INVESTIGACIONES ECONÓMICAS. vol. XXI (3), 1997, 473-500

SOME METHODS FOR COMPARING FIRST ORDER ASYMPTOTICALLY EQUIVALENT ESTIMATORS

M. DOLORES DE PRADA
LUIS M. BORGE
Universidad de Valladolid

Some methods for comparing asymptotically equivalent estimators are reviewed. We revise higher order asymptotic theory principles as well as procedures for determining Edgeworth approximations to the distribution functions. Different methods for determining the degree of proximity between estimators in probabilistic terms are discussed. On the one hand, direct comparisons among estimators by means of the stochastic difference between them. On the other hand, comparisons of the degree of proximity between their distribution functions. In the last context we present a method proposed by Robinson (1988) which simplifies this type of comparisons because it do not need to calculate higher order approximations. (*JEL* C10)

1. Introduction

For most estimation problems different procedures are available. In order to be able to choose among them we set up certain criteria or statistical properties that a "good" estimator should verify. Unfortunately, in econometrics, as in many others contexts, finite sample statistical properties of most estimators can not be established because the exact distribution function of these estimators is usually unknown. In such situations, choice among alternative estimators is based on their *asymptotic properties*, that is, on the behavior of the estimator when the sample size, T, goes to infinity. Central limit theorems are employed to obtain an approximated distribution, known as *the asymptotic distribution*, which is used to calculate the properties

Research funded by "Dirección General de Enseñanza Superior (DGES)", reference number PB95-0292. We are most thankful to Marta Risueño for her valuable suggestions and comments which have helped to improve the presentation of the article, and two anonymous referees for helpful comments and suggestions.

of the estimator. In that way, if we can prove that an estimator is efficient by using the asymptotic distribution, we would say that it is asymptotically efficient or first order efficient (FOE).

Under suitable regularity conditions, the maximum likelihood estimator (MLE) is asymptotically efficient. Any other estimator with the same asymptotic distribution would be asymptotically efficient too. Many classical estimators have the same limiting distribution as the MLE and therefore, from the point of view of their statistical properties, are asymptotically equivalent. In those cases, it would be of interest to have some procedures for comparing among these estimators.

Most of the literature related with this subject is based on the use of more refined approximations to the distribution function of an estimator than the asymptotic distribution: higher order approximations. The most widely used method to obtain these approximations is the Edgeworth expansions. There are several procedures to calculate them, the main difficulty being not to calculate the formal expansion but to show that these expansions are indeed higher order approximations to the distribution function.

Once these higher order approximations are available we can define the concept of higher order efficiency relating to the estimator that maximizes the probability of being concentrated about the true parameter value, calculating the probability with the higher order distribution function. To find an estimator that satisfies this property, the class of estimators to be considered must be reduced to those that are *asymptotically median unbiased*[1] (AMU). That means that if the estimator under study is not AMU it must be modified in such a way that it belongs to this class. As most of the FOE estimators are also second order efficient (SOE), third order efficiency will be the instrument to discriminate among all FOE estimators.

If, instead of looking for the best estimator within a class, we are interested in comparisons of *relative* efficiency between two estimators, Robinson (1988) proposes a method for carrying out these comparisons that avoid the use of higher order approximations. It also provides a way of comparing the estimators directly, instead of the estimator distributions, by calculating the stochastic difference between them. Both methods are related and in fact, theorem 3 which provides sufficient conditions for calculating the relative higher order efficiency between

[1] See definition 2 below.

two estimators is based on the proof of theorem 1 for determining the stochastic order of magnitude of the difference. For this reason the stochastic difference will be described first.

The organization of the paper is as follows: section two discusses the higher order approximation to the distribution function of an estimator. We illustrate heuristically how the Edgeworth expansions are calculated for the simple case of the standardized sum of independent and identically distributed (i.i.d.) random variables. Also we sketch a generalization for more complex statistics. Section three presents the basic concepts of higher order efficiency theory. Section four introduces the concept of stochastic order difference between estimators, presenting a theorem for establishing the order of magnitude of such a difference. In section five we present two theorems which provide sufficient conditions for determining higher order relative efficiency and, finally, in section six we summarize our conclusions.

2. Higher order asymptotic theory

When no exact distribution is available, inferences are usually based on the asymptotic distribution of a statistic. The exact distribution of the estimator, $F_T(x)$, is approximated by a distribution function, $F(x)$, such that

$$\lim_{T\to\infty} |F_T(x) - F(x)| = 0$$

at all continuity points x of $F(.)$. $F(x)$ is called the asymptotic distribution.

However, in many situations there are several estimators with the same limiting distribution. In such cases the asymptotic distribution is not a useful tool for discriminating among them. Nevertheless, it is usually possible to improve these approximations by obtaining higher order approximations to the distribution functions.

DEFINITION 1 A function $G_T(x)$ is said to be an approximation of $(r+1)$th order to the distribution function $F_T(x)$ if:

$$\forall\, x \quad \lim_{T\to\infty} T^{\frac{r}{2}} |F_T(x) - G_T(x)| = 0.$$

That is, the difference between both distributions is of order $o(T^{-\frac{r}{2}})$. Therefore, the asymptotic distribution is a first order approximation.

In order to obtain higher order approximations, two methods have been widely used: the Edgeworth and the saddlepoint expansions. Both are based on the fact that the cumulant generating function of an estimator, defined as the logarithm of the characteristic function, can often be expanded in a power series where the terms are usually increasing powers of $T^{-\frac{1}{2}}$. Then, the density function can be approximated, using the Fourier inverse[2], keeping the first terms of the series. The two methods differ at the point at which the cumulant function is expanded: the zero value for the first method and the "saddlepoint" value that maximizes the integrand in the Fourier formulae for the second. This later approach[3] usually provides more accurate approximations, especially in the tail of the distribution, but it requires a complete knowledge of the cumulant function. The Edgeworth expansion, which can be calculated from the first cumulants, have been much more widely used to obtain higher order approximations. Its general availability for a wide variety of estimators has made it the principal tool for the study of *higher order efficiency theory*, as well as for other statistical applications such as improving classical procedures for setting confidence intervals and critical regions, the study of the accuracy of bootstrap approximations, etc.

Historically, Chebyshev (1890) can be regarded as the first contribution which provides an approximation to the density function of a sum of independent and identically distributed (i.i.d.) random variables. Also, but with no relation to Chebyshev's work, Edgeworth (1905) considers the same problem, but goes further in extending it to other statistics and sets the bases of what nowadays are known as Edgeworth expansions. Cramer (1928) was the first to establish conditions for the expansion of the density function of a sum of i.i.d. random variables to be valid; Wallace (1958) suggests that these approximations can be generalized to functions of the sample moments, but does not provide a rigorous proof of this. Later on, the works by Chambers (1967), Chibisov (1972) and Battacharya and Ghosh (1978) establish conditions under which Edgeworth expansions are valid asymptotic expansions under different assumptions on the functional form and the distribution of the sample moments. A recent review of the state of the subject can be found in Bhattacharya and Puri (1995).

In econometrics, the use of the Edgeworth expansions begins in the seventies, though Sargan (1964b) has already commented on the use

[2] See equation [2], below.

[3] The saddlepoint approximations have been used, among others, by Daniels (1954, 1956), Phillips (1978) and Durbin (1980).

of the Gram-Charlier expansions[4] for approximating the distribution of the two stage least square estimator (2SLSE). Sargan (1970), Sargan and Mikhail (1971) and Mariano (1973) represent the first contributions in this area. Sargan (1975) extends the result of Chambers (1967) and proves a theorem on the validity of the Edgeworth expansions for estimators that are asymptotically normal. This theorem is generalized by Sargan (1976) to include the case of endogenous lagged variables. The same situation is treated by Phillips (1977, 1980, 1986)[5] and Sargan and Satchel (1986). The Edgeworth expansions are used in other fields of econometrics, among others, by Taniguchi (1983, 1991) in the framework of time series models with normal errors and by Linton (1991, 1995) in the field of semiparametric regression.

It is worth recalling that these expansions are asymptotic and therefore, for a given sample size, we can say nothing about the goodness of the approximation. In fact, Edgeworth expansions are not very accurate at the tails of the distribution and, in some cases, other approximations can provide better results[6]. What has made Edgeworth expansions the basis for studying higher order asymptotic properties is that they can be calculated for most of the commonly used estimators and therefore are suitable for comparing asymptotically efficient estimators. In what follows, we illustrate heuristically, how Edgeworth expansions are calculated. Following Rothenberg (1984) and Hall (1992), we will look at the relatively simple case of the standardized sum of i.i.d. continuous random variables. We will also sketch a generalization of these approximations for more complex statistics.

2.1 Edgeworth expansion for the sum of i.i.d. random variables

Suppose $X_1, X_2, \cdots$ is a sequence of i.i.d. random variables, with zero mean and variance one, with defined moments up to the $(s+2)$-th order. Let $f_X(x)$ be the common density function and $\varphi_X(t)$ its charac-

[4]These expansions are directly related to the Edgeworth expansions. The Gram-Charlier expansions have not been widely used because they need a greater number of cumulants than the Edgeworth expansion.

[5]Phillips (1977) established his theorem about the existence of Edgeworth expansions with the presence of lagged dependent variables under rather different conditions to Sargan (1976). He also suggests that the proof stated by Chambers (1967) is not strictly correct and supplementary conditions may be needed to ensure that the result goes through.

[6]A review of different methods can be found in Phillips (1980, 1983), Rothenberg (1984) and Bayaert (1987).

teristic function. The cumulants of X are defined as $k_r = i^{-r} K_X^{(r)}(0)$, $r = 1, 2, ..., s+2$, $(i = \sqrt{-1})$, where $K_X^{(r)}(0)$ is the r-th derivative of the cumulant generating function $K_X(t) = \log \varphi_X(t)$ evaluated at $t = 0$. The relationship between the characteristic and cumulant functions allows us to express the r-th cumulant as a homogeneous polynomial in the first r moments, with the first and second cumulants equal to the mean and variance respectively. Therefore the first cumulant is zero and the second is one.

Let $f_{S_T}(x)$, $\varphi_{S_T}(t)$ and $K_{S_T}(t)$ be the density, characteristic and cumulant functions of the standardized sum $S_T = T^{-1/2} \sum X_T$ and let $k_{r,T}$ be the r-th cumulant of S_T. Since $K_{S_T}(t) = TK_X(t/\sqrt{T})$ then $k_{r,T} = T^{1-r/2} k_r$, $r > 2$, therefore, we can say that the high-order cumulants of S_T are small when the sample size is large and also the first cumulant is zero and the second is one.

Expanding the cumulant function around the origin and using the Taylor expansion of the exponential function, the characteristic function φ_{S_T} can be written as

$$\varphi_{S_T}(t) = e^{-\frac{1}{2}t^2} \left\{ 1 + \frac{k_3}{6\sqrt{T}}(it)^3 + \left(\frac{k_4}{24T}(it)^4 + \frac{k_3^2}{72T}(it)^6 \right) + \cdots \right\} \quad [1]$$

If $\varphi_{S_T}(t)$ is absolutely integrable, we can get the density function by the inverse Fourier transform

$$f_{S_T}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_{S_T}(t) dt. \quad [2]$$

The Edgeworth expansion is obtained by substituting expression [1] for the characteristic function in the Fourier formulae, dropping high-order terms.

Using the fact that, if $\phi(x) = 1/2\pi \int_{-\infty}^{\infty} \exp(-itx) \cdot \exp(-t^2/2) dt$ is the density function of a Normal random variable with zero mean and variance one, the r-th derivative of $\phi(x)$ is $\phi^{(r)}(x) = 1/2\pi \int_{-\infty}^{\infty} (-it)^r \exp(-itx) \cdot \exp(-t^2/2) dt$. We can get the $o(T^{-1})$ Edgeworth expansion for the density function of S_T as

$$f_{S_T}(x) \simeq \phi(x) - \frac{k_3}{6\sqrt{T}} \phi^{(3)}(x) + \frac{k_4}{24T} \phi^{(4)}(x) + \frac{k_3^2}{72T} \phi^{(6)}(x). \quad [3]$$

This approximation can be written in terms of Hermite polynomials

$H_r(x)$, defined as $H_r(x) = (-1)^r \phi(x)^{-1} \phi^{(r)}(x)$

$$f_{S_T}(x) \simeq \phi(x) \left[1 + \frac{k_3 H_3(x)}{6\sqrt{T}} + \frac{3k_4 H_4(x) + k_3^2 H_6(x)}{72T} \right] \qquad [4]$$

To obtain the Edgeworth expansion for F_{S_T}, we integrate the last equation

$$F_{S_T}(x) \simeq \Phi(x) - \phi(x) \left[\frac{k_3 H_2(x)}{6\sqrt{T}} + \frac{3k_4 H_3(x) + k_3^2(x) H_5(x)}{72T} \right] \qquad [5]$$

where $\Phi(x)$ is the distribution function of a standard normal. Calculating the Hermite polynomials, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$, $H_5(x) = x^5 - 10x^3 + 15x$. The $o(T^{-1})$ Edgeworth expansion for F_{S_T} is:

$$F_{S_T} \simeq \Phi(x) + \frac{p_1(x)\phi(x)}{\sqrt{T}} + \frac{p_2(x)\phi(x)}{T} \qquad [6]$$

where: $p_1(x) = -\frac{1}{6} k_3 (x^2 - 1)$ and $p_2(x) = -\frac{1}{24} k_4 (x^3 - 3x) - \frac{1}{72} k_3^2 (x^5 - 10x^3 + 15x)$.

This approximation adds two terms to the normal asymptotic distribution. The term of order $T^{-\frac{1}{2}}$ corrects for the main effect of skewness, since it depends on the third cumulant, while the term of order T^{-1} corrects for the main effect of kurtosis and the secondary effect of skewness.

To guarantee that these expressions are $o(T^{-1})$ Edgeworth approximations to the true distribution, it is necessary that the high-order terms eliminated in [1] are *small enough.* However, as we can see in Feller (1971, p. 541): .

$$E\left[|X|^{s+2}\right] < \infty \quad \text{and} \quad \lim_{|t| \to \infty} |\varphi(t)| < 1 \qquad [7]$$

are sufficient conditions for the Edgeworth expansion of order $o(T^{-\frac{s}{2}})$ to be a valid asymptotic expansion[7]. This means that, for a fixed s, when $T \to \infty$, $\forall\, x$:

$$P[S_T \leq x] = \Phi(x) + \frac{p_1(x)\phi(x)}{\sqrt{T}} + \frac{p_2(x)\phi(x)}{T} + \cdots + \frac{p_s(x)\phi(x)}{T^{\frac{s}{2}}} + o(T^{-\frac{s}{2}})$$

[7]We say that an asymptotic expansion is valid when, if we truncate the series with a given number of terms, the remainder have a smaller order than the last term included. For a more rigorous discussion of this, see, among others, Wallace (1958)

where $p_j(x)$, $j = 1, 2, \cdots, s$, are polynomials of degree 3j-1, with coefficients that depend on the cumulants $k_3, k_4, \cdots, k_{j+2}$, but not on T, and with $p_1(x)$ and $p_2(x)$ defined as above. The condition about the moments implies that, if we want an approximation of order $o(T^{-\frac{s}{2}})$, the existence of the moments up to order $s+2$ is necessary. Thus, to obtain an approximation of order $o(T^{-1})$, we need the existence of the moments up to order 4, since, as we saw before, the cumulants of third and fourth order appear in the Edgeworth formulae [6]. The second condition in [7], called *Cramer's condition*, holds if the random variable X has a proper density function. Edgeworth expansions can also be calculated for situations where Cramer's condition is not satisfied, see for example Bhattacharya and Rao (1975, chapter 5).

2.2 Edgeworth expansions: more general statistics

The extension of the previous results to other sample moments is relatively simple and it could be proved that, under suitable regularity conditions, the distribution functions of the sample moments have Edgeworth expansions if the high-order population moments exist. The situation is more complex when we try to generalize the result to the most common estimators used in econometrics, because most of them can not be considered as a simple sum of independent random variables. However, a wide variety of estimators are functions of the first sample moments, which are $\sqrt{T}$-consistent estimators of the population moments.

Let p_T be the vector of sample moments and suppose that an estimator is defined by $g(p_T)$, where g is a smooth function. Let $g_s(p_T)$ be the approximation of $g(p_T)$ obtained by considering the terms of the Taylor expansion of $g(p_T)$ around the population moments, $\mu = E[p_T]$, which includes up to the $(s-1)$-th derivative. Since the stochastic difference between the sample and population moments is of order $O_p(T^{-\frac{1}{2}})$[8], we expect that $g_s(p_T)$ would be *near enough* to $g(p_T)$, and also, that the moments defined from $g_s(p_T)$ would be near to the moments of $g(p_T)$. In this situation we can apply a generalization of the Delta method[9], to calculate the Edgeworth expansion. This method was

[8] $X_T = O_p(f_T)$ if $\forall \varepsilon > 0$, $\exists c$, T_o such that $\forall T > T_o$, $P[|X_T| > cf_T] < \varepsilon$.

[9] Delta method. Let $\hat{\theta}_T$ be an estimator of a parameter θ_o such that $\sqrt{T}(\hat{\theta}_T - \theta_o) \xrightarrow{d} N[0, D]$, then, if g is a smooth function :$\sqrt{T}[g(\hat{\theta}_T) - g(\theta_o)] \xrightarrow{d} N[0, G_o D G_o']$ where $G_o = \frac{\partial g}{\partial \theta'}\big|_{\theta_o}$.

proposed by Wallace (1958), and years later, Battacharya and Ghosh (1978), Phillips (1977) and Sargan (1975 and 1976) proved the validity of these approximations under different hypotheses about $g(.)$ and the distribution of p_T.

Consider the statistics $Y_T = \sqrt{T}[g(p_T) - g(\mu)]$ and $Y_T' = \sqrt{T}[g_s(p_T) - g(\mu)]$ and let W_T be the difference between Y_T and Y_T', i.e., W_T is $\sqrt{T}$ times the rest of the terms in the Taylor expansion of Y_T. Under some assumptions about W_T (see, for example, Ghosh (1994), lema 2.1) the $o(T^{-(s-2)/2})$ Edgeworth expansions of the distribution function of Y_T' is the same as the $o(T^{-(s-2)/2})$ Edgeworth expansions of the distribution function of Y_T. Then, if we want to calculate the $o(T^{-1})$ approximation of the distribution function of Y_T we can do so by obtaining the approximation for Y_T' with $s = 4$.

Since Y_T' is a polynomial in $(p_T - \mu)$, it is relatively easy to calculate its cumulants[10]. Let $k_{r,T}$ be the r-th cumulant of Y_T' calculated up to $o(T^{-1})$. The r-th cumulant is of order $T^{1-r/2}$ and the first and second cumulants are of the form: $k_{1,T} = aT^{-1/2} + o(T^{-1})$ and $k_{2,T} = b_2 + b_1 T^{-1} + o(T^{-1})$, where a and b_1 denote suitable coefficients that depend on the moments of the observation but not on T and b_2 is the variance of the first order asymptotic distribution of Y_T' (or Y_T).

These cumulants are often considered the cumulants of Y_T calculated by the delta method. Nagar (1959) was the first to use this technique for approximating the moments of the distribution of the k-class estimators in the context of the simultaneous equation models. He noted that these estimators can be expanded as an increasing power series in $T^{-1/2}$, whose coefficients are polynomials of random variables with bounded moments. The procedure consists in keeping the first terms of the expansion and calculating the moments of the truncated series. In general, the moments of the truncated series Y_T' are not good approximations to the moments of Y_T and therefore must be interpreted as the moments of a distribution which approximates the distribution of Y_T. Sargan (1974 and 1976) analyzes the conditions under which these approximations will be valid.

Once the cumulants of Y_T' are calculated, following the same steps as for the standardized sum of i.i.d. random variables[11] (expanding the cumulant function, using the Fourier inverse, ...), we obtain an

[10] *Some conditions about the function g and the moments of the original data are needed.*

[11] We need to rescale the variable in order to ensure that the statistic has zero mean

analogue of the Edgeworth expansion [6],

$$F_{Y'_T} \simeq \Phi(x) + \frac{q_1(x)\phi(x)}{\sqrt{T}} + \frac{q_2(x)\phi(x)}{T}, \qquad [8]$$

where q_i, $i = 1, 2$, is a polynomial of degree no more than $3i - 1$, with coefficients that depend on the moments of the original observations.

In order to guarantee that this expression actually provides the $o(T^{-1})$ Edgeworth expansion of the distribution of Y_T, strong conditions over the function g and the sample moments p_T are required (see, for example, Battacharya and Ghosh (1984), theorem 2).

EXAMPLE. Let $X_1, X_2, \ldots$ be i.i.d. random variables with zero mean μ and variance σ^2. Consider the Student's t-statistic $Y_T = \sqrt{T}(\bar{X} - \mu)/s$, where $\bar{X}$ and s^2 are respectively the mean and variance of the sample $(X_1, X_2, \ldots, X_T)$.
Let $W_i = (X_i - \mu)/\sigma$, then Y_T can be written as

$$Y_T = \sqrt{T}\bar{W}[1 + T^{-1}\sum(W_i^2 - 1) - \bar{W}^2]^{-1/2}.$$

Expanding this expression and keeping the first terms, we can define the statistic:

$$Y'_T = \sqrt{T}\bar{W}\left[1 - \frac{1}{2}T^{-1}\sum(W_i^2 - 1) + \frac{1}{2}\bar{W}^2 + \frac{3}{8}\{T^{-1}\sum(W_i^2 - 1)\}^2\right]. \qquad [9]$$

It can be proved that the rest of the terms of the expansion satisfy the conditions that allow that the approximation to the distribution function of Y_T can be calculated from the distribution of Y'_T up to order $o(T^{-1})$.

The cumulants of Y'_T up to order $o(T^{-1})$ can be calculated from [9]

$$\begin{aligned}
k_{1,T} &= -\frac{1}{2}T^{-1/2}\gamma + o(T^{-1}) \\
k_{2,T} &= 1 + \frac{1}{4}T^{-1}(7\gamma^2 + 12) + o(T^{-1}) \\
k_{3,T} &= -2T^{-1/2}\gamma + o(T^{-1}) \\
k_{4,T} &= T^{-1}(12\gamma^2 - 2\delta + 6) + o(T^{-1}),
\end{aligned}$$

where γ and δ are the third and fourth cumulants of W_i respectively. From the definition of W_i, $\gamma = \sigma^{-3}k_3$, $\delta = \sigma^{-4}k_4$, where k_3 and k_4 are the third and fourth cumulants of X_i.

and unit variance.

The $o(T^{-1})$ Edgeworth expansion to the distribution function of Y_T is given by

$$F_{Y_T} \simeq \Phi(x) + T^{-1/2} q_1(x)\phi(x) + T^{-1} q_2(x)\phi(x),$$

where

$$q_1(x) = \frac{k_3}{6\sigma^3}(2x^2+1) \qquad \text{and}$$
$$q_2(x) = x\left[\frac{1}{12}\frac{k_4}{\sigma^4}(x^2-3) - \frac{1}{18}\frac{k_3^2}{\sigma^6}(x^4+2x^2-3) - \frac{1}{4}(x^2+3)\right].$$

Nagar's technique for approximating the moments is not the only method for constructing Edgeworth expansions, although it is the most common because it is the simplest. If, for example, the characteristic function of an estimator is known, that information may be used to solve the problem directly by expanding the characteristic function in a power series and applying the Fourier inversion formulae to the first terms of the expansion. This procedure is used, among others, by Anderson and Sawa (1973) to obtain the distribution of k-class estimators in simultaneous equation models, and by Taniguchi (1983, 1991) for the derivation of Edgeworth expansions for time series models under normality. In contrast, there are certain types of estimators which, due to their characteristics, are excluded from this type of procedures and require specific methods for calculating the Edgeworth approximation. For a recent survey of other methods and their applications, see Bhattacharya and Puri (1995).

In conclusion, Edgeworth approximations are not so difficult to calculate using some of the available procedures. Proofs that these approximations are valid are much more difficult and require stronger conditions than for calculating the formal expansions.

3. Higher order efficient estimators

In the statistical literature there are several ways of defining the concept of asymptotic efficiency or first order efficiency (FOE). For a long time, the concept has been based on the asymptotic variance of an estimator. It was thought that if we only consider the class of $\sqrt{T}$-consistent asymptotically normal estimator, $\hat{\theta}_T$, of a parameter θ and with i.i.d. observations, the variance of the limiting distribution of $\sqrt{T}(\hat{\theta}_T - \theta)$, $v(\theta)$, has a lower bound, that is the inverse of Fisher's information (per observation), $i(\theta) = E[(\partial \log f(x_i, \theta)/\partial\theta)^2]$, where

$f(x_i, \theta)$ is the common density function of the observations. If the asymptotic variance, $v(\theta)$ attained the lower bound then the estimator was said to be FOE.

But unfortunately, this result is not true without further conditions on the estimator. We need the additional assumption that the convergence to normality of $\sqrt{T}(\hat{\theta}_T - \theta)$ be uniform in compact intervals of θ, where uniformity here means: given any compact set Θ and $\varepsilon > 0$,

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \sup_{x} \left| P[\sqrt{T}(\hat{\theta}_T - \theta) < x] - F(x|0, v(\theta)) \right| < \varepsilon$$

where $F(\cdot|\mu, \sigma^2)$ is the normal distribution function with mean μ and variance σ^2. Under suitable regularity conditions on the probability density of i.i.d. observations an estimator $\sqrt{T}$-consistent, and uniformly asymptotically normal (CUAN), $\hat{\theta}_T$, is FOE if its asymptotic variance is the inverse of Fisher's information (Rao, 1973, p. 531).

This definition implies that the distribution function of an FOE estimator, $\hat{\theta}_T$, is at least as concentrated about the true parameter value as any other CUAN estimator θ_T^* in the sense that

$$P[-a \leq \sqrt{T}(\theta_T^* - \theta) \leq b] \leq P[-a \leq \sqrt{T}(\hat{\theta}_T - \theta) \leq b] + o(1). \quad [10]$$

We adopt this last statement of maximizing the probability within a class as the definition of FOE estimator because its generality and convenience for introducing the treatment of higher order efficiency nowadays.

To discriminate among all FOE estimators, it seems logical to establish a similar definition of asymptotic efficiency but using *higher order approximations* to the distribution function of the estimator, what is known as *higher order efficiency.* We will restrict our attention to the class of regular best asymptotically normal (RBAN) estimators, where regularity here means that the stochastic expansions are possible. That guarantees the validity of the Edgeworth type expansion of distributions using asymptotic moments (Takeuchi and Morimune, 1985. p. 178).

The term second order efficiency (SOE) was introduced by Rao (1961), but the concept, as well as the first results in this area, occurs in Fisher (1925), which proposes an second order efficiency measure, based on the loss of information, to discriminate between different asymptotically efficient estimators in a problem of estimation involving a multinomial population. Fisher stated, without any sort of proof, that the maximum likelihood estimator is SOE. Rao (1961, 1963) makes major progress by proving Fisher's result, introducing a more useful measure of SOE, restricting attention to what he calls Fisher consistent estimators. This result will be referred to as the Fisher-Rao theorem. It was not until the seventies that this theory was fully developed. Ghosh and Subramayan (1974) extended the results of Fisher and Rao to exponential families[12]. Third order asymptotic efficiency (TOE) has been discussed, among others, by Pfanzagl and Wefelmeyer (1978) (who adopted the terminology), Pfanzagl (1980), Ghosh, Sinha and Wieand (1980), Akahira and Takeuchi (1981) and Takeuchi and Morimune (1985). Ghosh (1994) offers a recent review of the state of the subject.

We follow the Pfanzagl (1980) and Akahira and Takeuchi (1981) treatment to establish the basis of higher order efficiency theory. Their definition is based on the requirement that the estimator is higher order asymptotically median unbiased (AMU).

DEFINITION 2 The estimator $\hat{\theta}_T$ is r-th order AMU if it satisfies:

$$\lim_{T\to\infty} T^{\frac{r-1}{2}} \left| P[\hat{\theta}_T \leq \theta] - \frac{1}{2} \right| = \lim_{T\to\infty} T^{\frac{r-1}{2}} \left| P[\hat{\theta}_T \geq \theta] - \frac{1}{2} \right| = 0.$$

From this condition, the following definition can be established:

DEFINITION 3 (Higher order efficient estimator) If a r-th order AMU, $\hat{\theta}_T$, satisfies that $\forall a, b > 0$, $\forall \theta$:

$$\lim_{T\to\infty} T^{\frac{r-1}{2}} \left[\begin{array}{c} P\left[-a < \sqrt{T}(\hat{\theta}_T - \theta) < b\right] - \\ P\left[-a < \sqrt{T}(\theta^*_T - \theta) < b\right] \end{array} \right] \geq 0$$

for every r-th order AMU estimator θ^*_T, then $\hat{\theta}_T$ is called r-th order asymptotically efficient[13].

[12]The results about SOE of both, Rao and Ghosh and Subramayan, can be considered as third order efficiency (TOE) in the following sense: they used the asymptotic expansion of the covariance matrix of an estimator which proceeds in powers of T^{-1}, whereas the asymptotic expansion of the distribution of an estimator proceeds in powers of $T^{-1/2}$. Hence it appears natural to use the term TOE for what corresponds to Rao's SOE.

[13]For the multiparameter case, $\theta \in \Theta \subset R^p$, an estimator, $\hat{\theta}_T$, is r-th order asymptotic efficient if

$$\lim_{T\to\infty} T^{\frac{r-1}{2}} \left[P\left[\sqrt{T}(\hat{\theta}_T - \theta) \in C\right] - P\left[\sqrt{T}(\theta^*_T - \theta) \in C\right]\right] \geq 0$$

for any convex set C containing the origin.

In some cases, as for example in the exponential family, the condition of higher-order AMU can be replaced by the condition of higher-order unbiasedness without altering the conclusions but, in general, it is impossible to replace the AMU condition by that of asymptotic unbiasedness.

The RBAN estimators are not usually higher order AMU. It is possible, however, to modify them (correcting for the bias), so that they become higher order AMU. This correction makes it possible to compare asymptotic distributions by equalizing the locations of estimators.

As is shown in Pfanzagl and Wefelmeyer (1978), an estimator that is first order efficient is already second-order efficient as long as it satisfies:

(a) The asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta)$ is normal, with a zero mean and variance covariance matrix the asymptotic inverse of the information matrix $\mathcal{I}(\theta)$. Thus, $\hat{\theta}_T$ is asymptotically efficient.

(b) The distribution of $\sqrt{T}(\hat{\theta}_T - \theta)$ admits the Edgeworth expansion up to $o(T^{-\frac{1}{2}})$ order.

(c) The estimator is second-order median unbiased.

A comprehensive treatment of this result can be found in Akahira and Takeuchi (1981) or in Battacharya and Deker (1990).

Any RBAN estimator adjusted to be second-order median unbiased, satisfies conditions a), b) and c), and therefore will be second order efficient (SOE). That means that the second term of the asymptotic expansion of their distribution functions are the same. Thus, since second order efficiency fails to distinguish among first order efficient (FOE) estimators, we will need to use the concept of third order efficiency to discriminate among them. Then, we have to compare the third order $O(T^{-1})$ terms of the asymptotic expansion. The natural question to ask is if the MLE, that under some regularity conditions is FOE, is also TOE.

Under suitable conditions, the bias-adjusted MLE is TOE within the class of RBAN estimators modified to be third order unbiased. This result is discussed for i.i.d. sampling, among others, by Ghosh and Subramayan (1974) and Pfangzagl and Wefelmeyer (1978, 1979) for cases where sufficient statistics exist. Akahira and Takeuchi (1981) establish more general results for the multiparameter exponential family. It proves the third-order efficiency of the extended MLE within the class of extended RBAN estimators modifies to be third order median unbiased, where an extended estimator means an estimator that is a special type of function of a sufficient statistic. Takeuchi and Morimune (1985) generalize this result to the case of non i.i.d. sample in the context of the multivariate normal regression model.

Of course, adjusted MLE is not the only optimal estimator. Any bias adjusted estimator which has the same $o(T^{-1})$ approximation to the distribution function as the adjusted MLE, is also TOE. However, these third order properties allow us to discriminate among FOE estimators. For example, the three state least square estimator (3SLSE) in the context of linear in variables simultaneous equations (LIVSE) models is dominated in third order efficiency by the full information maximum likelihood estimator (FIMLE) when both are properly corrected for the bias (Takeuchi and Morimune 1985).

It is important to emphasize that we can not say anything about the higher order efficiency of the MLE itself. The above results indicate that, if we correct for the bias, the MLE is TOE. Then, it is necessary to remark that *corrections for the bias* are essential if we want to be able to compare asymptotically equivalent estimators by their third order efficiency properties. In fact, the results in the literature comparing higher order expansions of the distribution of RBAN estimators fail to establish an order among them (see, for example, Anderson 1974 or Fujikoshi et al. 1982). However, there is a way to obtain *relative higher order efficiency* results avoiding the use of the asymptotic expansions. The method was introduced by Robinson and will be described in section 5. First, we will discuss a concept related to it and which is also useful for making relative comparisons between FOE estimators: the *stochastic difference between estimators.*

4. Stochastic order of magnitude of the difference between estimators

A completely different method of comparing asymptotically equivalent estimators is based on measuring how close alternative estimators are to another estimator considered as optimal (the MLE for example). One way of doing so consists in calculating the stochastic order of magnitude of the difference between the estimator being studied and the optimal estimator. This procedure is especially useful when there are different forms, from the computational point of view (Newton-Raphson, Gauss-Newton, etc.), to approximate a particular estimator, and we want to compare how good the approximations at each iteration are. But it could also be applied to compare any kind of asymptotically equivalent estimators.

DEFINITION 4 Let $\hat{\theta}_T$ and $\tilde{\theta}_T$ be two consistent estimators of the true parameter, θ_o. If $\hat{\theta}_T - \tilde{\theta}_T = O_p(l_T)$, we will say that l_T is the stochastic order of magnitude of the difference between $\hat{\theta}_T$ and $\tilde{\theta}_T$. That is:

$$\forall \varepsilon > 0 \; \exists C > 0, \; \exists T_o \text{ such that } P[\|\hat{\theta}_T - \tilde{\theta}_T\| > Cl_T] < \varepsilon \quad \forall T > T_o.$$

The stochastic difference l_T does not provide an absolute measure of "the goodness" of an estimator. It only establishes how close an estimator is to a target estimator. But with this information we can set up an order within alternative estimators.

There are few references in the econometric literature about this topic before Robinson (1988). The concept had mainly been used to study whether the order of the difference between estimators was small enough to allow for the use of some asymptotic results in establishing the asymptotic distribution of an estimator. However, there was no special interest in calculating the specific value of l_T. One exception is Sargan (1964a) which, in the context of simultaneous equation models, establishes that the stochastic order of magnitude of the difference between the full information maximum likelihood estimator (FIMLE) and the three stages least squares estimator (3SLSE) is $O_p(T^{-1})$. In a similar context, Robinson (1988) proves that the stochastic difference between the pseudo maximum likelihood estimator (PMLE) and the minimum distance estimator (MDE) is $O_p(T^{-\frac{3}{2}})$, meaning that, for large enough T, the MDE is a better approximation to the PMLE than the 3SLSE.

Robinson (1988) also provides the stochastic order of magnitude of the difference for a wide variety of estimators, especially for many of the iterative procedures suggested in the econometric literature. He improves the result of Hosoya and Taniguchi (1982) about the order of magnitude of the stochastic difference between the k-th iterated Newton-Raphson estimator and the PMLE. Besides the comparisons among different estimation procedures, Robinson (1988) establishes a general simple rule to obtain the order of magnitude of the difference between estimators.

THEOREM 1 *Let $f_T(\theta)$ and $g_T(\theta)$ be two functions such that, evaluated at $\hat{\theta}_T$ and $\tilde{\theta}_T$ respectively, they are zero, that is: $f_T(\hat{\theta}_T) = 0$ and $g_T(\tilde{\theta}_T) = 0$. Let $N(\varepsilon)$ be a neighbourhood of the parameter to* *be estimated* θ_0, $N(\varepsilon) = \{\theta : \|\theta - \theta_0\| < \varepsilon\}$. *Under the following assumptions:*

A.1 $f_T(\theta)$*has a zero in* $\hat{\theta}_T$, *such that*

$$\hat{\theta}_T = \theta_0 + o_p(1).$$

A.2 $g_T(\theta)$ *has a zero in* $\tilde{\theta}_T$, *such that*

$$\tilde{\theta}_T = \theta_0 + o_p(1)$$

and for some $\varepsilon > 0$ *the derivatives* $G_T(\theta) = \frac{\partial}{\partial\theta} g_T(\theta)$ *exist in the neighbourhood of* θ_0, $N(\varepsilon)$, *and verify that* $\forall\delta > 0$,

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} P[\sup_{\theta\in N(\varepsilon)} \|G_T(\theta) - G_T(\theta_0)\| \;>\; \delta] = 0$$
$$G_T(\theta_0) = G + o_p(1), \quad G \textit{ nonsingular.}$$

Then

$$\hat{\theta}_T - \tilde{\theta}_T = O_p\left(\|f_T(\hat{\theta}_T) - g_T(\hat{\theta}_T)\|\right) = O_p\left(\|g_T(\hat{\theta}_T)\|\right)$$

The theorem says that the stochastic order of magnitude of the difference between two consistent estimators can be obtained as the order of magnitude of the function that serves to determine an estimator, but evaluated in the other estimator. Thus, to calculate such differences, it is not necessary for the estimators to have explicit form.

The estimating functions $f_T(\theta)$ and $g_T(\theta)$ might arise from differentiating the likelihood or other objective function. One sufficient condition for the double limit condition in A2 is that $G_T(\theta)$ is uniformly continuous in large T with probability arbitrarily close to one.

EXAMPLE: The stochastic difference between the PMLE and the MDE (Robinson 1988).
Consider the linear in variables simultaneous equation system $AX_T = U_T$, where $A = (B, C)$ is the matrix of parameters, $X_T = (Y_T, Z_T)$ the matrix of observations and U_T the matrix of residuals. Suppose that the parameters of the system are a function of an unrestricted parameter vector θ, i.e., $B = B(\theta)$ and $C = C(\theta)$, and let $\Pi = -B^{-1}C$ be the parameters of the reduced form of the system $Y_T = Z_T\Pi' + V_T$. We suppose that the classical assumptions for this type of models hold.

The MDE and the PMLE estimators are the solutions of the estimating functions

$$\begin{aligned} g_T(\theta) &= \operatorname{tr} A_i \begin{pmatrix} \Pi \\ I \end{pmatrix} \frac{Z_T' X_T}{T} A' B^{-1'} \widetilde{\Omega}_T^{-1} B^{-1} \qquad \text{and} \\ f_T(\theta) &= \operatorname{tr} A_i \begin{pmatrix} \Pi \\ I \end{pmatrix} \frac{Z_T' X_T}{T} A' B^{-1'} \Omega^{-1} B^{-1} \end{aligned}$$

where $A_i = (\partial/\partial\theta_i)A$, θ_i being θ's ith element and $\widetilde{\Omega}_T$ is the estimator of the covariance matrix of the reduced form disturbance, Ω, calculated from the ordinary least square residuals, i.e., $\widetilde{\Omega}_T = T^{-1}\widetilde{V}_T'\widetilde{V}_T$, with $\widetilde{V}_T = Y_T - Z_T\widetilde{\Pi}_T'$ and $\widetilde{\Pi}_T = Y_T'Z_T(Z_T'Z_T)^{-1}$.

It is quite simple to prove that the estimating functions verify assumptions A1 and A2. Then, the stochastic order of magnitude of the difference between the PMLE and the MDE can be calculated from the order of magnitude of

$$f_T(\theta_T^{PMLE}) - g_T(\theta_T^{PMLE}) = \operatorname{tr} \widehat{B}^{-1}\widehat{A}_i \begin{pmatrix} \widehat{\Pi} \\ I \end{pmatrix} \frac{Z_T' X_T}{T} \widehat{A}' \widehat{B}^{-1'} (\widehat{\Omega}_T^{-1} - \widetilde{\Omega}_T^{-1})$$

where $\widehat{}$ indicates evaluated at θ_T^{PMLE}. As

$$\begin{aligned} &\widehat{B}^{-1}\widehat{A}_i \begin{pmatrix} \widehat{\Pi} \\ I \end{pmatrix} \frac{Z_T' X_T}{T} \widehat{A}' \widehat{B}^{-1'} = O_p(T^{-1/2}) \qquad \text{and} \\ &(\widehat{\Omega}_T^{-1} - \widetilde{\Omega}_T^{-1}) = O_p(T^{-1}) \end{aligned}$$

we can conclude that

$$\begin{aligned} \theta_T^{PMLE} - \theta_T^{MD} &= O_p(\| f_T(\theta_T^{PMLE} - g_T(\theta_T^{MD}))\|) = \\ O_p(T^{-1/2})O_p(T^{-1}) &= O_p(T^{-3/2}). \end{aligned}$$

As Rothenberg(1984) suggests, if the stochastic difference between two estimators is of a given order, it is reasonable to suppose that their distribution functions also differ by that order. Unfortunately, there is not a direct connection between both statements. Therefore, another way of comparing asymptotically equivalent estimators with respect to an optimal estimator would be to compare *how close* their distribution functions are. We will discuss this approach in the next section, where we will establish the concept of *higher order relative efficiency* between estimators and also, we will explain the method proposed by Robinson

(1988) for carrying out this type of comparison which is based on the proof of the previous theorem.

5. Comparisons of Relative Higher Order Efficiency

As we have seen in section three the study of higher order efficiency requires, first of all, that the distribution function of the estimators under consideration admits an Edgeworth approximation, which requires a complex proof to show that the error committed is of the required order, i.e., that the approximation is valid. Secondly, to have a higher order efficient statistic or to demonstrate that two statistics are ordered with respect to higher order efficiency often entails some corrections for median bias. Furthermore, comparisons of relative efficiency between two estimators, performed for finite samples and based on the corresponding Edgeworth approximations do not usually establish definite conclusions.

Robinson, in the paper mentioned in the previous section, introduces a simpler method for establishing the closeness between the distribution functions of two estimators related to the proof of theorem 1. In other words, the method provides a way of checking up to which order two estimators have the same higher order relative efficiency. The advantage of this procedure is that, in order to make the comparison, it is not necessary to assume the existence of the Edgeworth expansion nor is it necessary to perform bias corrections. On the other hand, its main inconvenience is that this procedure does not allow us to obtain conclusions as to whether an estimator is third order efficient or not, in the sense of determining whether it maximizes the probability of belonging to a convex set containing the origin within a class of estimators. The information provided by the procedure is the order up to which the distribution functions of two statistics coincide and therefore have the same efficiency up to that order. We shall begin by establishing when can we say that two estimators of a vector of parameters have equal efficiency of a certain order and then, we shall set up two theorems giving sufficient conditions for equal higher order efficiency.

DEFINITION 5 Two estimators $\hat{\theta}_T$ and $\tilde{\theta}_T$ of a vector of parameters $\theta_o(p \times 1)$, have equal $(r+1)$-th order efficiency if $\forall r \geq 0$, integer:

$$\lim_{T\to\infty} \left\{ T^{\frac{r}{2}} \left| P\left[\sqrt{T}(\hat{\theta}_T - \theta_o) \in A\right] - P\left[\sqrt{T}(\tilde{\theta}_T - \theta_o) \in A\right] \right| \right\} = 0 \quad [11]$$

for all compact convex subsets $A \subset R^p$. That is, the difference between both probabilities is $o(T^{-\frac{r}{2}})$.

The main contribution of Robinson (1988) to higher order efficiency theory is to provide sufficient conditions for [11] to hold. Using them we can avoid calculating the higher order asymptotic distribution of the estimators and therefore we can show the equal higher order efficiency between them without having to assure the existence of the Edgeworth expansions. These conditions are presented in the next theorem.

THEOREM 2 [14] *The conditions*

(i) $\sqrt{T}(\tilde{\theta}_T - \theta_o)$, *admits a density function that is bounded uniformly in large* T.

(ii)

$$\lim_{T\to\infty} T^{\frac{r}{2}} P\left[T^{\frac{r+1}{2}} \log \sqrt{T}\|\hat{\theta}_T - \tilde{\theta}_T\| \geq 1\right] = 0 \qquad [12]$$

are sufficient for $\hat{\theta}_T$ *and* $\tilde{\theta}_T$ *to have equal* $(r+1)$*-th order efficiency.*

The proof of this theorem is based on bounding the difference of probability given in equation [11].

PROOF:

$$\begin{aligned} P[\sqrt{T}(\hat{\theta}_T - \theta_o) &\in A] = \\ P[\sqrt{T}(\hat{\theta}_T - \theta_o) &\in A \wedge \sqrt{T}(\tilde{\theta}_T - \theta_o) \in A \cup A_\delta] \\ +P[\sqrt{T}(\hat{\theta}_T - \theta_o) &\in A \wedge \sqrt{T}(\tilde{\theta}_T - \theta_o) \notin A \cup A_\delta] \end{aligned} \qquad [13]$$

where A_δ is a reduced δ-neighbourhood[15] of A, and therefore, A and A_δ are disjointed, and if $\sqrt{T}(\hat{\theta_T} - \theta_o) \in A$ and $\sqrt{T}(\hat{\theta}_T - \theta_o) \notin A \cup A_\delta \Longrightarrow \sqrt{T}\|\hat{\theta}_T - \tilde{\theta}_T\| > \delta$.

We can, therefore, write [13] as:

$$P[\sqrt{T}(\hat{\theta}_T-\theta_o) \in A] \leq P[\sqrt{T}(\tilde{\theta}_T-\theta_o) \in A\cup A_\delta]+P[\sqrt{T}\|\hat{\theta}_T-\tilde{\theta}_T\| \geq \delta]$$

that is:

$$\begin{aligned} &P[\sqrt{T}(\hat{\theta}_T - \theta_o) \in A] - P[\sqrt{T}(\tilde{\theta}_T - \theta_o) \in A] \leq \\ &P[\sqrt{T}(\tilde{\theta}_T - \theta_o) \in A_\delta] + P[\sqrt{T}\|\hat{\theta}_T - \tilde{\theta}_T\| > \delta]. \end{aligned} \qquad [14]$$

[14] Forms part of Theorem 7, p.538, Robinson (1988)

[15] We call $A_\delta = \{x \in R^p \| \; 0 < d(x, A) < \delta\}$ a reduced δ-neighbourhood of a compact subset $A \subset R^p$, where $d(x, A)$ indicates the distance from the point x to the subset A.

On the other hand, if $\sqrt{T}(\tilde{\theta}_T - \theta_o)$ satisfies condition (i), the term $P[\sqrt{T}(\tilde{\theta}_T - \theta_o) \in A_\delta]$ is $O(\delta)$, since for a sufficiently large T we can find a value M such that $P[\sqrt{T}(\tilde{\theta}_T - \theta_o) \in A_\delta] \leq \delta M$. Choosing $\delta = T^{-r/2}(\log\sqrt{T})^{-1}$, the first term of the right hand side of the inequality [14] is of the required order and from condition (ii), the second term is too•

To summarize, in order to prove that two estimators have the same $(r+1)$-th order efficiency, it will be sufficient to prove that they satisfy conditions (i) and (ii) of theorem [2].

Condition (i) excludes lattice distributions. Condition (ii) is sufficient for $\hat{\theta}_T - \tilde{\theta}_T = o_p(T^{-(r+1)/2})$, which implies $\hat{\theta}_T - \tilde{\theta}_T = O_p(T^{-\alpha/2})$ if $\alpha \leq r+1$. But the theorem does not say that a given stochastic difference is necessary or sufficient for a certain level of efficiency equivalence. Nevertheless there is a close relationship between the O_p results and the degree of equal higher order efficiency. In general, $\hat{\theta}_T - \tilde{\theta}_T = O_p(T^{-\alpha/2})$ and $r+1 < \alpha$ suggest, under suitable additional conditions, equal $(r+1)$-th order efficiency. This condition (ii) may be verified using the following theorem.

Theorem 3 *Let $N(\varepsilon) = \{\theta : \|\theta - \theta_o\| < \varepsilon\}$ and let $\hat{\theta}_T$ and $\tilde{\theta}_T$ be estimators of θ_o which satisfy the assumptions A.1 and A.2 of theorem 1. The conditions:*

$$\begin{aligned} P\left[T^{\frac{r+1}{2}}\log\sqrt{T}\|g_T(\hat{\theta}_T\| \geq \varepsilon\right] &= o(T^{-\frac{r}{2}}) \\ P[\hat{\theta}_T \notin N(\varepsilon)] &= o(T^{-\frac{r}{2}}) \\ &\text{and} \\ P[\tilde{\theta}_T \notin N(\varepsilon)] &= o(T^{-\frac{r}{2}}) \end{aligned}$$

are sufficient for

$$\lim_{T\to\infty} T^{\frac{r}{2}} P\left[T^{\frac{r+1}{2}}\log\sqrt{T}\|\hat{\theta}_T - \tilde{\theta}_T\| > 1\right] = 0$$

The theorem is based on the proof of theorem 1.

Proof:

Let $f_T(\theta)$ and $g_T(\theta)$ be the estimating functions of $\hat{\theta}$ and $\tilde{\theta}$, which satisfy assumptions A.1 and A.2. Choosing $\varepsilon > 0$ to satisfy A2, with probability arbitrarily close to one, for a sufficiently large T_o and all $T > T_o$, $\hat{\theta}_T$ and $\tilde{\theta}_T$ belong to $N(\varepsilon)$ and there exists a $\tilde{G}_T$ such that:

$$g_T(\hat{\theta}_T) = \tilde{G}_T(\hat{\theta}_T - \tilde{\theta}_T). \qquad [15]$$

where $\tilde{G}_T = G + o_p(1)$. From [15]

$$\begin{aligned} P(\|\hat{\theta}_T - \tilde{\theta}_T\| > 1) &\leq P\left[\|\tilde{G}_T^{-1}\|\|g_T(\hat{\theta}_T)\| > 1\right] \\ &= P\left[\|g_T(\hat{\theta}_T)\| > \|\tilde{G}_T\| \wedge \|\tilde{G}_T\| \geq \varepsilon\right] \\ &+ P\left[\|g_T(\hat{\theta}_T)\| > \|\tilde{G}_T\| \wedge \|\tilde{G}_T\| < \varepsilon\right] \\ &\leq P\left[\|g_T(\hat{\theta}_T)\| \geq \varepsilon\right] + P\left[\|\tilde{G}\| < \varepsilon\right]. \end{aligned} \quad [16]$$

On the other hand, if $\hat{\theta}_T$ and $\tilde{\theta}_T \in N(\varepsilon) \quad \Rightarrow \quad \tilde{G}_T \geq \varepsilon$. Then

$$P(\|\tilde{G}_T\| < \varepsilon) \leq P\left(\hat{\theta}_T \notin N(\varepsilon)\right) + P\left(\tilde{\theta}_T \notin N(\varepsilon)\right)$$

therefore

$$\begin{aligned} P\left(T^{\frac{r+1}{2}} \log \sqrt{T}\|\hat{\theta}_T - \tilde{\theta}_T\| > 1\right) &\leq P\left(T^{\frac{r+1}{2}} \log \sqrt{T}\|g_T(\hat{\theta}_T)\| \geq \varepsilon\right) \\ &\quad + P\left(\hat{\theta}_T \notin N(\varepsilon)\right) \\ &\quad + P\left(\tilde{\theta}_T \notin N(\varepsilon)\right). \end{aligned} \quad [17]$$

Then, a sufficient condition for the left hand side to be $o(T^{-\frac{r}{2}})$is that the three terms of the right side of this inequality are of that order ·

From the previous theorems we can conclude that in order to prove that $\hat{\theta}_T$ and $\tilde{\theta}_T$ have equal $(r+1)$-th order efficiency, one has to prove that, for large enough T, $\sqrt{T}(\tilde{\theta}_T - \theta_o)$ has a bounded density function, and that the three conditions of theorem 3 are $o(T^{-\frac{r}{2}})$. This procedure allows us to prove that two estimators have the same relative higher order efficiency, without having to carry out asymptotic expansions or having to correct for the bias. This means that, although it may be a laborious task to prove the conditions established in theorem 3, this method represents a great simplification with respect to the traditional way of making higher order efficiency comparisons. Furthermore, this method permits comparisons of relative higher order efficiency up to any order.

When we have a reference estimator considered as *optimal* we can choose among alternative FOE estimators by calculating the relative higher order efficiency of each of them with respect to the *optimal*.

Making use of this argument, Robinson, in the LIVSE context, proved that the Minimum Distance estimator (MDE) has the same second order efficiency as the Pseudo Maximum Likelihood estimator (PMLE). Prada (1996) extended Robinson's result to the $(k-1)$-th iterated MDE, proving that it has the same k-th order efficiency as the PMLE.

As we mentioned earlier, Takeuchi and Morimune (1985) proved that in the LIVSE model the 3SLSE is third order inefficient. It is well known that iterating the estimator (I3SLSE) it does not converge to the PMLE[16]. However Morimune and Sakata (1993) proposes a modified 3SLS (M3SLS) that is third order efficient. The word 'modified' does not refer correcting for the bias[17] and it seems more natural to call it *four stage least square estimator* (4SLSE) because it is defined as follows: first estimating the 3SLSE and them reestimating not only the structural form residual covariance matrix as in the I3SLSE but also the estimator of the reduced form parameters. This result implies that 4SLSE has the same higher order efficiency as the first-iterated MDE. It would be interesting to compare the $(k-1)$-th iterated MDE with I3SLSE, but considering iterations in the sense defined by Morimune and Sakata.

EXAMPLE (Robinson 1988) The PMLE and the MDE have the same second order efficiency.

Let θ_T^{MD} and θ_T^{PMLE} be MDE and the PMLE defined as in example 2. The estimators satisfy assumptions A.1 and A.2, therefore the condition (a) $\sqrt{T}(\theta_T^{MD}-\theta_o)$ admits a density that is bounded uniformly in large T, (b) $P[\theta_T^{PMLE}\notin N(\varepsilon)]=o(T^{-1/2})$, (c) $P[\theta_T^{MD}\notin N(\varepsilon)]=o(T^{-1/2})$, and (d) $P[T\log\sqrt{T}\|g_T(\theta_T^{PMLE})\|\geq\varepsilon]=o(T^{-1/2})$, are sufficient for both estimators to have the same second order efficiency.

For any $\varepsilon>0$ there exists a $\delta>0$ such that

$$\begin{aligned} P[\theta_T^{PMLE}\notin N(\varepsilon)] \;\leq\; C_1\Big[&P[\|T^{-1}Z_T'Z_T-E(z_tz_t')\|\geq\delta] \\ &+P[\|T^{-1}V_{To}'V_{To}-\Omega\|\geq\delta] \\ &+P[\|T^{-1}Z_T'V_{To}\geq\delta]\Big] \end{aligned} \qquad [18]$$

[16] The stochastic order of magnitude between the PMLE and the k-th iterated 3SLSE is also $O_p(T^{-1})$, as for $k=1$.

[17] In fact, in this case, bias correction is not necesary because M3SLE and FIMLE have the same second order term of their asymptotic expansions and therefore a bias correction is unnecessary.

$$P[\theta_T^{MD} \notin N(\varepsilon)] \leq C_2 \Big[P[\|T^{-1}Z_T'Z_T - E(z_t z_t')\| \geq \delta] + P[\|T^{-1}V_{To}'V_{To} - \Omega\| \geq \delta] + P[\|T^{-1}Z_T'V_{To} \geq \delta] \Big] \quad [19]$$

$$P[T \log \sqrt{T} \|g_T(\theta_T^{PMLE})\| \geq \varepsilon] \leq C_3 \left[P[T \log \sqrt{T} \|T^{-1}Z_T'V_{To}\|^3 \geq \delta] + o(T^{-1/2}) \right] \quad [20]$$

The zero subscript indicate evaluated at θ_o. By Markov's inequality, with proper assumptions about the moments of z_t and v_{ot}, and making use of Von Bahr and Esseen inequality, one can prove that the right hand sides of [18], [19] and [20] are $o(T^{-1/2})$.

6. Conclusions

To discriminate among FOE estimators we use the concept of higher order efficiency which is usually based on the Edgeworth-type expansion of the estimator distribution function. Therefore, we have to restrict attention to the class of FOE estimators that admits stochastic expansion (RBAN). An estimator is said to be higher order efficient if it is maximally concentrated about the true parameter value, using the higher order approximation to its distribution function. In general, for this estimator to be found, it is necessary to adjust it for the bias to make it AMU.

As in the class of RBAN estimators, any adjusted for the bias FOE estimator is also SOE for establishing an order among them we need to compare third order terms of the asymptotic expansion of their distribution. The difficulties of this procedure arise from the need to calculate the Edgeworth expansion and, what is more complex, to prove that the expansion is valid.

Another way to compare FOE estimators is using the concept of relative higher order efficiency, which consists in obtaining the degree up to which the distribution functions of two estimators coincide. Robinson (1988) provides a method which does not need the higher order approximations for the distribution functions to be calculated (i.e. you do not need to assume the existence of the Edgeworth expansions). Moreover, the procedure permits a direct comparison of the estimators without having to make corrections for the bias. But with this method we can not say if an estimator is the best within its class.

In conclusion, both procedures provide different information about the higher order properties of estimators. Therefore, it is not possible to say which procedure is the best, but Robinson's approach seems to be the simplest for carrying out higher order comparisons.

References

Akahira, M. and K. Takeuchi (1981): "Asymptotic efficiency of statistical estimators: concepts and higher order efficiency", *Lecture Notes in Statistics* (Vol. 7), New York: Springer-Verlag.

Anderson, T.W.(1974): "An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system", *Journal of the Americal Statistical Association* 69, pp. 565-573.

Anderson, T.W. and T. Sawa (1973): "Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions", *Econometrica* 41, pp. 683-714.

Bayaert, A. (1987): "Distribución en muestra pequeña del estimador parcialmente restringido de la forma reducida: evidencia teórica y experimentación de Monte-Carlo", Tesis Doctoral, Facultad de Ciencias Económicas y Empresariales, Universidad Autónoma de Madrid.

Bhattacharya, R.N. and M. Denker (1990), *Asymptotics Statistics,* Birkhäuser, Boston.

Bhattacharya, R.N. and J. K. Ghosh (1978): "On the validity of the formal Edgeworth expansion", *Annals of Statistics* 6, pp. 434-451.

Bhattacharya, R.N. and M. L. Puri (1995): "Asymptotic expansions in statistics: a review of methods and appications", *Advances in Econometrics and Quantitative Economics* Maddala, G.S.; Phillips, P.C.B. and Srinivasan, T.N. (eds), Cambridge. Blackwel Publishers Inc., pp. 88-122.

Bhattacharya, R.N. and R.R. Rao (1975), *Normal Approximation and Edgeworth Expansion,* New York, Wiley.

Chambers, J.M. (1967): "On methods of asymptotic approximation for multivariate distributions", *Biometrika* 54, pp. 367-283.

Chebyshev, P.L. (1890): "Sur deux théorèmes Relatifs aux probabilitès", *Acta Mathematica* 14, pp. 305-315.

Chibisov, D.M. (1972): "An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion", *Theory of Probability and Applications* 17, pp. 620-630.

Cramér, H. (1928): "On the composition of elementary errors", *Skand. Aktnarietidskrift* 11, pp. 13-74 and pp. 141-180.

Daniels, H.E. (1954): "Saddlepoint approximations in statistics", *Annals of Mathematical Statistics* 25, pp. 631- 650.

Daniels, H.E. (1956): "The approximate distribution of serial correlation coefficients", *Biometrika* 43, pp. 169-185.

Durbin, J. (1980): "The approximate distribution of serial correlation coefficients calculated from residuals on fourier series", *Biometrika* 67, pp. 335-350.

Edgeworth, F. Y. (1905): "The law of error", *Transactions of the Cambridge Philosophical Society* 20, pp. 36-65 and pp. 113-141.

Feller, W. (1971), *An Introduction to Probability Theory and its Applications,* (Vol. 2) New York: Wiley.

Fisher, R.A. (1925): "Theory of statistical estimation", *Proceedings of the Cambridge Philosophical Society* 22, pp. 700-725.

Fujikoshi, Y.; K. Morimune; N. Kunimoto and M. Taniguchi (1982): "Asymptotics expansions of the distribution of the estimates of coefficients in a simultaneous equation system", *Journal of Econometrics* 18, pp. 191-205.

Ghosh, J.K. (1994), *Higher Order Asymptotics,* Hayward, CA: Institute of Mathematical Statistics.

Ghosh, J.K.; B. K. Sinha and H. S. Wieand (1980): "Second order efficiency of the MLE with respect to any Bowl-Shaped loss function", *Annals of Statistics* 8, pp. 506-521.

Ghosh, J.K. and K. Subramanyam (1974): "Second order efficiency of maximun likelihood estimators", *Sankhya A* 36, pp. 325- 358.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion,* New York; Springer-Verlag.

Hosoya, Y. and M. Taniguchi (1982): "A central limit theorem for stationary processes and the parameter estimation of linear processes", *Annals of Statistics* 10, pp. 132-153.

Linton, O.B. (1991): "Edgeworth approximation in semiparametric regression models", Ph.D. Thesis, Departament of Economics, UC Berkeley.

Linton, O.B. (1995): "Second order approximation in the partially linear regression model", *Econometrica* 63, pp. 1079-1112.

Mariano, R.S. (1973): "Approximations to the distribution functions of the ordinary least-squares end two-stage estimators in the Case of Two included endogenous variables", *Econometrica* 41, pp. 67-77.

Morimune, K. and S. Sakata (1993): "Modified three-stage least squares estimator which is third-order efficient", *Journal of Econometrics* 57, pp. 257-276.

Nagar, A.L. (1959): "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations", *Econometrica* 27, pp. 573-595.

Pfanzagl, J. (1980): "Asymptotic expansions in parametric statistical theory", *Developments in Statistics* (vol. 3), Krishnaiah, P.R. (ed.); New York: Academic Press.

Pfanzagl, J. and W. Wefelmeyer, (1978): "A third order optimum property of maximum likelihod estimators", *Journal of Multivariate Analysis* 8, pp. 1-29.

Pfanzagl, J. and W. Wefelmeyer, (1979): Addendum to: "A third order optimum property of maximum likelihod estimators", *Journal of Multivariate Analysis* 9, pp. 179-182.

Phillips, P.C.B. (1977): "A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators", *Econometrica* 45, pp. 1517-1534.

Phillips, P.C.B. (1978): "Edgeworth and Saddlepoint approximations in the first-order noncircular autoregression", *Biometrika* 65, pp. 91-98.

Phillips, P.C.B. (1980): "Finite sample theory and the distributions of alternative estimators of the marginal propersity to consume", *Review of Economic Studies* 47, pp. 183-224.

Phillips, P.C.B. (1983): "ERA's: a new approach to small sample theory", *Econometrica* 52, pp. 1505-1526.

Phillips, P.C.B. (1986): "Large deviation expansions in econometrics", *Advances in Econometrics* 5, pp. 199-226.

Prada, M.D. (1996): "Comparación de eficiencia de orden superior entre estimadores: resultados asociados al estimador de distancia mínima iterado en el contexto de las ecuaciones simultaneas lineales", Tesis Doctoral, Facultad de Ciencias Económicas y Empresariales, Universidad de Valladolid.

Rao, C.R. (1961): "Asymptotic efficiency and limiting information", *Proc. Fourth Berkeley Symp. Math. Statist. Probab;* Univ. California Press, Berkeley, pp. 531-546.

Rao, C.R. (1963): "Criteria of estimation in large samples", *Sankhya A* 25, pp. 189-206.

Rao, C.R. (1973), *Linear Statistical Inference and its Applications,* New York: John Wiley & Sons.

Robinson, P.M. (1988): "The stochastic difference between econometric statistics", *Econometrica* 56, pp. 531-548.

Rothenberg, T.J. (1984): "Approximating the distributions of econometric estimators and test statistcs", *Handbook of Econometrics* (Vol. 2), Griliches, Z. and Intriligator, M.D. (eds.), Amsterdam: North-Holland, pp. 881-935.

Sargan, J.D. (1964a): "Three-stage least squares and full maximun- likelihood estimates", *Econometrica* 32, pp. 77-81.

Sargan, J.D. (1964b): "An approximate distribution of the two stage least squares estimators", *Abstract in Econometrica* 32, p. 660.

Sargan, J.D. (1970): "The finite sample distribution of FIML estimators", (Paper presented to The World Congress of the Econometric Society, Cambridge).

Sargan, J.D. (1974): "On the validity of Nagar's expansions for the moments of econometric estimators", *Econometrica* 42, pp. 169-176.

Sargan, J.D. (1975): "Gram-Charlier approximations applied to t-ratios of k-class estimators", *Econometrica* 43, pp. 326-346.

Sargan, J.D. (1976): "Econometric estimators and the Edgeworth approximation", *Econometrica* 44, pp. 421-448; and (1977) "Erratum" 45, p. 272.

Sargan, J.D. and W.M. Mikhail (1971): "A general approximation to the distribution of instrumental variables estimates", *Econometrica* 39, pp. 131-169.

Sargan, J.D. and S.E. Satchell (1986): "A theorem of validity of Edgeworth expansions", *Econometrica* 54, pp. 189-214.

Takeuchi, K., and K. Morimune (1985): "Third-order efficiency of the extended maximun likelihood estimators in a simultaneous equation system", *Econometrica* 53, pp. 177-200.

Taniguchi, M. (1983): "On the second order asymptotic efficiency of estimators of Gaussian ARMA processes", *Annals of Statistics* 11, pp. 157-169.

Taniguchi, M. (1991): "Higher order asymptotic theory for time series analysis", *Lecture Notes in Statistics* (Vol. 68), New York: Springer-Verlag.

Wallace, D.L. (1958): "Asymptotic approximations to distributions", *Annals of Mathematical Statistics* 29, pp. 635-654.

Resumen

En este artículo revisamos algunos métodos para comparar estimadores asintóticamente equivalentes, los principios de la teoría asintótica de orden superior y diferentes procedimientos para calcular la aproximación de Edgeworth a funciones de distribución. Discutimos diferentes formas de determinar el grado de proximidad entre estimadores: comparando directamente los estimadores a partir de la diferencia estocástica entre los mismos y comparando la proximidad entre sus funciones de distribución. En este último contexto presentamos el resultado de Robinson (1988) que simplifica notablemente las comparaciones ya que no es necesario calcular las aproximaciones de orden superior.