RAIRO. RECHERCHE OPÉRATIONNELLE

THUAN V. TRUONG

The Kolmogorov-Smirnov distance as criterion of choice of estimators with application to the first order auto-regressive case : a Monte Carlo study

RAIRO. Recherche opérationnelle, tome 18, n° 3 (1984), p. 297-307

<http://www.numdam.org/item?id=RO_1984__18_3_297_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 18, n° 3, août 1984, p. 297 à 307)

THE KOLMOGOROV-SMIRNOV DISTANCE AS CRITERION OF CHOICE OF ESTIMATORS WITH APPLICATION TO THE FIRST ORDER AUTO-REGRESSIVE CASE: A MONTE CARLO STUDY (*)

by THUAN V. TRUONG (¹)

Abstract. — *For any stochastic model, the general interest is not only to produce the "best" possible estimator of the regression parameter but also to use it to make inference. The finite sample distribution of the estimator on which the inference can be based, is usually unknown, but its asymptotic distribution is known. It has been a common practice in Econometrics to use an estimated asymptotic distribution to make inference in small sample. Obviously, this will give misleading inference unless the estimated asymptotic distribution can well approximate the exact finite sample distribution.*

Given the present practice of making inference in the absence of our knowledge of the finite sample distribution of an estimator, if we have to choose an estimator from a set of estimators, it seems one should choose the one for which the estimated asymptotic distribution is closest to its exact distribution. This can be contrasted with the well known criterion of choice such as the mean squared error (MSE). Though it may be possible, it does not seem likely that these two criteria will produce the same choice.

We examine the five most used estimators of the regression coefficients of the regression model with the error following the first-order autoregressive process, namely Ordinary Least Squares (OLS), Cochrane-Orcutt (CO), Cochrane-Orcutt modified by Prais and Winston (PW), Durbin and Maximum Likelihood (ML) estimators. Adopting the well known measure of distance between two distributions by Kolmogorov and Smirnov as a measure of closeness of the two distributions, we compute the distance between the asymptotic distribution and its small sample estimate for each of these estimators. Due to analytical complexity, we resort to Monte Carlo study. General conclusion that can be drawn from our study is that OLS should never be preferred, and even though all other estimators are comparable over the entire range of the autocorrelation parameter, PW seems to be preferable. This can be contrasted to the well known conclusion using MSE as the criterion that OLS may be preferred when the order of the autocorrelation coefficient does not exceed .30 and for larger values of this coefficient, all other estimaros are comparable, but possibly the ML estimator may be prefered.

Keywords: Asymptotic distribution; distance; first order autoregressive model; Monte Carlo technique.

Résumé. — *Pour un modèle économétrique donné, quand plusieurs méthodes d'estimation existent, on se trouve devant un problème de choix, donc de définition du meilleur estimateur. Le critère de choix le plus souvent employé est celui de l'erreur quadratique moyenne. Cette étude propose l'utilisation de la distance de Kolmogorov-Smirnov et compare au moyen de la méthode de simulation Monte Carlo, les estimateurs les plus utilisés des coefficients de regression du modèle auto-régressif du premier ordre.*

Mots clés : Distribution asymptotique; distance; modèle autoregressif du premier ordre; technique Monte Carlo.

(*) Received July 1982.
I would like to thank an anonymous referee for very helpful comments. As usual, the remaining mistakes, if any, are entirely mine.
(¹) University of Baltimore, 1420 North Charles Street, Baltimore, MD 21201, U.S.A.

1. INTRODUCTION

For any stochastic model, the usual concern is not only to produce the "best" possible estimator of an unknown parameter but also to use the estimators to make inferences about this parameter. For most of the econometric models that are in use, the finite sample distribution(s) of the estimator(s) is (are) not known. Fortunately, however, the asymptotic distributions of these estimators are known. In all these cases it has been a common practice to use a finite sample estimate of this asymptotic distribution for inference purposes.

The implication of using the estimated asymptotic distribution in place of the actual distribution of an estimate may be shown by the following example.

Assume that the estimated asymptotic distribution of the estimate $\hat{\beta}$ of β is a t-distribution. Let ESE be the estimated standard error and STA be the following ratio:

$$\text{STA} = (\text{Estimate} - \text{True Value})/\text{ESE}.$$

According to the common practice for a small sample size, a confidence interval at 100 $(1-\varepsilon)$ percent level is given by:

$$\hat{\beta} \pm t_{\varepsilon/2}\,\text{ESE}. \tag{1}$$

If the exact distribution of STAT corresponding to $\hat{\beta}$ were known, then the "exact" confidence interval should be: $\hat{\beta} \pm d_{\varepsilon/2}$ ESE, where $d_{\varepsilon/2}$ is computed from the actual distribution which is assumed to be symmetric for illustration purposes. Thus the common practice approximates the coefficient $d_{\varepsilon/2}$ by $t_{\varepsilon/2}$ that is computed from the estimated asymptotic distribution of $\hat{\beta}$. In this case the $100(1-\varepsilon)\%$ cofidence interval as given by (1) is misleading because such intervals will not include the true value $100\ (1-\varepsilon)\%$ of the time unless $d_{\varepsilon/2} = t_{\varepsilon/2}$.

In general the confidence coefficient corresponding to (1) should be different from the stated coefficient of $100\ (1-\varepsilon)\%$. The magnitude of this error depends on the closeness of $d_{\varepsilon/2}$ to $t_{\varepsilon/2}$.

Given the present practice of making inference in Econometrics, we like to argue that given a choice between two estimators for which only asymptotic distributions are known, one should choose the estimator whose exact distribution can be most closely approximated by its estimated asymptotic distribution. To make the idea concrete, suppose that we have two estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ of β with the same asymptotic normal distribution. From each of these estimates, the common practice confidence intervals of β are given by

$\hat{\beta}_1 \pm t_{\varepsilon/2}\,\mathrm{ESE}_1$ and $\hat{\beta}_2 \pm t_{\varepsilon/2}\,\mathrm{ESE}_2$ with the same $t_{\varepsilon/2}$ in both cases since the asymptotic distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ are the same. Let the correct confidence intervals be $\hat{\beta}_1 \pm d_{1\,\varepsilon/2}\,\mathrm{ESE}_1$ and $\hat{\beta}_2 \pm d_{2\,\varepsilon/2}\,\mathrm{ESE}_2$. Therefore the closer $d_{i\,\varepsilon/2}$ $(i=1, 2)$ is to $t_{\varepsilon/2}$ the less misleading the inference, and by implication the better the estimator. In other words, loosely speaking, given the present practice of making inference in Econometrics, the least misleading inference is provided by the estimator whose exact distribution is "closest" to the estimated asymptotic distribution. This can be contrasted with the well known mean squared error (MSE) criterion where an estimate is judged by its closeness to the true parameter value. The MSE criterion may not satisfy the requirement that the chosen estimator will provide better approximation to a true inference and thus it may not be desirable to consider it as a good criterion in the choice of inference-oriented estimators.

In conclusion the common practice of using MSE to select an estimator may fail to provide a criterion for choosing estimators to be used for inference. Such a criterion should reflect some idea of the "closeness" of two distributions.

2. THE CONCEPT OF DISTANCE

The "closeness" of two distributions can be defined in terms of a measure of distance between them. To be useful, such a distance d between two probability distributions has to be defined not only on the real line, but on any abstract measure spaces. It has to be a metric and has to satisfy some fundamental statistical restrictions.

Lets (X, S) be a probability space. If a distance $d(\mu, \nu)$ between two measure μ and ν defined on (X, S) is a metric, then $d(\mu, \nu)$ satisfies the following properties:

(i) $d(\mu, \nu) \geqq 0$;

(ii) $d(\mu, \nu)=0$ if and only if $\mu=\nu$;

(iii) $d(\mu, \nu)=d(\nu, \mu)$;

(iv) $d(\mu, \nu) \leqq d(\mu, \xi)+d(\xi, \nu)$ for any probability measure defined on the same space (X, S). This relation is often referred to as a triangle inequality. Obviously an ideal distance should be zero when $\mu=\nu$ and should be maximum when μ and ν are most apart, in symbols $\mu \perp \nu$. Also, although it is not a major restriction, d should take only finite values.

There are a number of distances that satisfy the above mentioned properties (Adhikari, 1956 and Ali, 1966). But they are not often mathematically tractable. Thus, in the subsequent discussion the only distance we propose as a measure of closeness between two distributions is the Kolmogorov-Smirnov distance. Its definition is:

$$d(\mu, \nu) = \operatorname*{Sup}_{E \in S} |\mu E - \nu E|.$$

3. THE KOLMOGOROV-SMIRNOV DISTANCE AS CRITERION OF CHOICE OF ESTIMATORS OF THE REGRESSION PARAMETERS OF THE FIRST ORDER AUTOREGRESSIVE MODEL

The objective of this section is to provide a comparison of various estimations of the regression parameters, not in terms of the MSE of their distributions, but in terms of the Kolmogorov-Smirnov distance between the t-distribution (with appropriate degree of freedom) and the actual distribution of what has been defined as STAT.

Let's introduce the model. The first order autoregressive case may be written as:

$$Y_t = \beta_1 + \beta_2 x_t + u_t \quad \text{with } u_t = \rho u_{t-1} + v_t, \qquad t = 1, \ldots, T,$$

Since an analytical treatment seems to be intractable, the Monte Carlo technique will be used. And for the purpose of this technique, further assumptions are necessary. They are made in such a way that our results can be compared with those of two previous studies by Rao and Griliches (1969) and Beach and McKinnon (1978). Thus, from here on, we assume that:

(i) the v_t's follow a normal distribution with zero mean and variance .0036;

(ii) the two regression parameters β_1 and β_2 are both set equal to 1;

(iii) we experiment with 20 different values of ρ in the open interval $(-1,1)$: $-.90, -.80, \ldots, -.10, 0, .10, \ldots, .90, .99$;

(iv) two sample sizes 20 and 50 are chosen. For each sample size, there is one and only one set of values of x, each of which is considered as an independent observation, and is generated from an exponential trend plus stochastic error,

$$x_t = \exp(.04\, t) + w_t, \qquad t = 1, \ldots, T,$$

where the disturbance term w_t is identically, independently, and normally distributed with zero mean and variance .0009.

In each experiment we choose a value of ρ, a sample size and a sample of x, and generate 1,000 samples of Y. For each sample of Y and given the values of x, the five most known techniques of estimation of the regression coefficients of the model, namelky Ordinary Least Squares (OLS), Cochrane-Orcutt (CO), Crochane-Orcutt modified by Prais and Winston (PW), Durbin and Maximum-Likelihood (ML) methods, are used to get the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ of the regression parameters and their estimated standard errors ESE_1, ESE_2. And for each of these methods the statistic.

$$STAT_i = (\hat{\beta}_i - \beta_i)/ESE_i, \qquad i = 1, 2,$$

is then computed. Finally for each estimator the distance $DIST_i$ $(i=1, 2)$ of the actual distribution of $STAT_i$ $(i=1, 2)$ from the corresponding t-distribution is computed. (Since the computations of the estimators, and their standard errors, ESE_i are too well known procedures, we won't recall them here. But references are given for the interested reader.)

The results of this Monte Carlo study are recorded in two tables.

In one hand, Table I gives the values of $DIST_i$ $(i=1, 2)$ for each value of ρ, each sample size T and for each of the five methods of estimation. In the other hand, computation techniques for three out of the five estimators considered are iterative. In all of our computations, we require a five-digit accuracy for $\hat{\rho}$. With this degree of accuracy, the number of iterations depend not only upon the true value of ρ, but also upon the sample size. For each sample size, we have 20 experiments each corresponding to a specific true value of ρ. For each experiment we have one thousand samples. For each sample we record the number of iterations needed by each of the iterative technique to compute the relevant estimates. For each experiment we have recorded the minimum, maximum and average number of iterations. These are recorded in Table II for twenty values of ρ and two sample sizes $T=20$ and 50.

The most important results are summarized below.

TABLE I

Distance of the distribution of stat from the corresponding T-distribution

ρ	Distance	OLS	CO	PW	Durbin	ML	T
−.90	$DIST_1$	.221,56	.023,58	.021,38	.021,69	.023,30	20
	$DIST_2$	.220,14	.275,80	.013,41	.023,56	.030,14	
	$DIST_1$	.266,09	.023,79	.019,20	.021,46	.024,30	50
	$DIST_2$	.256,09	.015,75	.012,20	.014,20	.011,26	
−.80	$DIST_1$	.183,77	.036,85	.019,90	.040,53	.040,91	20
	$DIST_2$	.183,40	.046,30	.014,45	.044,95	.049,13	
	$DIST_1$	.219,07	.022,60	.020,34	.020,07	.021,32	50
	$DIST_2$	.212,83	.016,51	.014,32	.013,87	.014,20	

TABLE I (continued)

ρ	Distance	OLS	CO	PW	Durbin	ML	T
−.70	$DIST_1$ $DIST_2$	.159,64 .161,46	.017,81 .018,84	.019,25 .022,65	.019,31 .020,74	.016,77 .020,77	20
	$DIST_1$ $DIST_2$	.191,75 .191,85	.020,85 .024,33	.019,50 .021,84	.018,54 .022,20	.020,30 .022,36	50
−.60	$DIST_1$ $DIST_2$	.150,82 .148,36	.035,23 .035,64	.034,78 .025,43	.034,78 .038,77	.031,68 .033,85	20
	$DIST_1$ $DIST_2$	.162,44 .160,28	.032,22 .025,28	.031,02 .026,70	.036,94 .024,91	.029,65 .025,33	50
−.50	$DIST_1$ $DIST_2$	.107,64 .113,24	.053,63 .041,74	.019,85 .026,34	.050,69 .041,92	.039,82 .036,00	20
	$DIST_1$ $DIST_2$	.133,11 .132,68	.018,28 .023,99	.019,36 .017,98	.020,20 .025,72	.019,41 .018,09	50
−.40	$DIST_1$ $DIST_2$	.090,02 .091,92	.026,96 .026,74	.050,89 .038,33	.026,85 .034,10	.031,12 .027,50	20
	$DIST_1$ $DIST_2$	.130,29 .128,24	.050,01 .041,05	.052,11 .035,35	.049,60 .039,27	.051,71 .035,65	50
−.30	$DIST_1$ $DIST_2$	.087,03 .092,98	.045,29 .041,70	.019,78 .022,06	.044,72 .042,32	.060,51 .050,37	20
	$DIST_1$ $DIST_2$	.098,16 .098,46	.017,17 .020,80	.019,76 .022,63	.021,65 .022,97	.020,46 .023,22	50
−.20	$DIST_1$ $DIST_2$	.063,73 .062,77	.020,62 .023,07	.039,67 .028,13	.019,65 .025,59	.022,53 .026,42	20
	$DIST_1$ $DIST_2$	.058,73 .064,53	.038,96 .028,08	.035,55 .025,09	.041,92 .029,27	.035,20 .025,30	50
−.10	$DIST_1$ $DIST_2$	.047,56 .045,91	.037,50 .037,61	.019,03 .025,89	.039,71 .040,45	.050,70 .048,24	20
	$DIST_1$ $DIST_2$	.033,84 .038,43	.019,40 .025,34	.017,69 .022,85	.017,55 .024,98	.017,59 .022,71	50
.00	$DIST_1$ $DIST_2$	.036,67 .025,01	.028,67 .032,84	.031,27 .030,22	.027,11 .033,25	.026,77 .031,34	20
	$DIST_1$ $DIST_2$	.023,32 .024,54	.032,07 .029,05	.033,70 .031,66	.029,74 .029,83	.033,73 .031,43	50
.10	$DIST_1$ $DIST_2$	.040,62 .045,41	.044,32 .041,68	.040,91 .037,05	.042,50 .041,47	.045,84 .052,63	20
	$DIST_1$ $DIST_2$	.048,21 .053,44	.040,74 .036,63	.035,42 .040,93	.038,92 .035,81	.036,03 .041,85	50

TABLE I (continued)

ρ	Distance	OLS	CO	PW	Durbin	ML	T
.20	$DIST_1$ $DIST_2$	.055,39 .060,47	.059,95 .064,40	.027,69 .029,80	.060,00 .064,97	.064,01 .069,11	20
	$DIST_1$ $DIST_2$	.061,80 .062,31	.027,26 .029,63	.032,44 .031,01	.027,04 .029,02	.033,58 .031,20	50
.30	$DIST_1$ $DIST_2$	.068,57 .067,82	.044,88 .044,94	.036,95 .038,06	.045,86 .046,15	.046,27 .046,87	20
	$DIST_1$ $DIST_2$	.093,49 .085,51	.035,98 .037,33	.036,99 .034,43	.032,31 .033,73	.036,39 .034,50	50
.40	$DIST_1$ $DIST_2$	.103,34 .096,34	.078,17 .068,39	.043,19 .046,30	.079,14 .069,29	.064,23 .053,76	20
	$DIST_1$ $DIST_2$	.117,04 .111,49	.042,57 .045,76	.037,51 .039,67	.041,76 .042,72	.035,71 .038,99	50
.50	$DIST_1$ $DIST_2$	.138,81 .136,47	.088,62 .082,58	.045,74 .039,20	.083,55 .078,20	.080,58 .080,81	20
	$DIST_1$ $DIST_2$	.141,24 .143,52	.045,43 .039,13	.042,26 .040,03	.043,38 .037,06	.041,96 .038,96	50
.60	$DIST_1$ $DIST_2$	.170,16 .170,53	.106,74 .107,64	.052,27 .057,63	.102,34 .101,40	.097,60 .091,89	20
	$DIST_1$ $DIST_2$	.159,56 .168,45	.052,07 .057,45	.049,70 .052,25	.049,45 .053,30	.047,72 .050,47	50
.70	$DIST_1$ $DIST_2$	.223,06 .221,33	.131,48 .129,13	.067,34 .066,71	.128,83 .129,23	.136,73 .139,97	20
	$DIST_1$ $DIST_2$	.224,68 .219,80	.067,38 .066,65	.083,89 .076,74	.067,10 .063,70	.086,36 .076,46	50
.80	$DIST_1$ $DIST_2$	.235,22 .239,65	.149,73 .146,12	.104,46 .096,36	.139,52 .137,36	.153,64 .146,16	20
	$DIST_1$ $DIST_2$	.265,05 .257,89	.104,51 .096,25	.092,05 .099,69	.131,33 .095,01	.094,72 .100,28	50
.90	$DIST_1$ $DIST_2$	.304,50 .294,80	.191,49 .175,35	.149,65 .134,87	.174,67 .168,53	.197,18 .184,51	20
	$DIST_1$ $DIST_2$	.313,55 .314,79	.149,60 .134,83	.150,97 .130,26	.148,61 .135,43	.150,13 .132,40	50
.99	$DIST_1$ $DIST_2$	.383,21 .316,12	.291,46 .195,07	.307,14 .212,01	.286,24 .199,06	.307,80 .214,79	20
	$DIST_1$ $DIST_2$	.445,90 .379,74	.307,13 .211,96	.327,07 .217,46	.305,47 .210,38	.330,48 .214 86	50

TABLE II

Number of iterations

ρ	CO						PW						ML					
	T=20			T=50			T=20			T=50			T=20			T=50		
	Min	Max	Ave	Min	Max	Ave	Min	Max	Ave	Min	Max	Ave	Min	Max	Ave	Min	Max	Ave
−.90	2	10	4.28	2	6	3.24	2	6	3.24	2	3	2.51	2	11	4.42	2	5	3.21
−.80	3	10	4.54	3	6	3.24	3	6	3.24	2	3	2.50	3	11	4.70	3	6	3.21
−.70	2	10	4.40	2	7	3.18	2	7	3.18	2	3	2.46	3	11	4.83	2	7	3.20
−.60	2	10	4.44	2	9	3.24	2	9	3.24	2	3	2.45	2	11	5.08	2	9	3.23
−.50	2	10	4.56	2	6	3.18	2	6	3.18	2	3	2.43	2	11	5.04	2	7	3.21
−.40	3	10	4.36	2	6	3.22	2	6	3.22	2	3	2.38	2	11	5.06	2	9	3.22
−.30	2	10	4.42	2	8	3.19	2	8	3.19	1	3	2.29	2	11	5.06	2	11	3.21
−.20	2	10	4.59	2	7	3.17	2	7	3.17	1	3	2.21	2	11	5.16	2	8	3.13
−.10	2	10	4.37	2	10	3.16	2	10	3.16	1	3	2.15	2	11	5.28	2	11	3.10
0	2	10	4.35	2	8	3.09	2	8	3.09	1	3	2.09	2	11	5.21	2	8	3.06
.10	2	10	4.35	2	8	3.12	2	8	3.12	1	3	2.11	2	11	5.63	2	11	3.11
.20	2	10	4.32	2	7	3.20	2	7	3.20	1	3	2.19	2	11	5.69	2	7	3.17
.30	2	10	4.32	2	10	3.31	2	10	3.31	1	4	2.33	2	11	5.69	2	11	3.29
.40	2	10	4.49	2	8	3.39	2	8	3.39	2	4	2.44	2	11	5.95	2	11	3.37
.50	2	10	4.36	2	9	3.56	2	9	3.56	2	4	2.60	2	11	5.86	2	10	3.49
.60	2	10	4.85	2	8	3.70	2	8	3.70	2	4	2.70	2	11	5.94	2	10	3.63
.70	2	10	4.76	2	9	3.86	2	9	3.86	2	4	2.78	2	11	6.20	2	8	3.70
.80	2	10	4.48	2	10	4.23	2	10	4.23	2	7	2.95	2	11	6.17	2	11	3.95
.90	3	10	4.30	2	10	4.57	2	10	4.57	2	9	3.22	3	11	6.59	2	11	4.36
.99	2	10	4.11	2	10	4.58	2	10	4.58	2	10	3.47	2	11	6.89	2	11	4.79

3.1. Numbers of iterations

(i) *Co method:* With the sample of size 20, seventeen out of 20 times, the minimum number of iterations is 2 and only three times this number is 3; the maximum number of iterations is 10 twenty times and the average varies between 4.11 and 4.85. With a sample of size 50, nineteen times the minimum number of iterations is 2 and once it is 3, the maximum is between 6 and 10 and the average is less than 3.5 fourteen times, between 3.5 and 4 three times and between 4 and 4.60 three times.

(ii) *PW method:* The PW method is also an iterative process. But from the computational point of view, the "extra observation" that is thrown in makes the PW method a more efficient method than the CO method.

With the sample of size 20, the minimum number of iterations is almost always 2, the maximum of it is between 6 and 10 and almost always the average is less than 3.7, except for the five extreme positive values of ρ. With the sample of size 50, the minimum number of iterations drops to 1 for all $|\rho| \leqq .30$; the maximum is often 3, reaching 4 five times and is larger than 7 two times which also occurs with extreme positive values of ρ; the average number of iterations varies between 2.11 and 3.47.

(iii) *ML method:* With the sample of size 20, the average number of iterations increases almost regularly and monotonically from 4.42 to 6.89 as ρ goes from $-.90$ to .99. Almost the same occurs with sample size 50 for which the same average varies from 3.21 to 4.79. Almost always the minimum number of iterations is 2 except at $\rho=-.80$, $-.70$, and $-.90$ with sample of size 20 and at $\rho=-.80$ with the sample of size 50 for which this minimum is 3. The maximum of number of iterations is always 11 when sample is of size 20, and is between 5 and 11 when the sample is of size 50.

3.2. Effects of sample sizes on distances

With OLS an increase in the sample size does not accompany with a significant reduction in $DIST_i$ ($i=1, 2$). As a matter of fact, the increase in the sample size from 20 to 50 improves (the smaller the better) $DIST_1$ only for three values of ρ ($\rho=-.10$, 0, .60) and improves $DIST_2$ only for two values of ρ ($\rho=-.10$, .70). Thus, the reliability of the OLS method does not increase with the sample size.

Similar conclusion can be maintained for the PW method. However, it is clear that the reliability of the remaining methods of estimation, namely the CO, Durbin and ML methods, does increase with the sample size.

3.3. Effects of the autoregressive parameter on distances

(i) *OLS method:* In this case, with a sample size of 20, $DIST_1$ increases as ρ deviates from zero. Equal deviations of ρ in either direction invarably affects these distances almost equally if $|\rho| < .0.4$. If $|\rho| > 0.4$, the effects seem more serious if the deviation is in the positive direction than if it is in the negative direction. Similar conclusions can be reached for $DIST_1$ with sample size of 50 and for $DIST_2$ with both sample sizes of 20 and 50.

(ii) *The other methods:* For each of the four remaining methods, we find both $DIST_1$ and $DIST_2$ with both sample sizes are rather insensitive to the deviations of ρ in the negative direction, whereas these distances are sensitive to the deviations of ρ in the positive direction. It implies that inference errors using these estimators may be more serious when ρ is positive than when ρ is negative.

3.4. Comparison of the estimators in terms of their distances

(i) *Sample size of* 20: It appears clearly that $DIST_1$ is the largest for the OLS estimator if ρ is not too close to zero, possibly if $|\rho| > .20$. If ρ is close to zero, $DIST_1$'s for all the estimators are comparable. If ρ is negative, $DIST_1$'s for the four estimators CO, PW, Durbin and ML are almost indistinguishable. However, if ρ is positive, it seems that $DIST_1$ for the PW estimator is the smallest.

Thus, in estimating β_1, we conclude that except when ρ is close to zero, the OLS estimator should not be preferred. If ρ is negative, there is no clear distinction between the four estimators CO, PW, Durbin and ML. However if ρ is positive, then the PW estimator is clearly preferable. But it should be recognized that PW estimators may give seriously misleading inference if ρ is close to 1.

Similar conclusion can be reached for $DIST_2$.

(ii) *Sample Size of* 50: Again, and very clearly, we find that the OLS estimation of β_1 should not be recommended unless ρ is close to zero. And also as in the case of the sample size of 20, we find that the four other estimators of β_1 can not be distinguished when ρ is negative. However, in contrast to the case of sample size of 20 where PW estimator was preferred to the other estimators when ρ is positive, we find that the sample size of 50, the four estimators are indistinguishable.

Similar conclusions can be reached for $DIST_2$.

4. CONCLUDING REMARKS

Our remarks are twofold. First, in the case of the first order autoregressive model, the use of the Kolmogorov-Smirnov distance leads to results somewhat different from those based on the MSE criterion. Mainly (i) the OLS method should never be used unless one have a prior information that the correlation coefficient is very close to zero, possibly less than .20 in absolute value, and (ii) although all other methods seem to be equivalent to one another, the PW method may be singled out as the best. Second, it is quite possible to use the Kolmogorov-Smirnov distance as choice criterion in other econometrics models for which more than one methods of estimation are available.

REFERENCES

1. B. P. ADHIKARI and D. D. JOSHI, *Distance-discrimination et résumé exhaustif*, Publications de l'Institut de Statistique de l'Université de Paris, Vol. 5, 1956, pp. 57-74.
2. S. M. ALI and S. D. SILVEY, *A General Class of Coefficients of Divergence of One Distribution from Another*, The Journal of Royal Statistical Society, Sr. B, Vol. 58, 1966, pp. 131-142.
3. C. M. BEACH and J. G. MACKINNON, *A Maximum Likelihood Procedure for Regression with Autocorrelated Errors*, Econometrica, Vol. 46, No. 1, 1978, pp. 51-58.
4. J. DURBIN, *Estimation of Parameters in Time-Series Regression Models*, The Journal of Royal Statistical Society, Sr. B, Vol. 22, 1960, pp. 139-153.
5. L. RAO and Z. GRILICHES, *Small-Samples Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors*, Journal of the American Statistical Association, Vol. 64, 1969, pp. 253-272.