characteristic function (c.c.f.) to estimate diffusion processes. Let $\{y_t\}$ be a scalar diffusion process and $\psi_\theta(\tau|y_t)$ be the c.c.f. of y_{t+1} conditional on y_t. Singleton shows that GMM based on moments

$$\left(e^{i\tau y_{t+1}} - \psi_\theta(\tau|y_t)\right)\omega(y_t)$$

is asymptotically efficient for an appropriate choice of the instrument ω. The efficiency results from the fact that the true score lies in the space spanned by these moment conditions. When y_t is solution of an affine diffusion, the c.c.f. is known. Otherwise, $\psi_\theta(\tau|y_t)$ can be estimated via simulations. This requires simulating from the conditional distribution of y_{t+1} conditional on y_t as described in Subsection 3.2. This line of research is pursued by Carrasco, Chernov, Florens and Ghysels (2001). They apply the generalization of GMM proposed by Carrasco and Florens (2000) to exploit the full continuum of moments resulting from the c.c.f.. This method requires the introduction of a regularization parameter to be able to invert the covariance matrix.

One method that we have omitted in this survey (because it is not based on moment conditions) is the simulated maximum likelihood (SML) estimation. SML has been widely applied to the estimation of discrete choice models, see e.g. the survey by Kamionka (1998). Fermanian and Salanié (2001) propose a nonparametric SML estimator based on a kernel estimation of the likelihood. They show that this method delivers an efficient estimator in the static case.

Finally, the next generation of simulated methods combines the techniques from classical and Bayesian statistical inference, they use importance sampling, Gibbs sampling, Metropolis-Hastings algorithm and Markov Chain Monte Carlo (MCMC) methods, see the special issue of *Econometrics Journal* (1998) on simulation methods. Hajivassiliou and McFadden (1998) use Gibbs resampling techniques to perform a classical estimation of a limited dependent variable model. This method yields a criterion that is smooth in the parameter of interest and therefore easier to optimize. Schennach (2001) combines Exponential Tilting and Empirical Likelihood to estimate latent variable models. Contrary to MSM, her simulation method using Metropolis-Hastings algorithm does not require the parametric or semi-parametric specification of the DGP of the latent variables. MCMC has been applied e.g. to the estimation of Markov-switching models by Billio, Monfort, and Robert (1999) and of diffusions by Elerian, Chib, and Shephard (2001).

We now review two points that need to be kept in mind when applying simulated methods. First, the resulting estimators are consistent only if the simulations have been drawn from the true DGP. This requires the specification of the full model including the dynamic of the latent variable. Some attempts have been made to relax this assumption: Dridi and Renault (1999) in a dynamic context and Schennach (2001) in a static model. Second, except in very special cases, MSM based on ad hoc moment conditions is not efficient. As discussed in Section 4.2, efficiency can be reached if the score is in the linear span of the moment conditions. In general, this will be possible only if an infinity of moment conditions is used. Three important examples of efficient estimators are the following. First, the EMM based on SNP where the dimension of λ increases. Second, MSM based on polynomial functions of increasing order. Third, MSM estimation using the characteristic function. All these methods will deliver an asymptotically efficient estimator if the number of moment conditions is allowed to increase at a certain rate. Hence they require the use of a smoothing parameter that take the form of a truncation parameter (to control the number of moments) or a regularization parameter (as in Carrasco et al., 2001).

ACKNOWLEDGMENTS

We are grateful to the two editors, Eric Ghysels and Alastair Hall for helpful comments. We also wish to thank Nour Meddahi.

References

[1] An, M. Y. and M. Liu (2000) "Using Indirect Inference to Solve the Initial-Conditions Problem", *Review of Economics and Statistics*, 82, 656-667.

[2] Andersen, T., L. Benzoni, and J. Lund (2001) "An Empirical Investigation of Continuous-Time Equity Return Models", forthcoming in *Journal of Finance*.

[3] Andersen, T., H-J. Chung, and B. Sorensen (1999) "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study", *Journal of Econometrics*, 91, 61-87.

[4] Andersen, T. and J. Lund (1997) "Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate", *Journal of Econometrics*, 77, 343-378.

[5] Amaro de Matos, J. (2001) "MSM Estimators of European Options with Jumps", *Mathematical Finance*, 11, 189-203.

[6] Asea, P. and S. Turnovsky (1998) "Capital income taxation and risk-taking in a small open economy", *Journal of Public Economics*, 68, 55-90.

[7] Andersen, T., H-J. Chung, and B. Sorensen (1999) "Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study", *Journal of Econometrics*, 91, 61-87.

[8] Billio, M., A. Monfort, and C.P. Robert (1999) "Bayesian estimation of switching ARMA models", *Journal of Econometrics*, 93, N.2, 229-256.

[9] Broze, L., O. Scaillet, and J-M. Zakoian (1998) "Quasi-Indirect Inference for Diffusion Processes", *Econometric Theory*, 14, 161-186.

[10] Calzolari, G., F. De Iori, and G. Fiorentini (1998) "Control Variants for Variance Reduction in Indirect Inference: Interest Rate Models in Continuous Time", *Econometrics Journal*, 1, C100-C112.

[11] Carrasco, M. and J. P. Florens (2000a) "Generalization of GMM to a continuum of moment conditions", *Econometric Theory*, 16, 797-834.

[12] Carrasco, M. and J. P. Florens (2002) "Efficient GMM Estimation Using the Empirical Characteristic Function", mimeo, University of Rochester.

[13] Carrasco, M., M. Chernov, J. P. Florens, and E. Ghysels (2001) "Estimating jump-diffusions with a continuum of moment conditions", working paper.

[14] Chernov, M. and E. Ghysels (2000) "A Study towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation", *Journal of Financial Economics*, 56, 407-458.

[15] Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen (2001) "Alternative Models for Stock Price Dynamics", mimeo, Business School of Columbia University.

[16] Chow, Y-F. and M. Liu (1999) "Long Swings with Memory and Stock Market Fluctuations", *Journal of Financial and Quantitative Analysis*, 34, 341-367.

[17] Chumacero, R. (1997) "Finite Sample Properties of the Efficient Method of Moments", *Studies in Nonlinear Dynamics and Econometrics*, 2, 35-51.

[18] Chung, C. (2001) "Testing Target Zone Models Using the Efficient Method of Moments", forthcoming in *Journal of Business and Economic Statistics.*

[19] Collard, F., P. Fève, and C. Perraudin (2000) "Solving and Estimating Dynamic Models under Rational Expectations", *Computational Economics*, 15, 201-221.

[20] Collard, F., P. Fève, F. Langot, and C. Perraudin (2001) "A Structural Model of US Aggregate Job Flows", forthcoming in *Journal of Applied Econometrics.*

[21] Dai Q. and K. Singleton (2001) "Specification Analysis of Affine Term Structure Models", *Journal of Finance*, 55, 1943-1978.

[22] Dhaene, G., C. Gouriéroux, and O. Scaillet (1998) "Instrumental Models and Indirect Encompassing", *Econometrica*, 66, 673-688.

[23] Devroye, L. (1986) *Non-Uniform Random Variate Generation*, Springer Verlag, NY.

[24] Dridi, R., E. Renault (1999) "Semiparametric Indirect Inference", working paper, Universite des Sciences Sociales de Toulouse.

[25] Eichenbaum, M., L.P. Hansen, and K. Singleton (1988) "A time series analysis of representative agent models of consumption and leisure choice under uncertainty", *The Quaterly Journal of Economics*, 51-78.

[26] Elerian, O., S. Chib, and N. Shephard (2001) "Likelihood Inference For Discretely Observed Nonlinear Diffusions", *Econometrica*, 69, 959-993.

[27] Duffie, D. and K. Singleton (1993) "Simulated Moments Estimation of Markov Models of Asset Prices", *Econometrica*, 61, 929-952.

[28] Fermanian, J-D. and B. Salanié (2001) "A Nonparametric Simulated Maximum Likelihood Estimation Method", working paper, Crest, France.

[29] Galindo, A. (2000) "Estimating Credibility in Colombia's Exchange-Rate Target Zone", *Journal of Development Economics,* 63, 473-484.

[30] Gallant, A. R (1980) "Explicit Estimators of Parametric Functions in Nonlinear Regression", *Journal of the American Statistical Association*, 75, 182-193.

[31] Gallant, A. R., D. Hsieh, and G. Tauchen (1997) "Estimation of Stochastic Volatility Models with Diagnostics", *Journal of Econometrics,* 81, 159-192.

[32] Gallant, A. R., C.-T. Hsu, and G. Tauchen (1999) "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance", *The Review of Economics and Statistics*, 81, 617-631.

[33] Gallant, A. R. and J. R. Long (1997) "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-squared", *Biometrika*, 84, 125-141.

[34] Gallant, R. and D. Nychka (1987) "Semi-Nonparametric Maximum Likelihood Estimation", *Econometrica*, 55, 363-390.

[35] Gallant, A. R., and G. Tauchen (1996) "Which Moments to Match?", *Econometric Theory*, 12, 657-681.

[36] Gallant, A. R., and G. Tauchen (1998) "Reprojecting Partially Observed Systems with Application to Interest Rate diffusions", *Journal of the American Statistical Associations*, 93, 10-24.

[37] Gallant, A. R., and G. Tauchen (1999) "The relative efficiency of method of moments estimators", *Journal of Econometrics*, 92, 149-172.

[38] Gallant, A. R., and G. Tauchen (2001) "Efficient Method of Moments", mimeo, University of North Carolina and Duke University.

[39] Gallant, A. R., and G. Tauchen (2002) "Simulated Score Methods and Indirect Inference for Continuous-time Models", forthcoming in the *Handbook of Financial Econometrics.*

[40] Geweke, J. (1989) "Bayesian Inference in Econometric Models Using Monte Carlo Integration", *Econometrica*, 57, 1317-1339.

[41] Geweke, J. (1996) "Monte Carlo Simulation and Numerical Integration" in *Handbook of computational economics.* Eds H. Amman, D. Kendrick, and J. Rust. Elsevier, New York, 791-800.

[42] Ghysels, E., L. Khalaf, and C. Vodounou (2001) "Simulation Based Inference in Moving Average Models", forthcoming in *Annales d'Economie et de Statistique.*

[43] Gouriéroux, C., A. Monfort (1991) "Simulation Based Inference in Models with Heterogeneity", *Annales d'Economie et de Statistique*, N. 20/21, 69-107.

[44] Gouriéroux, C., A. Monfort (1993) "Simulation Based Inference, A survey with special reference to panel data models", *Journal of Econometrics*, 59, 5-33.

[45] Gouriéroux, C., A. Monfort (1996) *Simulation Based Econometric Methods*, CORE Lectures, Oxford University Press, New York.

[46] Gouriéroux, C., A. Monfort, and E. Renault (1993) "Indirect Inference", *Journal of Applied Econometrics*, 8, S85-S118.

[47] Hairault, J-O., F. Langot and F. Portier (1997) "Time to implement and aggregate fluctuations", *Journal of Economic Dynamics and Control*, 22, 109-121.

[48] Hajivassiliou, V. and D. McFadden (1998) "The Method of Simulated Scores for the Estimation of LDV Models", *Econometrica*, 863-896.

[49] Hajivassiliou, V. and P. Ruud (1994) "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of econometrics.* Eds: R. Engle and D, McFadden, North Holland, New York, 2384-2441.

[50] Hansen, L. (1982) "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

[51] Hansen, L. (1985) "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators", *Journal of Econometrics*, 30, 203-238.

[52] Heaton, J. (1995) "An empirical investigation of asset pricing with temporally dependent preference specifications", *Econometrica*, 63, 681-717.

[53] Iannizzotto, M. and M. Taylor (1999) "The Taget Zone Model, Non-Linearity and Mean Reversion: Is the Honeymoon Really Over?", *Economic Journal*, 109, C96-C110.

[54] Kamionka, T. (1998) "Simulated maximum likelihood estimation in transition models", *Econometrics Journal*, 1, C129-C153.

[55] Jonsson G. and P. Klein (1996) "Stochastic fiscal policy and the Swedish business cycle", *Journal of Monetary Economics*, 38, 245-268.

[56] Judd, K. (1998) *Numerical Methods in Economics*, MIT University Press, Cambridge.

[57] Laffont, J-J., H. Ossard, and Q. Vuong (1995) "Econometrics of First-Price Auctions", *Econometrica*, 63, 953-980.

[58] Lee, B-S. and B. Ingram (1991) "Simulation estimation of time-series models", *Journal of Econometrics*, 47, 197-205.

[59] Liesenfeld, R. and J. Breitung (1999) "Simulation Based Method of Moments", in *Generalized Method of Moment Estimation.* Eds: L. Matyas, Cambridge University Press.

[60] Liu, M. (2000) "Modeling Long Memory in Stock Market Volatility", *Journal of Econometrics*, 99, 139-171.

[61] Magnac, T., J-M. Robin, and M. Visser (1995) "Analysing Incomplete Individual Employment Histories Using Indirect Inference", *Journal of Applied Econometrics*, 10, S153-S169.

[62] McFadden, D. (1989) "A method of simulated moments for estimation of discrete response models without numerical integration", *Econometrica*, 47, 995-1026.

[63] McFadden, D. and K. Train (2000) "Mixed MNL Models for Discrete Response", *Journal of Applied Econometrics*, 15, 447-470.

[64] Michaelides, A. and S. Ng (2000) "Estimating the rational expectations model of speculative storage: A Monte Carlo comparison of three simulation estimators", *Journal of Econometrics*, 96, 231-266.

[65] Monfardini, C. (1998) "Estimating Stochastic Volatility Models through Indirect Inference", *Econometrics Journal*, 1, C113-C128.

[66] Nagypal, E. (2000) "Learning-by-Doing versus Learning About Match Quality: Can We Tell them Apart?", mimeo, Stanford University.

[67] Pakes, A. and D. Pollard "Simulation and the asymptotics of optimization estimators", *Econometrica,* 57, 1027-1057.

[68] Pastorello, S., E. Renault, and N. Touzi (2000) "Statistical Inference for Random-Variance Option Pricing", *Journal of Business and Economic Statistics*, 18, 358-367.

[69] Schennach, S. (2001) "Information-Theoretic Estimation of Latent Variable Models", mimeo, University of Chicago.

[70] Sieg, H. (2000) "Estimating a Bargaining Model with Assymmetric Information: Evidence form Medical Malpractice Disputes", *Journal of Political Economy*, 108, 1006-1021.

[71] Singleton, K. (2001) "Estimation of Affine Pricing Models Using the Empirical Characteristic Function", *Journal of Econometrics*, 102, 111-141.

[72] Smith, A. (1993) "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions", *Journal of Applied Econometrics*, 8, 563-584.

[73] Solibakke, P. B. (2001) "A Stochastic Volatility Model Specification with Diagnostics for Thinly Traded Equity Markets", *Journal of Multinational Financial Management*, 11, 385-406.

[74] Stern, S. (1997) "Simulation-Based Estimation", *Journal of Economic Literature*, XXXV, 2006-2039.

[75] Tauchen, G. (1997) "New Minimum Chi-Square Methods in Empirical Finance", in *Advances in Econometrics*, *Seventh World Congress*, eds. D. Kreps and K. Wallis, Cambridge University Press, Cambridge.

[76] Topa, G. (2001) "Social Interactions, Local Spillovers and Unemployment", *Review of Economics Studies*, 68, 261-295.

[77] Zhou, H. (2002) "Finite Sample Properties of EMM, GMM, QMLE, and MLE for a Square-Root Interest Rate Diffusion Model", *Journal of Computational Finance*, Vol. 5, No 2.