Journal of Quantitative Economics, Vol. 9 No. (1), January 2011

IMPORTANCE OF NON-PARAMETRIC DENSITY ESTIMATION IN ECONOMETRICS WITH ILLUSTRATIONS*

T. KRISHNA KUMAR[1]
JOSEPH M. MARKMANN[2]

Abstract

Econometrics deals with joint and conditional probability density functions and the mean value of the latter, the regression equation. When these densities take nonparametric form we get rich results. In this paper we retrace the work we did in mid seventies and recast them in today's perspectives. We also raise a few very interesting research issues that could be of some interest for newcomers to this field. Even before Efron wrote his first influential paper on bootstraps the senior author used simulated data generated from a Monte Carlo exercise and estimated the entire sampling distribution of estimators whose small sample distributions were not known and used them for statistical inference. One important result the authors showed then was that such an empirically estimated sampling distribution could approximate reasonably well the exact sampling distribution and its asymptotic approximation in known case derived by Anderson and Sawa in a two endogenous equations case. This raised hopes that the method should be useful in other cases when we do not know the exact small sample distributions of econometric estimators, setting the stage for bootstraps use in econometrics. One question that is often raised against the nonparametric inference is how one would formulate null hypotheses and test them. This issue is addressed in this paper and a suggestion is offered. Another issue that is addressed in this paper is the rate of convergence of different nonparametric density estimators to their true distribution

Keywords: Kernel density, Non-parametric inference, empirical likelihood, nonlinear least squares estimators, hypotheses testing in nonparametric case, empirical analogue of Rao's score

JEL Classification: C01, C14, C!5, C18

[1] T. Krishna Kumar is Retired Professor of Economics, Indian Statistical Institute, Bangalore and Guest Faculty, Indian Institute of Management, Bangalore. Email:tkkumar@gmail.com,

[2] Joseph M. Markmann is a CEO of an accounting firm, 2530 Huntingdon Pike, Huntingdon Valley, in Doyles Town, Pennsylvania.

This paper is written by Kumar and it relies on some results from an earlier work, the doctoral dissertation work of Markmann in early seventies under the guidance of Kumar. Hence Markmann is shown as a co-author of this paper.

* Paper presented at the Conference on Quantitative Approaches to Public Policy' jointly organized by Centre for Public Policy, Indian Institute of Management, Bangalore (IIMB); Indira Gandhi Institute of Development Research (IGIDR), Mumbai, India; and Centre for Globalisation Research, School of Business and Management, Queen Mary, University of London, UK, August 10-12, 2009 at Indian Institute of Management, Bangalore. Although this is an area of research the authors pursued more than three decades ago, the topic is found to be useful today in a variety of econometric applications including development economics and public policy (See Deaton (1991, 1993) and Pagan and Ullah (1999). We relate our earlier work to some recent developments, and identify several interesting unexplored problems on nonparametric estimation of density functions. The authors are grateful to an anonymous referee for his comments that helped us in revising the paper.

1. Introduction

In a very broad sense econometrics deals with interdependence between economic variables generated through an underlying probabilistic data generating process. This is true even with several econometric applications needed for public policy. The underlying probabilistic mechanism needs a description. In most of the cases this is done in two different ways. First, it is described by the joint conditional distribution of all the endogenous variables conditional on the given values of the exogenous variables. Second, it is described by a regression model with some unknown parameters entering into the specification of the conditional mean and conditional variance of the dependent variable, conditional on the independent variables. In addition one assumes that the error term or the residual term of the regression has a probability density with some additional unknown parameters. In either type of model there are two ways of specifying the unknown density function, either as a parametric density function or a nonparametric density function.[3]

In Section 2 we describe briefly parametric and nonparametric density estimation and nonparametric regression methods. In Section 3 we emphasize the importance of nonparametric density estimation. In Section 4 we present a few researchable issues on nonparametric kernel density estimation. In Section 5 we compare the parametric and nonparametric methods of estimating the probability density. In Section 6 we present analytic results of Anderson and Sawa that are needed for the next section. In Section 7 we present a comparison of parametric and nonparametric density estimation for the econometric problem of estimating the parameters of a two endogenous variable econometric model of Anderson and Sawa. In that Section we establish, with that example the advantage of using nonparametric kernel density estimation. Finally in Section 8 we make some concluding remarks.

2. Parametric and Nonparametric Density Estimation

In the development of classical statistical theory parametric probability distributions played a significant role. The Pearsonian system of frequency curves represents many probability density functions (bell shaped J-shaped and U-shaped, and one may see Johnson, Kotz, and Balakrishnan (1994)). A class of density functions belonging to an exponential family played an important role in classical theories and methods of statistics. If the probability density function belongs to the exponential family then there exist minimal sufficient statistics that attain the minimum variance limit set by the Cramer-Rao bound. In actuarial statistics, dealing with failure or failure rates, it is found to be inadequate to have regression models with errors distributed as a Normal distribution. The Normal distribution is replaced, in such cases, by a member of the exponential class of distributions giving rise to what is termed as Generalized Linear Model. This generalized linear model includes models with dependent variables having binomial and Poisson distributions, thus encompassing probit, logit, and Poisson regression models.

---

[3] There is another alternative which is semi-parametric, where a portion is specified parametrically and another portion nonparametrically.

In its simplest form the parametric method of estimating a density function consists of first assuming that the sample observations come from a population whose density function has a known (given) functional form and then estimating the parameters of that functional form. Understanding income and consumption distributions is quite crucial for public policy. Income and consumption distributions are so-estimated using Lognormal and Pareto density functions. If the prior knowledge on the density function is not rich enough to specify the functional form of the density function, the parametric method of density-function estimation consists of choosing one functional form from a number of different functional forms allowed as possible candidates for the density function and estimating the parameters of that functional form. Curve-fitting using the method of moments, *a la* Karl Pearson, is such a procedure. This procedure is described by Elderton and Johnson (1969). The functional forms of the members of the Pearsonian family of density functions appear on page 45 of Elderton and Johnson (1969). All these densities (f(x)) satisfy the differential equation

$$\frac{d\ln\{f(x)\}}{dx} = \frac{x-a}{b_0 + b_1x + b_2x^2} \qquad \text{... (2.1)}$$

These densities include such popular densities as the Normal, Student-t, Beta, Gama, Chi-Square, and F, and their truncated forms. A more recent source that gives a variety of parametric density functions is Johnson, Koltz, and Balakrishna (1994). For a given functional form of a Pearsonian density we shall distinguish between two methods of estimation of its parameters. One method uses the method of moments and the other that of maximum likelihood.

We shall distinguish between two broad types of nonparametric methods of estimating density functions. The first one is based on the notion that orthogonal polynomials form a basis in a function space and that any continuous density can be expressed as a linear function of those orthogonal polynomials. This method likewise has a long history starting from the 1904 contribution of Edgeworth (1904). The second method, which is more recent, is based on the contribution of Rosenblatt (1956). It consists of taking the sample density which gives a mass of l/n (where n is the sample size) to each sampled observation and a mass of zero elsewhere, and smoothing it by using a convolution with another weighting function or kernel. We shall label these methods, respectively, as the method of orthogonal functions and the kernel method. Within each of these broad categories of methods there may be many estimators, these corresponding to the various choices of orthogonal functions or of the kernel. The reader is referred to Wegman (1972) for a survey of the nonparametric methods, Prakasa Rao (1983), and Silverman (1992) for later and more rigorous treatment of the subject.

The concept behind a non-parametric Kernel type density estimator can be explained in intuitive terms. First, split the sample space into a large number of finite number of mutually exclusive and collectively exhaustible intervals or “windows”; second assume a prior probability density within that interval (the kernel); third, take the sample relative frequency in that interval and obtain the posterior density (for the interval) conditional on the observation lying in that interval. Finally obtain the weighted (the weights being the probabilities that the observation lies in each one of the intervals) sum the posterior densities so obtained for all intervals to arrive at the posterior density in the entire sample space.

Wegman (1972) noted that the choice of trigonometric functions as orthogonal functions used as kernels performed better than other choices of orthogonal functions. In our Monte Carlo investigations reported in Section 5 we use, therefore, as representative of the method of orthogonal functions, an estimator, due to Kronmal and Tarter, which uses trigonometric functions, specifically being a sum of sines and cosines. Details of this method can be found in Kronmal and Tarter (1968).

Given a sample $x_1,...x_n$ from a population with an unknown density function f(x), the kernel method of estimating the density function consists of choosing a kernel or weight function w(x) and a bandwidth h (that depends on n in such a way that $h \to o$ as $n \to \infty$ to obtain an estimate $\hat{f}_n(x)$ given by

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{j=1}^{n} w\left(\frac{(x-x_j)}{h}\right) \quad ... (2.2)$$

The measure of distance between $f_n(x)$ and the true density f(x) is taken to be the Mean Integrated Squared Error (MISE), $J(f_n)$, defined as

$$J(f_n) = \int_{-\infty}^{+\infty} f(x) - f_n(x)^{\ 2}\, w(x)dx \quad ... (2.3)$$

Here $w(x)$ is a weight function. Rosenblatt showed that to minimize J for a given kernel $w(x)$, the optimal bandwidth must satisfy the relation

$h = Kn^{-1/5}$ where

$$K = \frac{2^{2/5}\{\int w^2(v)dv\}^{1/5}}{\{\int (f''(x))^2 dx\}^{1/5}\{\int w(u)u^2 du\}^{2/5}} \quad ... (2.4)$$

Thus the optimal bandwidth depends on the kernel chosen and on the true unknown density function.

In our empirical investigation reported in Section 5 and 7 we consider four kernel estimators. The first kernel estimator is the so-called naive estimator or Rosenblatt estimator with the kernel

$$w(x) = \frac{1}{2} \text{ if } |x| \le 1 \text{ and } w(x) = o \text{ otherwise.}$$

For which

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} \quad ... (2.5)$$

Here $F_n(x)$ is the sample cumulative distribution function. The second kernel estimator is based on an optimal kernel among the class of all kernels that are of the form of a density function symmetric about zero with unit variance. This kernel is due to Epanechnikov (1969) and has w(x) in the form

$$w(x) = \frac{3}{4\sqrt{5}}(1 - \frac{x^2}{5}) \text{ if } |x| \le \sqrt{5}$$
$$= 0 \quad \text{otherwise} \quad ... (2.6)$$

The third kernel is based on a weight function which is h times the normal density with mean zero and standard deviation h. That is,

$$w\left(\frac{x}{h}\right) = \frac{1}{\sqrt{2\pi}} \text{Exp}\left(-\frac{x^2}{2h^2}\right) \qquad \text{... (2.7)}$$

And

$$\hat{f}_n(x) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{nh}\right)\sum_{i=1}^{n} \text{Exp}\left(\frac{(x-x_i)^2}{2h^2}\right) \qquad \text{... (2.8)}$$

This estimator was proposed by Specht (1971) who suggested approximating each of the exponential terms under summation using Taylor series and retaining only a certain number of terms. We instead use the closed form expression (2.8) and thus avoid any errors due to the approximation. The fourth kernel estimator is the Kromwal-Tartar (1968) estimator.

For the first two kernel estimators we used, where feasible, the optimal bandwidth given by $h=Kn^{-1/5}$ with K as given by (2.4). In the uniform distribution case this formula breaks down since $\int (f''(x))^2 dx = 0$. Here, we have determined optimal h experimentally through cubic interpolation. For the Specht estimator, the h which minimizes MISE becomes too complex to evaluate except when the underlying distribution is normal. Thus, for the normal distribution, we again determine optimal h values experimentally through cubic interpolation.

While the results reported in Section 5 were computed in 1974 (thirty six years ago) only for these four kernel densities there have been several other alternative nonparametric density estimators that have been suggested in the literature.[4] Just to give some examples, Adamowski (1989) performed simulation exercises, such as what we report in division V, and based on those he recommended using a symmetric kernel given by:

$$w(x) = \frac{3(1-x^2)}{4} \text{..................for } |x| \le 1 \qquad \text{... (2.9)}$$
$$= 0 \text{ elsewhere}$$

Such an estimator is called Adamowski kernel estimator. This is also called Epanechnikov-2 estimator, given its similarity in form to Epanechnikov's estimator.

The kernel for Bi-weight or quartic kernel density function is given by:

$$w(x) = \frac{15}{16}(1-x^2)^2 \text{ for } |x| < 1 \qquad \text{...(2.10)}$$
$$w(x) = 0 \text{ otherwise}$$

The kernel for triangular kernel density is given by:

---

4 We wrote a Fortran IV program in 1974 to perform all the numerical calculations. Today many statistical software programs, such as STATA, SYSTAT, etc incorporate kernel density estimation. The Kernel called rectangular in STATA and uniform in other literature is our naïve or Rosenblatt estimator, the one called normal in STATA is our Specht estimator, while the Cosine Kernel in STATA is similar to Kromwal and Tarter estimator. Epan 2 estimator in STATA is Adamowski estimator. In addition there are some other kernel estimators-Bi-weight, Parzen, and Traingular in STATA that we have not used. Hence anyone interested in extending our investigation with a larger set of kernel density functions or pursuing some of the interesting questions raised in this article can do so easily using GAUSS, STATA, and other econometric software.

$$\begin{aligned} w(x) &= 1-|x| \text{ for } |x| < 1 \\ w(x) &= 0 \text{ otherwise} \end{aligned} \qquad ...(2.11)$$

Standard kernel density estimation is known to create local biases that flatten peaks and lift up the troughs in the density. While working on this topic more than three and half decades ago we had wondered why the same Kernel should be used throughout the range of the variable, and why one should use the same window width for all intervals when we know the shape of the observed relative frequency. In fact one can think of varying either the function used for the kernel, *viz*, w(x) or one may use a variable window width with the same kernel, or one can vary both. One example of using different kernels in different ranges is a very old kernel estimator suggested by Parzen (1962) with the kernel given below:

$$w(x) = \left\{ \begin{array}{c} \frac{4}{3} - 8x^2 + 8|x|^3 \text{ for } |x| \le \frac{1}{2} \\ \frac{8(1-|x|^3)}{3} \text{ for } \frac{1}{2} < |x| \le 1 \\ 0 \text{ otherwise} \end{array} \right\} \qquad ...(2.12)$$

During the last two decades we do find in the literature some developments in this direction. Terrell and Scott (1992) showed the advantage of using a variable bandwidth. Guo (1991) demonstrated that fitting nonparametric densities to the frequency of floods using Adamovsky's symmetric kernel performs poorly at the tails. As extrapolation is needed at the tails to forecast floods the traditional nonparametric density estimation does not suit the purpose. He suggested using a different kernel at the tails and showed that extrapolation at the tails is better if an extreme value distribution kernel is used for the tail portion.

Kernel regression or nonparametric regression is actually the conditional mean of the dependent variable given the explanatory variables when the densities used to form the expectation are kernel densities. Thus the nonparametric regression is a function of the sample and does not involve any parameters. One can show that almost all curve-fitting methods between two variables, such as the nearest neighbour, spline regression etc are all special cases of nonparametric regressions for suitable choices of the kernel.

The Nadaraya-Watson nonparametric regression between y and x can be written as:

$$E(Y/X) = \hat{m}_h(x) = \frac{\sum_{i-1}^{n} \{ \frac{w\left(\frac{(x-X_i)}{h}\right) Y_i}{\sum_{i=1}^{n} w\left(\frac{(x-X_i)}{h}\right)} \}}{} \qquad ...(2.10)$$

As mentioned above, there is no reason why the same kernel and the same bandwidth should be used throughout the range for both the joint and marginal kernel densities, as in equation (2.10). Although it is argued that the choice of kernel is not as important as the choice of the bandwidth, that observation was based on using the same kernel in the entire range. It has been observed that the tail estimation can be improved by taking an extreme value distribution kernel. Hence it is worthwhile to consider modified Nadaraya-Watson estimators with different bandwidth and kernel choices for the two densities in the denominator and the numerator, and also with different kernels for different portions of the sample range. Specht (1991) introduced variable bandwidth density to arrive at General Regression Neural Network Model (GRNN). One may also consider an extension of Nadaraya-Watson nonparametric regression by using the notion of fractal regression and combining it with kernel estimation (Sen (2005)).

3. Why Nonparametric Inference?

Any particular parametric choice is quite narrow, even if such choice is made on the basis of approximating the true density by a linear function of a class of densities that span the space of continuous density functions, as in the case of Edgeworth's approximation. This is so as such approximations may not work equally well in the entire range. It is true that by making the model nonparametric we are expecting the sample to tell us everything about the underlying pattern. For that to work well we need a large sample. In data mining, where we are dealing with very large data sets, the data could be generated by a mixture of distributions. Fitting any one parametric density to such data could give rise to a poor fit of the data to the model. A nonparametric density specification could provide us a better pattern recognition and hence better model.[5] Any density we specify is an approximation to the unknown true density that must have generated the data. A description of observed relative frequency as an estimate of the unobservable theoretical probability density is quite common. The observed sample frequency is subject to sampling variability and any abnormality in the sample must be eliminated. For analytic convenience also it is desirable to smooth the rugged sample frequency. These considerations require that in each interval we consider a weighted average of a smooth density specified *a priori*, like the kernel, and the sample frequency. This is what the kernel density does. One may see Bierens (1987), Deaton (1993), Hardle and Linton (1994) and Pagan and Ullah (1999) for reviews on the use of nonparametric density estimation in econometrics.

There are a few major statistical problems that require density estimation. These are: smoothing of observed relative frequencies; estimating the probability density of an estimator whose exact distribution is not known; estimation of the critical values for hypotheses testing of the distribution of an estimator, whose exact small sample distribution is not known. The first type of problems arises quite frequently. An observed consumption or income distribution may have to be smoothed to purge any sample aberrations in the observed frequencies. While eliciting the prior probability beliefs a Bayesian may obtain such prior beliefs as discrete frequency distributions. She may want to smooth it and purge it of any aberration before proceeding to evaluate the posterior density functions. The second type of problem is also quite common in the econometrics literature as many econometric estimators do not have closed-form analytic

---

5 In one of the first verifications of this point the authors showed through a computer simulation experiment that if a true density did not belong to a known class of parametric densities the nonparametric density performs better than the parametric density (See Kumar and Markmann (1975a). The summary findings of that work were reported in Prakasa Rao (1983). The detailed results of the 1975 study are reproduced in Section 5 below.

probability density function when the sample sizes are finite (small).[6] One may first need to gain an understanding of the nature of such exact finite sample density functions of econometric estimators through simulation and fitting a smooth curve for the observed sample relative frequency.

The third type of problem is encountered when one tabulates the critical values for some test statistics whose small sample distributions are not known. At present such critical values are first calibrated from a few simulations with a few sample sizes using sample relative frequency distributions. Then a regression is run of such critical values on the sample size, and using that regression the tabulated values for other sample sizes are interpolated. One may see Mackinnon (1993) for an elaboration of this procedure. As an alternative one may obtain the smooth kernel density from the simulated sample results, and from such a kernel density one may calibrate the critical values. The extrapolating for other sample sizes may be done using these critical values in a regression. One interesting question worth exploring is which of the two procedures is preferable: the present practice as suggested by Mackinnon or estimating a kernel density to the simulated data and then derive the critical values? The latter procedure could reduce sampling variation through smoothing. The third alternative one uses is to obtain bootstraps estimates of standard errors. But in order to use those estimates to make the inference one must have some knowledge of the sampling distribution of the estimators, and kernel density functions are useful for this purpose.

4. Some Interesting Researchable Issues on Nonparametric Density Estimation[7]

Although it has been documented that changing the kernel did not significantly change the goodness of fit of a nonparametric density overall in terms of integrated squared error, it is possible that changing the kernel can alter how good the fit is in different segments of the density. Hydrologists find an extreme value distribution kernel more useful for the tails. Can one use data exploration for choosing the right kind of kernel, including a mixture of different types of kernels for different ranges extending the idea of Parzen? For instance can one use a bell shaped kernel in those ranges where the sample frequency is bell-shaped, an exponential kernel where the sample frequency is either declining or increasing, and a uniform kernel where the sample frequency is flat, and so on? Can one improve the kernel density estimator by varying the kernel, the same way an improvement was achieved by varying the bandwidth?

In any window the sample provides much more information than simple sample frequency in that window, which is used in kernel estimator of equation (2.2). For instance, in any window we know how many observations are in upper half of the window and how many are in

[6] It is in this context that the senior author started using density estimation, both parametric and subsequently nonparametric (Sengupta and Kumar (1965), Gapinski and Kumar (1972) and Kumar and Gapinski (1974), Kumar and Markmann (1975b).

[7] As the authors are not very actively researching in this area of nonparametric density estimation it is possible that some of the issues raised here are already addressed in the literature. Our Google search has not revealed any. If there are any studies that already addressed these issues the authors will be grateful to the reader for bringing that to their attention.

the lower half. If these two numbers are the same then the typical kernel density expression is alright. If the ratio of the number in upper half of the window to the lower half is greater (smaller) than unity then the sample information shows that the density has a positive slope(negative slope). Thus one can have sample estimates of the slope of the density at the midpoints of all the windows. As we know that any curve is the envelope of its tangents we can obtain another independent estimate of the density by estimating these slopes and forming an envelope. Are the two density estimators so arrived at equivalent or do they differ? If they differ, can we take these two estimators and derive a new estimator that is better than the usual kernel density?

There are several other interesting issues on statistical inference with respect to nonparametric density estimation. On a first impression one may think that there are no parameters to estimate in a nonparametric situation, and there are no statistics, no sufficient statistics, and no hypothesis tests of the usual type. But strictly speaking those parametric hypotheses actually refer to parameters of the density function which in turn reflect some aspects of the density such as its mean value, slope or curvature etc. One can postulate in place of a parameter certain hypotheses about the density function, such as its mean, first derivative, its second derivative, etc. One such is the hypothesis that the nonparametric density is equal to a parametric density. Another is that the nonparametric density has the same first and second derivatives at a given point in sample space as a specified parametric density.

In the spirit of Fisher and Rao one may pose questions on the information content of a sample on unknown density function with no parametric form. We can ask the following questions-"What is the information that the sample provides on the unknown slope of the density function at any point? What is a sufficient statistic for the slope of the density?" In order to answer this question can one use the first principles and show that the nonparametric density function, given a statistic for slope, does not depend on the slope? One can make hypotheses regarding the mean of the posterior distribution and whether such mean is dependent on a conditioning explanatory variable or not. Hence hypotheses can be formulated even when we deal with nonparametric densities. These types of tests with nonparametric densities can be performed using the empirical likelihoods (Owen (2001), empirical likelihood ratios, empirical analogues of Rao's score (Gonzalez and Ullah (2001), Ullah (2003)). One may also see Ermakov (1990) for the general issue of testing of hypotheses with nonparametric density functions.

The parametric inference is based on conditional likelihood, conditional on the specified parametric model being true. As most often our specification of a parametric model could be far from the true model that generated the data our parametric inference could have its own limitations on that account. It has been observed that Cramer-Rao bound could be improved. What is the Cramer-Rao limit to variance of an entire parametric density estimator, treating the density as a function of the parameter? What is its equivalent when the density is a nonparametric density? Can one improve Cramer-Rao limit to variance of a density using a nonparametric estimator? We know that under certain conditions a nonparametric density estimator could be closer to the true density than any parametric specification. Le Cam (1964) introduced the concept of approximate sufficiency. Can we use the empirical likelihood and define the concepts of sufficient statistic, complete statistic, and ancillary statistic in the context of a nonparametric specification from the first principles?

Having presented the importance of nonparametric density estimation and some interesting researchable issues we now present some results obtained by the authors in the early seventies (thirty six years ago!) and not published in their entirety. These results were based on an incomplete doctoral thesis of the second author under the guidance of the first author and presented in two parts at the Third World Congress of the Econometric Society in August 1975 and at the Annual Meetings of the Econometric Society in December 1975 (Kumar and Markmann (1975a, 1975b).

5. Comparison of Parametric and Nonparametric Methods of Estimating the Probability Density

In this section the performance of alternate parametric and nonparametric methods of estimating the probability density will be examined using Monte Carlo experiments. The Monte Carlo experiment consists of drawing a random sample of a given sample size from a completely specified distribution belonging to the Pearsonian family of distributions and replicating this procedure several times [here 25 times for all sample sizes except 500 where 10 replications were used].

The Monte Carlo experiments reported in this section were designed, within the constraints of our computer resources in those days (early seventies when we used our own Fortran IV programming on punched cards and used a mainframe CDC computer, the jobs being executed in a batch mode!). Our objective then was to be complementary to results presented in the previously-mentioned paper of Wegman (1972). There he examined the behavior of some nonparametric estimators for certain underlying densities, using as his criterion for comparison the average (over replications) of mean squared error computed at the sample observations (AMSE).

The densities that are common to both Wegman's and our study are the Normal (0, 1), Uniform (0, 1), and Exponential (0, 1). The estimators common to the two studies are the naive estimator and the Kronmal-Tarter estimator since Wegman demonstrated that these outperformed the others used in his study for all the distributions he considered. The inclusion of the Pareto density and exclusion of the Cauchy were based on a requirement we imposed that the maximum likelihood estimates be obtainable in a simple closed form involving no iterative solution of nonlinear equations.

Tables 1-4 present the average mean squared errors (with standard deviations in parentheses) for alternate estimators and for different sample sizes. We summarize our conclusions presented in Kumar and Markmann (1975a, and 1994). Since the four underlying densities considered are members of the Pearson family, we would expect the best-fitting density of the Pearsonian system to be of the same form as the true density or a close approximation to it. By fitting the parametric form of the true density only, rather than all the Pearsonian forms, we save considerable computing effort and obtain results quite close to what can be expected by considering all members of the Pearsonian family and choosing the best-fitting. The tables show that for the four densities considered the parametric methods of density estimation are substantially superior in terms of AMSE to the nonparametric methods considered, and thus imply that the computer algorithm of Gapinski and Kumar (1972) for selecting the best-fitting Pearsonian curve is practical when the true population density is a member of the Pearsonian family or a density approximating a member of that family. The tables also reveal that the nonparametric density estimator given by Specht uniformly outperforms other nonparametric density estimators. The results are summarised in the following figure for Pareto distribution. The figures for other distributions are quite similar qualitatively and can be plotted from Table 1.

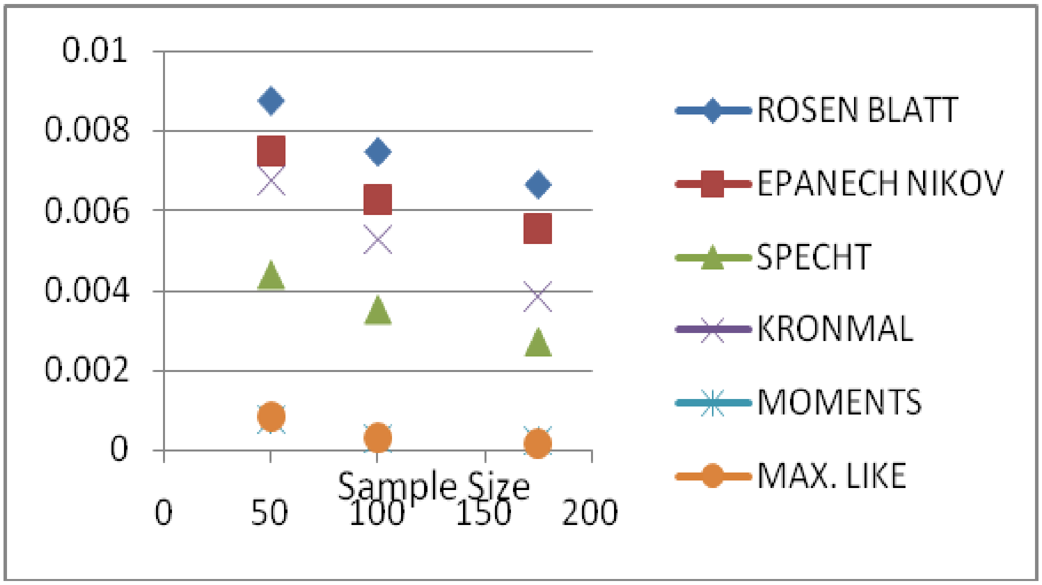

Figure 1. Average Mean Squared Error of Alternate Estimators for the Pareto Distribution

Table 1. Normal (0, 1)

Sample Size	*Rosen Blatt*	*Epanech Nikov*	*Specht*	*Kronmal*	*Moments*	*Max. Like*
50	.001976 (.001460)	.001705 (.001492)	.001783 (.001474)	.001584 (.001335)	.000975 (.000949)	.000975 (.000949)
100	.001132 (.000797)	.001142 (.000777)	.001066 (.000763)	.001559 (.001128)	.000933 (.001304)	.000933 (.001304)
175	.000767 (.000530)	.000715 (.000564)	.000721 (.000537)	.001000 (.001538)	.000338 (.000372)	.000338 (.000372)
250	.000755 (.000314)	.000704 (.000323)	.000737 (.000331)	.000547 (.000552)	.000260 (.000244)	.000260 (.000244)
500	.000424 (.000252)	.000415 (.000275)	.000401 (.000272)	.000403 (.000324)	.000144 (.000121)	.000144 (.000121)

Table 2. Uniform (0, 1)

Sample Size	*Rosen Blatt*	*Epanech Nikov*	*Specht*	*Kronmal*	*Moments*	*Max. Like*
50	.052622 (.015571)	.046437 (.015927)	.043307 (.013018)	.056869 (.010433)	.003062 (.004675)	.002305 (.003329)
100	.047751 (.020628)	.044199 (.022152)	.038094 (.017882)	.052350 (.015525)	.001375 (.001574)	.000528 (.000596)
175	.038241 (.014195)	.036897 (.014850)	.032923 (.013314)	.050259 (.006160)	.001681 (.003127)	.000353 (.000785)
250	.030724 (.009544)	.029088 (.009405)	.026592 (.008613)	.048789 (.004794)	.000543 (.000609)	.000112 (.000145)
500	.018101 (.005034)	.015955 (.005207)	.015493 (.004968)	.047411 (.004300)	.000457 (.000589)	.000055 (.000082)

Table 3. Exponential (0, 1)

Sample Size	*Rosen Blatt*	*Epanech Nikov*	*Specht*	*Kronmal*	*Moments*	*Max. Like*
50	.068024 (.014130)	.056684 (.012221)	.025831 (.011688)	.051413 (.035441)	.015132 (.019471)	.006407 (.010939)
100	.057096 (.008711)	.047013 (.008830)	.022277 (.009044)	.031710 (.012974)	.011074 (.026586)	.001086 (.002352)
175	.052003 (.007040)	.043470 (.006372)	.017646 (.007418)	.026999 (.009583)	.006533 (.008191)	.001052 (.001388)
250	.048891 (.004909)	.040798 (.004790)	.015264 (.004816)	.022916 (.005989)	.004386 (.006411)	.000588 (.000679)

Table 4. Pareto (3.0, 7.5)

Sample Size	*Rosen Blatt*	*Epanech Nikov*	*Specht*	*Kronmal*	*Moments*	*Max. Like*
50	.008733 (.002107)	.007489 (.001982)	.004396 (.001481)	.006767 (.002823)	.000763 (.001243)	.000830 (.001542)
100	.007468 (.001752)	.006267 (.001524)	.003511 (.001618)	.005273 (.001958)	.000294 (.000535)	.000301 (.000491)
175	.006654 (.001219)	.005522 (.001123)	.002693 (.001028)	.003844 (.001220)	.000250 (.000401)	.000166 (.000300)

An auxiliary Monte Carlo experiment was also set up to examine the relative performance of the parametric and nonparametric methods when the underlying true density does not belong to the Pearsonian family. Due to the limited computer budget the experiment was limited to only one or two sample sizes and to only 3 or 5 replications for each sample size. These preliminary investigations demonstrate clearly that the overwhelming superiority of the parametric method observed in Tables 1-4 might not be upheld if the true density does not belong to the Pearsonian System. In Section 7 we demonstrate that this is in fact true in that case.

Another comment regarding these results is also appropriate. Our tables display a generally favourable performance by maximum likelihood estimation as compared to the method of moments. As noted earlier, however, for all the distributions we have included in this study, the maximum likelihood estimators of parameters exist in closed form with no iterative procedure involved. Such, however, is not generally the case, for example, maximum likelihood estimation of most of the Pearsonian densities requires iterative solution of a system of nonlinear equations. It is not clear, then, that the maximum likelihood technique, with its concomitant problems, is superior to the method of moments. An interesting topic for further research is to extend the class of densities to include those for which there is no closed form solution to maximization of the likelihood and use an iterative algorithm for estimation of maximum likelihood estimators.

As mentioned earlier our interest in this topic arose due to our ignorance on small sample densities of econometric estimators. To illustrate the usefulness of this kernel density estimation method we took an econometric example. We compared kernel density estimation with the exact density and the Normal approximation. These results are reported in the next two sections.

6. Analytic Results in a Two-Endogenous-Variable Model

If one were confident that the distribution of the estimator under consideration was a member of the Pearsonian family or closely related thereto, then generating Monte Carlo data and choosing the best-fitting of the Pearsonian types (Gapinski and Kumar (1972)) would yield a tool for statistical inference. However, in those finite-sample situations where exact densities have been derived the densities are seen to be quite complicated, being incomprehensible doubly infinite series in some cases. Thus, there appears no reason to believe or argue that, in a given model, the distribution of an estimator will be Pearson-like.

The aim of this section is to describe a simple model in which the exact distribution of the 2SLS estimator has been derived and in which an analytic approximation, as well, has been advanced. In the next section we utilized the Gapinski-Kumar algorithm for fitting the best type from the family of Pearsonian densities to the 2SLS estimates generated by the Monte Carlo experiments. We also fit the nonparametric density estimators studied above. Results are presented in tabular form and two Figures so that comparisons among the exact, approximate, and estimated densities are facilitated.

For the simple two-equation, linear econometric model where two endogenous variables occur in the equation being estimated, the coefficient of one endogenous variable is specified to be one, and all the predetermined variables are exogenous, the exact distribution of the 2SLS estimator of the coefficient of one endogenous variable has been obtained by Richardson (1968) and Sawa (1969). The exact distribution involves multiple infinite series and is difficult to interpret, but Sawa (1969) has graphed the density for differing parameter combinations on the basis of calculations from a finite number of terms of an infinite series expression.

The main result of the Anderson-Sawa paper (1973) is an asymptotic expansion of the distribution function of the k-class estimator in the case of two included endogenous variables. The density of the approximate distribution yields more insight into the nature of the estimator than the exact density since it (the approximation) is a normal density multiplied by a polynomial. The first correction term to the normal involves a cubic divided by the square root of the sample size while other correction terms involve polynomials of higher degree divided by higher powers of the square root of the sample size. Numerical evaluations for samples of size 20 and 10, as presented in Anderson-Sawa, of the exact normal approximation, and Anderson-Sawa approximation are presented in the first 3 columns of Tables 5 and 7. In the model underlying these results, using the Anderson-Sawa notation, the exact density depends on five parameters: 1) T, the sample size; 2) K_1, the number of exogenous variables included in the relevant equation; 3) K_2, the number of exogenous variables excluded, 4) α, a structural parameter dependent upon the covariance matrix of the reduced form disturbances and the parameter (β) being estimated; 5) Ψ, a non-centrality parameter which measures the effect of the excluded exogenous variables beyond the effect of the included exogenous variables. In tables 5 and 7 the true parameter values chosen by Anderson and Sawa are $K_1 = 1$, $K_2 = 4$, $\alpha = 0.6$, and $\Psi = 4.0$. In Table 5, T = 20, while in Table 7, T = 10. For further details of the econometric model itself the reader is referred to Anderson and Sawa (1973).

7. Estimated Densities in the Two-Endogenous Variable Model

To enable us to determine how well density estimation fares in an econometric setting we use the same model specified above and generate Monte Carlo observations on the 2SLS estimator of the coefficient whose exact and approximate densities are already derived.

In generating the Monte Carlo observations on the 2SLS estimator the following procedure was used for selecting the parameters of the model, the exogenous variables and the disturbance terms. The coefficients of the two endogenous variables and the noncentrality parameter were chosen to assume the same values as in the Anderson-Sawa paper. The parameters of the reduced form equations were selected by trial and error to satisfy the equation $(\pi'_{22}\ A_{22.1}\ \pi_{22})/w_{22} = \Psi$ and $\pi_{21} = \beta\pi_{22}$. The vectors of exogenous variables were chosen to be orthonormal with $A = Z'Z = I$. The disturbance terms of the reduced form equations were assumed to come from independent normal distributions with mean zero and unit standard deviation. For each parameter combination and the sample size 300 T disturbances were generated giving rise to 300 replications of samples of size T. For each of these replications the model was estimated using a 2SLS computer program. For each parameter combination and sample size 300 2SLS estimates of the structural parameter β were obtained. These 300 observations on the 2SLS estimator were utilized to estimate the probability density of the 2SLS estimator using the Pearsonian Curve-Fitting algorithm of Gapinski and Kumar (1972) and the nonparametric methods of estimating the density discussed by us in an earlier paper (Kumar and Markmann (1975)).

Table 5. Anderson-Sawa Table 1 (2SLS)

T=20, K_1=1, K_2=4, α=0.6, Ψ=4.0

a-α	Exact	Normal	Approx.	Type 3	Pearsonian Fits		Type 8
					Type 5	Normal	
-0.5	0	0.002	0	0	0	0.002	0.009
-0.4	0.011	0.028	0.007	0.001	0.003	0.027	0.046
-0.3	0.174	0.227	0.209	0.129	0.136	0.241	0.227
-0.24	0.573	0.561	0.657	0.593	0.587	0.649	0.561
-0.2	1.052	0.945	1.168	1.185	1.165	1.101	0.965
-0.18	1.35	1.18	1.469	1.547	1.524	1.379	1.236
-0.14	2.005	1.72	2.129	2.305	2.286	1.999	1.901
-0.1	2.612	2.282	2.726	2.933	2.931	2.608	2.64
-0.08	2.846	2.536	2.95	3.146	3.15	2.863	2.976
-0.06	3.013	2.739	3.073	3.262	3.279	3.062	3.249
-0.04	3.111	2.921	3.174	3.29	3.312	3.189	3.429
-0.02	3.108	3.025	3.158	3.23	3.254	3.236	3.495
0	3.037	3.061	3.061	3.092	3.115	3.198	3.441
0.02	2.898	3.025	2.892	2.89	2.909	3.078	3.271
0.04	2.702	2.921	2.668	2.641	2.655	2.886	3.007
0.06	2.466	2.739	2.405	2.364	2.372	2.635	2.676
0.08	2.205	2.536	2.122	2.074	2.076	2.344	2.309
0.1	1.935	2.282	1.838	1.786	1.784	2.031	1.937
0.14	1.414	1.72	1.311	1.258	1.249	1.409	1.265
0.18	0.971	1.18	0.89	0.831	0.821	0.88	0.761
0.2	0.788	0.945	0.722	0.66	0.652	0.668	0.577
0.24	0.501	0.561	0.465	0.401	0.395	0.356	0.319
0.3	0.235	0.227	0.225	0.173	0.172	0.114	0.124
0.4	0.057	0.028	0.049	0.034	0.035	0.01	0.025
0.5	0.012	0.002	0.006	0.005	0.006	0	0.005
AMSE		0.0393	0.0052	0.0232	0.0236	0.01	0.0387

In Tables 5 and 7 we present numerical evaluations, at the same points as shown by Anderson and Sawa for the sample sizes 20 and 10 respectively, of those Pearsonian densities which fit the Monte Carlo data. Since generally more than one Pearsonian type will fit a set of data, one will desire to use some criterion for choosing "the" appropriate curve. Here we know the exact density and can again use a criterion such as Mean Squared Error as shown in the final row of the tables. Figure 2 below presents the Anderson-Sawa approximation and the best fitting parametric Pearsonian density for sample size T=20. Similar picture is seen with sample size 10. The AMSE associated with Anderson-Sawa approximation is 0.005 whereas the AMSE associated with the Pearsonian density is 0.01. The picture clearly shows that around the mean the densities are quite close. Only at the right tail the Pearsonian density is poor, suggesting nonparametric density could be better.

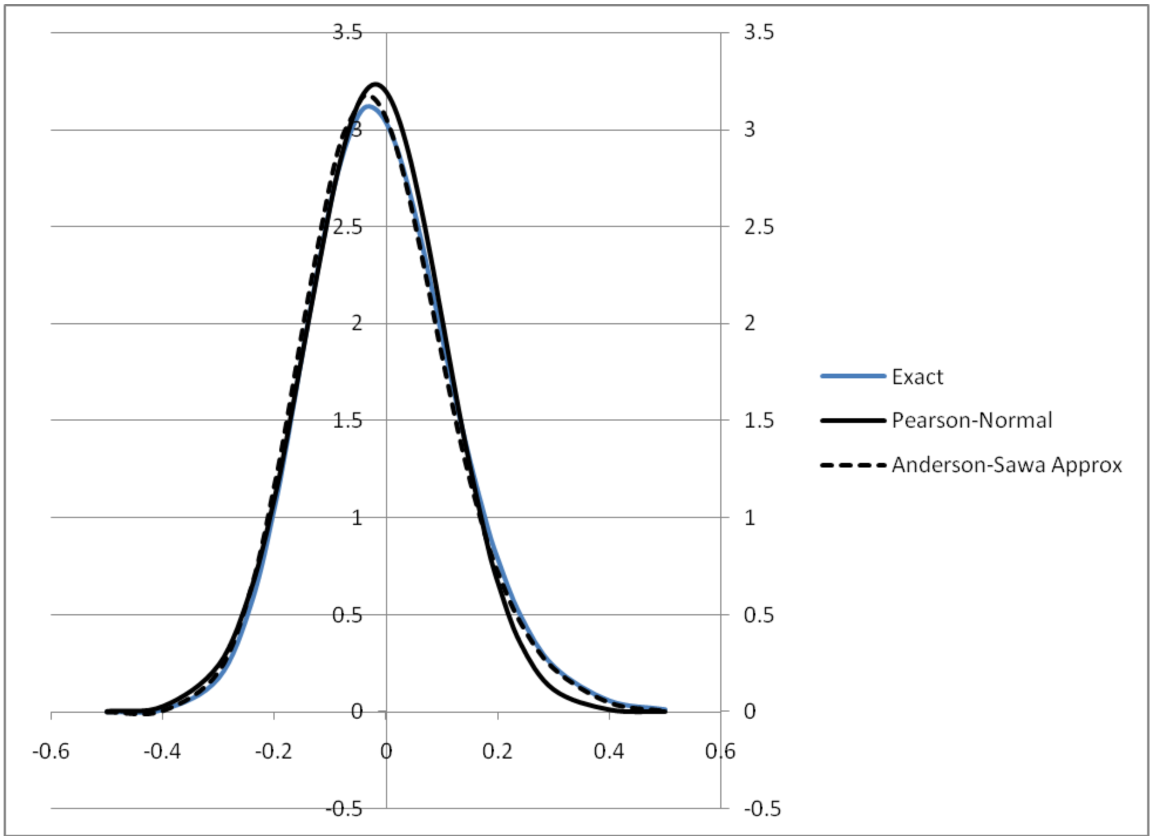

Figure 2. Comparison of Anderson-Sawa approximation and Best-fitting Pearsonian Density

In Tables 6 and 8, we present results for the nonparametric density estimators considered here, again using the same Monte Carlo generated data on the above two-equation model for sample size 20 and 10 respectively. For these estimators we have simulated a realistic situation in assuming that the exact density is unknown and hence so is the optimal bandwidth. However, one must make a decision as to what bandwidths will be used. Our "arbitrary" procedure was to take that bandwidth which would be used if the underlying density were the normal (with the true mean and variance estimated by the sample mean and variance). However, in Table 6 we also show the experimentally-determined optimal behavior of the nonparametric estimators. That is, the second group of results in that reflect usage of the bandwidth which minimizes AMSE. The goodness of fit of the nonparametric density and the Anderson and Sawa approximation are shown in Figure 3 below:

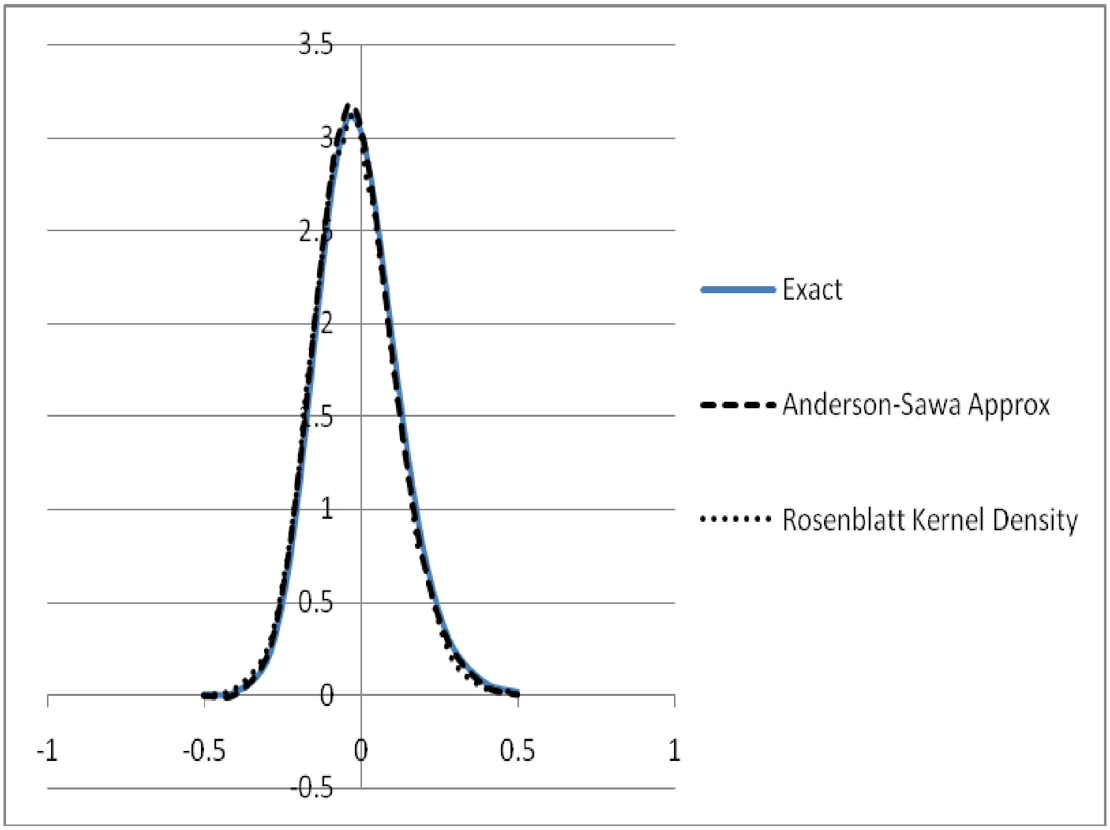

Figure 3. Comparison of Nonparametric Kernel Density and Anderson and Sawa Approximation

From Figures 3 it is clear that the poor fit at the right tail observed in Figure 2 is overcome by fitting a nonparametric kernel density. Thus using nonparametric kernel density is a much better option than using the parametric Pearsonian density. It is a good alternative to Edgeworth expansion of the density.

Summary of Findings

The results set forth in Tables 5 through 8 appear to imply the following, though they constitute perhaps too small a set of results on which to base any firm conclusions. In terms of ASE the Anderson-Sawa approximation provides a significantly more powerful tool for approximating the exact density [and thus for inference] than the normal approximation. None of the parametric or nonparametric estimators considered here performs as well as the Anderson-Sawa approximation in either case [T=10 or T=20]. Relative to the Anderson-Sawa approximation, some of the estimated densities appear to perform rather well. In particular, the Rosenblatt and Epanechnikov estimates outperform the best-fitting parametric for the case of T=10, and perform almost as well for T=20 [in fact, better using the proper bandwidth]. This result might appear surprising in light of the results displayed in Tables 1 through 4, but, as pointed out above, there is no basis for assuming that the exact density of 2SLS estimator in the cases considered here [particularly for T=10] closely resembles a Pearson-type density. The Specht estimator exhibits poor performance with the pragmatically chosen bandwidth, but is competitive with Rosenblatt and Epanechnikov when the bandwidths are optimal ones. In general, the Kronmal-Tarter estimator displays poor behavior and is not itself a true density here since it yields negative values in some instances.

Table 6. Anderson-Sawa Table 1 (2SLS)

T=20, K_1=1, K_2=4, α=0.6, Ψ=4.0

	Exact	*Rosenblatt*	*Epanechnikov*	*Specht*	*Kronmal-Tarter*	*Rosenblatt*	*Epanechnikov*	*Specht*
a-α		*Using "Arbitrary" Bandwidths*				*Using "optimal" Bandwidths*		
-0.5	0	0	0	0	0.041	0	0	0.002
-0.4	0.011	0.046	0.033	0	0.045	0.037	0.051	0.049
-0.3	0.174	0.207	0.212	0.211	0.079	0.241	0.296	0.292
-0.24	0.573	0.689	0.616	0.31	0.558	0.685	0.747	0.727
-0.2	1.052	1.148	1.125	1.296	1.142	1.167	1.236	1.219
-0.18	1.35	1.4	1.441	1.528	1.491	1.556	1.521	1.512
-0.14	2.005	2.272	2.205	2.208	2.224	2.111	2.111	2.125
-0.1	2.612	2.754	2.804	3.027	2.867	2.704	2.638	2.655
-0.08	2.846	3.03	3.061	3.217	3.105	2.889	2.808	2.844
-0.06	3.013	3.121	3.203	3.295	3.268	2.944	2.939	2.968
-0.04	3.111	3.098	3.186	3.406	3.344	3.093	3.021	3.02
-0.02	3.108	3.121	3.178	3.311	3.331	3.111	3.014	3.002
0	3.037	3.098	3.106	3.17	3.231	3	2.903	2.918
0.02	2.898	3.007	2.959	2.609	3.053	2.759	2.762	2.776
0.04	2.702	2.548	2.667	3.006	2.809	2.63	2.567	2.585
0.06	2.466	2.479	2.397	2.885	2.516	2.352	2.348	2.355
0.08	2.205	2.02	2.102	2.126	2.193	2.111	2.111	2.101
0.1	1.935	1.951	1.849	1.3	1.86	1.796	1.861	1.841
0.14	1.414	1.262	1.255	1.488	1.231	1.333	1.371	1.349
0.18	0.971	0.895	0.904	0.822	0.734	0.815	0.908	0.932
0.2	0.788	0.734	0.726	0.733	0.551	0.722	0.733	0.753
0.24	0.501	0.39	0.373	0.304	0.309	0.444	0.479	0.457
0.3	0.235	0.138	0.14	0.2	0.16	0.167	0.195	0.193
0.4	0.057	0.023	0.031	0.014	0.051	0.037	0.054	0.048
0.5	0.012	0.023	0.029	0.087	-0.022	0.019	0.02	0.025
AMSE		0.0117	0.0112	0.0635	0.0264	0.0084	0.0097	0.0087

Table 7. Anderson-Sawa Table 2 (2SLS)

T=10, K_1=1, K_2=4, α=0.6, Ψ=4.0

a-α	*Exact*	*Normal*	*Approx.*	*Pearsonian Fits*		*Type 8*
				Type 3	*Normal*	
-0.5	0.027	0.055	0.028	0.01	0.037	0.049
-0.4	0.169	0.206	0.217	0.128	0.189	0.182
-0.3	0.644	0.576	0.755	0.659	0.655	0.587
-0.24	1.131	0.927	1.252	1.246	1.137	1.053
-0.2	1.494	1.201	1.606	1.685	1.516	1.455
-0.18	1.67	1.344	1.776	1.896	1.709	1.671
-0.14	1.979	1.622	2.065	2.254	2.069	2.097
-0.1	2.192	1.867	2.255	2.483	2.351	2.448
-0.08	2.252	1.969	2.303	2.537	2.446	2.571
-0.06	2.28	2.052	2.317	2.548	2.505	2.647
-0.04	2.275	2.113	2.253	2.519	2.525	2.673
-0.02	2.239	2.151	2.246	2.451	2.504	2.645
0	2.175	2.163	2.163	2.349	2.444	2.567
0.02	2.087	2.151	2.056	2.219	2.347	2.443
0.04	1.979	2.113	1.928	2.068	2.218	2.28
0.06	1.856	2.052	1.787	1.9	2.064	2.09
0.08	1.722	1.969	1.635	1.724	1.889	1.88
0.1	1.583	1.867	1.479	1.545	1.702	1.663
0.14	1.302	1.62	1.179	1.196	1.316	1.24
0.18	1.037	1.344	0.912	0.885	0.955	0.872
0.2	0.916	1.201	0.796	0.749	0.794	0.717
0.24	0.701	0.927	0.602	0.521	0.523	0.469
0.3	0.452	0.576	0.397	0.281	0.248	0.232
0.4	0.201	0.206	0.195	0.084	0.052	0.063
0.5	0.085	0.055	0.083	0.021	0.007	0.016
AMSE		0.0475	0.0062	0.0289	0.0265	0.0522

Table 8. Anderson-Sawa Table 2 (2SLS)

$T=10$, $K_1=1$, $K_2=4$, $\alpha=0.6$, $\Psi=4.0$

a-α	*Exact*	*Rosenblatt*	*Epanechnikov*	*Specht*	*Kronmal-Tarter*
-0.5	0.027	0.036	0.031	0.009	-0.089
-0.4	0.169	0.161	0.146	0.112	0.098
-0.3	0.644	0.734	0.726	0.629	0.758
-0.24	1.131	1.182	1.248	1.376	1.315
-0.2	1.494	1.629	1.627	1.67	1.69
-0.18	1.67	1.791	1.816	1.682	1.865
-0.14	1.979	2.185	2.171	2.132	2.164
-0.1	2.192	2.31	2.41	2.64	2.37
-0.08	2.252	2.346	2.469	2.77	2.431
-0.06	2.28	2.453	2.484	2.766	2.46
-0.04	2.275	2.346	2.418	2.559	2.458
-0.02	2.239	2.346	2.319	2.24	2.425
0	2.175	2.292	2.216	2.029	2.361
0.02	2.087	2.077	2.072	1.956	2.269
0.04	1.979	1.97	1.959	1.932	2.151
0.06	1.856	1.826	1.859	1.929	2.011
0.08	1.722	1.737	1.768	1.832	1.853
0.1	1.583	1.612	1.678	1.688	1.582
0.14	1.302	1.307	1.354	1.507	1.319
0.18	1.037	0.985	0.991	0.948	0.858
0.2	0.916	0.895	0.832	0.685	0.79
0.24	0.701	0.501	0.497	0.465	0.492
0.3	0.452	0.233	0.221	0.175	0.176
0.4	0.201	0.072	0.08	0.083	-0.001
0.5	0.085	0.054	0.047	0.007	0.056
AMSE		0.0112	0.0159	0.0492	0.0269

8. Concluding Remarks

We pointed out that there exists little knowledge on the sampling distribution of certain econometric estimators (particularly in small-sample situations). Any knowledge gained on such distributions will aid econometricians in the two fundamental inferential problems of choosing the best estimator under a given choice criterion, and in developing tests of significance. Such knowledge can be gained from the capital-intensive approach of generating estimates from Monte Carlo experiments and estimating the unknown density function. However, numerous alternatives for estimating the unknown density exist, and only through studying the behavior of the alternatives can a decision be made as to which technique to choose. We consider this paper to be only a step in that direction, and, of course, have made no firm recommendations. Further, we do not imply that estimating density functions will ever be a substitute for theoretical derivations of exact densities or even good theoretical approximations. But in those situations where neither of these exist, and statistical inference must be carried out, the technique appears to offer a viable alternative (Kumar and Gapinski (1974).

It is necessary to examine in greater detail the connection between Edgeworth approximation, kernel estimation of densities and Efron's bootstrap approach (Efron (1982)). Can one method enrich the other? What are the interconnections between them? For example, Efron's bootstrap test will give better accuracy than the Edgeworth's approximation if the asymptotic distribution of the test statistic does not depend on the parameters of the model, i.e., the test statistic is pivotal (see Beran (1988)). If the test statistic is not pivotal, is inference based on a kernel estimator better than the one based on the bootstrap method?

References

Adamowsky, K (1989), “A Monte Carlo Comparison of Parametric and Nonparametric Estimation of Flood Frequencies”, *Journal of Hydrology*, Vol. 108, pp.295-308.

Anderson, T.W. and Sawa, T. (1973), "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions", *Econometrica*, Vol.41, No.4, July, pp.683-714.

Bierens, H.J. (1987), “Kernel Estimators of Regression Functions”, in T.F. Bewley, *Advances in Econometrics, Fifth World Congress,* Vol. 1, Cambridge University Press.

Deaton, A.S. (1991), “Rice-prices and income distribution in Thailand: A Nonparametric analysis”, *Economic Journal,* 99, pp.1-37.

Deaton, A.S. (1993), “Data and econometric tools for development economics”, in J.R. Berman and T.N. Srinivasan, *Handbook of Development Economics*, Vol. 3.

Edgeworth, F.Y. (1904), "The Law of Error", *Transactions of Cambridge Philosophical Society*, Vol.20, pp.36-113.

Elderton, W.P. and Johnson, N.L., (1969), *Systems of Frequency Curves*, Cambridge University Press, London.

Epanechnikov, V.A, (1969), "Nonparametric Estimates of a Multivariate Probability Density", *Theoretical Probability and Application*, pp. 153-58.

Ermakov, M.S. (1990), “Asymptotically Minimax Tests for Nonparametric Hypotheses Concerning the Distribution Density”, *Journal of Mathematical Sciences*, Vol. 52, No. 2, pp.1072-3374.

Gapinski, J.H. and Kumar, T.K., (1972), "A Pearsonian Curve-Fitting Algorithm", *Econometrica*, 1972, p.963.

Gonzalez, Rivera-Gloria and Aman Ullah (2001), “Rao's score test with nonparametric density estimators”, *Journal of Statistical Planning and Inference*”, Vol. 97, No.1, pp.85-100.

Guo, S.L. (1991), “Nonparametric Variable Kernel Estimation with Historical Floods and Palaeoflood Information”, *Water Resource Research*, Vol.27, No.1, pp.91-98.

Hardle, W., and O. Linton (1994), "Applied Nonparametric Methods", Chapter 38 in R.F. Engel and D.F. MacFadden (Editors), *Handbook of Econometrics,* Elsevier Science B.V. Vol. 4, pp. 2295-2339.

Jiang, George J., and John L. Knight (1997), "A Nonparametric Approach to the Estimation of Diffusion Processes, *Econometric Theory,* Vol. 13, No. 5, pp.615-645.

Johnson, N.L., (1949), "Systems of Frequency Curves Generated by Methods of Translation", *Biometrika*, pp. 149ff.

Johnson, N.L., Kotz, and N. Balakrishnan (1994), *Continuous Univariate Distributions*, *Volume I, 2nd Ed., John Wiley and Sons*

Kronmal, R. and Tarter, M., (1968), "The Estimation of Probability Densities and Cumulatives by Fourier Series Methods", *Journal of the American Statistical Association*, pp.925-52.

Kumar, T.K. and Gapinski, J.H., (1974), "Nonlinear Estimation of the CES Production Function: Sampling Distributions and Tests in Small Samples", *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 1973, and in *Southern Economic Journal*, Vol.41, No.2, October, pp.258-66.

Kumar, T.K. and Markmann, J.M.,(1975a): "Estimation of Probability Density Functions: A Monte Carlo Comparison of Parametric and Nonparametric Methods", Paper Presented at the Third World Congress of the Econometric Society, Toronto, Canada, August, 1975. (Cited by Rao, BLSP (1983) and Jiang and Knight (1997))

Kumar, T.K. and Markmann, J.M., (1975b), "Nonparametric Density Estimation with An Econometric Application", Paper Presented at the Annual Meetings of the Econometric Society, 1975.

Kumar,T.K. and Markmann, J.M. (1994), "Estimation of Probability Density Functions With an Econometric Application", (with J M Markmann), *Indian Journal of Applied Economics*, Vol.3, Nos. 3 & 4, April-September 1994, pp.220-241.

Le. Cam, L. (1964), "Sufficiency and Approximate Sufficiency", *Annals of Mathematical Statistics*, Vol. 35, No 4, pp. 1419-1455 (Section 7)

Mackinnon, James G. (1993), "Critical Values for the Cointegration Tests". In Robert F. Engle, Clive William John Granger, *Long-run economic relationships: readings in cointegration,* Cambridge University Press, 1991, pp.267-274.

Owen, A.B. (2001), *Empirical Likelihood*, Chapman and Hall

Pagan, Adrian, and Aman Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press

Parzen, E. (1962), "On estimation of probability density function and mode", *Annals of Mathematical Statistics*, Vol. 33, pp 1065-1076.

Prakasa Rao, B.L.S. (1983), *Nonparametric Functional Estimation*, Academic Press. London.

Richardson, David H., (1968): "The Exact Distribution of a Structural Coefficient", *Journal of the American Statistical Association*, pp.1214-26.

Rosenblatt, Murray, (1956), "Remarks on Some Nonparametric Estimates of a Density Function", *Annals of Mathematical Statistics*, pp.832-35.

Rosenblatt, Murray, (1971), "Curve Estimates", *Annals of Mathematical Statistics*, pp. 1815-42.

Sawa, Takamitsu, (1969), "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators", *Journal of the American Statistical Association*, pp.923-37.

Sen, B. (2005), “Estimation and Comparison of Fractile Graphs Using Kernel Smoothing Techniques”, *Sankhya*, 67, pp. 305-334.

Sengupta, J.K. and T.K. Kumar (1965), “An Application of Sensitivity Analysis to a Linear Programming Problem", (with J.K.Sengupta), *Unternehmensforschung*, Vol.9, No.1, 1965, pp. 18-36.

Silverman, B.W.(1992), *Density Estimation for Statistics and Data Analysis*, London, Chapman and Hall,

Specht, Donald F., (1971), "Series Estimation of a Probability Density Function", *Technometrics*, 1971, pp. 409-24.

Specht, Donald. F. (1990) "Probabilistic neural networks," *Neural Networks*, Vol. 3, No.1, pp. 110-118.

Specht Donald F. (1991). “A General Regression Neural Network”, *IEEE Transactions on Neural Networks*, Vol. 2, No.6, 568–576.

Terrel, George R., and David W. Scott (1992), “Variable Kernel Density Estimation”, *Annals of Statistics*, Vol. 20, No.3, pp.1236-1265.

Ullah, Aman (2003), “Nonparametric Kernel Method of Estimation and Hypothesis Testing” In Baltagi, Badi H. (Ed), *A Companion to Theoretical Econometrics,* Blackwell Publishers, Chapter 20.

Wegman, E.J., (1972a), "Nonparametric Probability Density Estimation I: A Summary of Available Methods", *Technometrics*, pp. 523-45.

Wegman, E.J. (1972b), "Nonparametric Probability Density Estimation II: A Comparison of Density Estimation Methods", *Journal of Statistical Computation and Simulation*, pp. 225-45.

♦♦♦