MONASH

20/93

M O N A S H

U N I V E R S I T Y

AUSTRALIA

THE USE OF INFORMATION CRITERIA FOR MODEL SELECTION BETWEEN MODELS WITH EQUAL NUMBERS OF PARAMETERS

Simone D. Grose and Maxwell L. King

Working Paper No. 20/93

December 1993

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0388 9

THE USE OF INFORMATION CRITERIA FOR MODEL SELECTION BETWEEN MODELS WITH EQUAL NUMBERS OF PARAMETERS

Simone D. Grose and Maxwell L. King

Working Paper No. 20/93

December 1993

DEPARTMENT OF ECONOMETRICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

The Use of Information Criteria for Model Selection between Models with Equal Numbers of Parameters

Simone D. Grose and Maxwell L. King*

Department of Econometrics

Monash University

Clayton, Victoria 3168

Australia

Abstract

Information criteria (IC) are used widely to choose between competing alternative models. When these models have the same number of parameters, the choice simplifies to the model with the largest maximized log-likelihood. By studying the problem of selecting either first-order autoregressive or first-order moving average disturbances in the linear regression model, we present clear evidence that a particular model can be unfairly favoured because of the shape or functional form of its log-likelihood. We also find that the presence of nuisance parameters can adversely affect the probabilities of correct selection. The use of Monte Carlo methods to find more appropriate penalties and the application of IC procedures to marginal likelihoods rather than conventional likelihoods is found to result in improved selection probabilities in small samples.

* This research was supported in part by an Australian Research Council grant. We are grateful to Tim Fry, David Harris and Catherine Scipione for helpful suggestions and to Alan Morgan for research assistance.

1. Introduction

Often in statistics and particularly in non-experimental disciplines such as econometrics we are forced to use the available data to make a choice between a number of competing alternative models. One approach is to use a series of pairwise hypothesis tests. As Granger et al. (1993) note, this has a number of limitations. They argue as others do that such model building decisions should be based on well-thought-out model selection procedures. The consensus in the literature seems to be for the use of an information criterion (IC) based on minus the maximized log-likelihood function plus a penalty function for the number of parameters in the model. There is little agreement about what this penalty function should be. For example, for the Akaike IC (AIC) it is q and for Schwarz's Bayesian IC (BIC) it is qlog(n)/2 where q is the number of parameters in the model and n is the sample size. BIC does have an asymptotic justification but little has been written about small-sample based penalty functions. The small-sample based correction to AIC proposed by Hurvich and Tsai (1989) involves a modification to the penalty function which is still a function only of q and n.

When the choice is between models with the same number of parameters, it appears that all IC are in agreement - the model with the largest maximized log-likelihood is chosen. While this may be justified by asymptotic arguments, in small samples it is not clear it is entirely appropriate. It is possible for certain models to be favoured purely because of the shape or functional form of their log-likelihood function. The main aim of this paper is to investigate this question.

We have chosen to conduct our investigation on the problem of selecting between first-order autoregressive (AR(1)) and first-order moving average (MA(1)) disturbances in the linear regression model. There is a vast literature on the former model that dates back to the work by Cochrane and Orcutt (1949) and Durbin and Watson (1950, 1951). Subsequently, the MA(1) disturbance model has been recognised by econometricians as a possible alternative for both economic reasons (see for example Nicholls, Pagan and Terrell 1975; Rowley and Wilton 1973 and Sims 1974) and statistical reasons such as the Durbin-Watson test having good power against MA(1) disturbances. This has led to the development of a number of tests of AR(1) disturbances against MA(1) disturbances and vice versa (see for example King 1983, 1987; King and McAleer 1987; Burke, Godfrey and Tremayne 1990 and Silvapulle and King 1991).

Our chosen problem of selecting between AR(1) and MA(1) disturbances in the linear regression model is invariant to transformations involving a change of scale of the dependent variable and the addition of a known linear combination of the regressors. We can find a maximal invariant statistic whose distribution depends only on the autocorrelation parameter of the AR(1) or MA(1) process. A secondary aim of this paper is to consider whether using the likelihood of the maximal invariant improves inferences for this model selection problem by providing a better treatment of nuisance parameters. Ara and King (1993) have shown this likelihood is equivalent to the marginal likelihood for the parameter in question. There is considerable evidence (see Tunnicliffe Wilson 1989 and Ara and King) that maximum likelihood estimates based on the marginal likelihood are less biased than their conventional counterparts.

The plan of this paper is as follows. In section 2, we present the two competing models and discuss two properties that an ideal model selection procedure should have. The section concludes by reporting a Monte Carlo experiment conducted to investigate whether there are problems with a particular model being favoured and also with the treatment of nuisance parameters. Strong evidence of both problems is found. Section 3 discusses ways of dealing with each of these problems. The results of a Monte Carlo experiment conducted to evaluate these suggestions are reported in section 4. Some concluding remarks are made in the final section.

2. Are selection probabilities affected by the shape of the likelihood function?

2.1 Introduction

Consider the linear regression model

$$y = X\beta + u , \tag{1}$$

where y is an nx1 vector, X is an nxk nonstochastic matrix of rank $k < n$ and β is a kx1 parameter vector. We will answer the above question in the context of the selection problem in which the elements of the nx1 disturbance vector u are either generated by the stationary AR(1) process

$$u_t = \rho u_{t-1} + e_t , \qquad |\rho| < 1 , \qquad t = 1,\dots,n \tag{2}$$

where $u_0 \sim N(0,\sigma^2/(1-\rho^2))$ and $e = (e_1,\dots,e_n)' \sim N(0,\sigma^2 I_n)$, or by the MA(1) process

$$u_t = e_t + \gamma e_{t-1} , \qquad t = 1,\dots,n \tag{3}$$

where $e^* = (e_0,e')' \sim N(0,\sigma^2 I_{n+1})$. Under (2), $u \sim N(0,\sigma^2\Sigma(\rho))$, where $\Sigma(\rho)$ is the nxn matrix whose $(i,j)^{th}$ element is $\rho^{|i-j|}/(1 - \rho^2)$. Under (3), $u \sim N(0,\sigma^2\Omega(\gamma))$ where $\Omega(\gamma)$ is the nxn tridiagonal matrix with $1 + \gamma^2$ as the main diagonal elements and γ as the nonzero off-diagonal elements. For this selection problem, β and σ^2 are nuisance parameters.

Our focus is on all model selection procedures which can be applied in the form: choose the model which minimizes

$$- L_i(\hat{\theta}) + f(n_i, q_i) \tag{4}$$

where $L_i(\hat{\theta})$ is the maximized log-likelihood function of the i^{th} model and $f(n_i, q_i)$ is a penalty function which depends purely on the number of observations n_i and the number of parameters q_i in the i^{th} model. The two competing models in our selection problem have the same number of unknown parameters in which case (4) reduces to choosing the model with the largest maximized likelihood. Under (1) and (2), the concentrated log-likelihood is

$$- \frac{n}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma(\rho)| - \frac{n}{2} \log\left\{(y - X\hat{\beta})'\Sigma^{-1}(\rho)(y - X\hat{\beta})/n\right\} - \frac{n}{2} \tag{5}$$

where

$$\hat{\beta} = (X'\Sigma^{-1}(\rho)X)^{-1}X'\Sigma^{-1}(\rho)y . \tag{6}$$

The maximized log-likelihood is (5) with ρ taking that value, $\hat{\rho}$, which maximizes (5). In the case of (1) and (3), $\Omega(\gamma)$ replaces $\Sigma(\rho)$ in (5) and (6) and now (5) is maximized with respect to γ.

Observe that when $\rho = 0$ and $\gamma = 0$, (2) and (3) become the same model, namely $u_t = e_t$. In other words, the parameter spaces of the two models intersect when $\rho = 0$ and $\gamma = 0$. If the choice is restricted to only (2) and (3), then a good model selection procedure would be indifferent between the two models; i.e., the probability of choosing (2) when $u_t = e_t$ would be one half. We will call this property 1. Another highly desirable property would be that this probability increases as $|\rho|$ increases from zero when (2) is the true model. Similarly for the MA(1) disturbance model. This will be called property 2.

2.2 Monte Carlo Experiment

In order to check whether procedures of the form (4) have these properties in our case, we conducted a computer simulation experiment. The Monte Carlo method was used to estimate probabilities of correct selection when regression disturbances are generated from the AR(1) model (2) for ρ = -0.9, -0.8, ..., -0.1, 0, 0.1, ..., 0.9, and from the MA(1) model (3) for the same range of γ values. The following X matrices with n = 20 and 50 were used:

X1 : (nx1). The constant dummy as the only regressor.

X2 : (nx2). The constant dummy and time trend.

X3 : (nx3). The eigenvectors corresponding to the three smallest eigenvalues of the Durbin-Watson (DW) nxn A_1 matrix and hence to the upper bound of the DW statistic. A_1 is a tridiagonal matrix with 2's down the main diagonal, -1's on the off-diagonal and 1's in the top left and bottom right elements.

X4 : (nx3). The eigenvectors corresponding to the zero and two largest eigenvalues of the nxn DW A_1 matrix and hence to the lower bound of the DW statistic.

X5 : (nx3). The first n observations of Durbin and Watson's (1951, p.159) consumption of spirits example.

X6 : (nx3). A constant dummy, the quarterly Australian Consumer Price Index commencing 1959(1) and the same index lagged one quarter.

X7 : (nx3). A constant dummy, quarterly Australian private capital movements and quarterly Australian Government capital movements.

X8 : (nx6). A full set of quarterly seasonal dummy variables plus quarterly seasonally adjusted Australian household disposable income and private final consumption expenditure commencing 1959(4).

X9 : (n×5). A constant dummy plus four independent trending regressors generated as

$$x_{ti} = z_{ti} + 0.25t$$

where z_{ti}; $t = 1,\ldots,n$; $i = 2,\ldots,5$; are mutually independent AR(1) time series generated from

$$z_{ti} = 0.5z_{t-1,i} + \eta_{ti}$$

and $\eta_{ti} \sim IN(0,1)$; $t = 1,\ldots,n$; $i = 2,\ldots,5$.

These design matrices cover a range of behaviour. X1 is the special case of the Gaussian time-series model with unknown mean. A rough estimate of the disturbance autocorrelation is given by (2-d)/2 where d is the DW statistic. We therefore expect X3 and X4 to show some extreme behaviour. X5 is based on annual data while X6, X7 and X8 use quarterly data. The regressors of X7 are strongly seasonal with two seasonal peaks per year plus some large fluctuations. We also estimated probabilities of correct selection in the pure time-series model with no regressors; i.e. $y_t = u_t$. This case was included in order to assess the influence of the $X\beta$ term on the probabilities of making a correct selection.

Maximum likelihood estimates under (1) and (2) were computed using Beach and MacKinnon's (1978) algorithm, while Pesaran's (1973) transformation, together with the IMSL non-linear maximization subroutine, DUVMIF, were used to obtain maximum likelihood estimates under (1) and (2). The error variance, σ^2, was set to unity and the components of β were set to zero because one can show, using the results of Breusch (1980), that each of the estimated probabilities is invariant to the values taken by β and σ^2. Five thousand replications were used and pseudo-random N(0,1) variates were generated as described by King and Giles (1984).

2.3 Results

Selected estimated probabilities of correctly choosing the AR(1) error model for each of the design matrices for the regression model and also for the pure time-series model are given in Table 1. The corresponding probabilities of correctly choosing MA(1) errors are presented in Table 2. Plots of these probabilities for the X9 regressor matrix are given in Figure 1. Note that the probabilities for $\rho = 0$ in Table 1 and the corresponding probabilities for $\gamma = 0$ in Table 2 sum to one.

The most striking feature of the results for the regression model is that the probabilities of selecting the AR(1) error model when $\rho = 0$ are all significantly below 0.5. They range from 0.105 to 0.333 when n = 20 and from 0.293 to 0.415 when n = 50, with the largest probabilities occurring when there are no nonconstant regressors (X1). These probabilities increase significantly as n increases *ceteris paribus*. As expected, X3 and X4 show the most extreme behaviour particularly when n = 50. Another feature is that for n = 20 and a given design matrix with nonconstant regressors, the lowest probability of correctly choosing the AR(1) model does not occur at $\rho = 0$ but typically at $\rho = -0.2$ or $\rho = -0.4$ while for X4 it is at $\rho = 0.2$. For n = 50, the lowest probability always occurs at $\rho = 0$, even when the results for $\rho = 0.1, -0.1$ that have been omitted from Table 1, are included in the analysis. A similar but less pronounced pattern may be seen in Table 2. There the lowest probabilities of correctly choosing the MA(1) model typically occur at $\gamma = 0.2$ when n = 20 and always at $\gamma = 0$ when n = 50. Furthermore, all probabilities in Table 2 are greater than 0.5. Because probabilities at $\gamma = 0$ decrease significantly towards 0.5 as n increases, we find that for a wide range of γ values, the probability of correctly choosing the MA(1) model decreases as n increases.

With respect to the results for the pure time-series model $y_t = u_t$, we again see that the probabilities of selecting the AR(1) model when $\rho = 0$ are below 0.5. However they are much closer to 0.5 than for the regression model and show signs of rapid convergence to 0.5 as n increases. Furthermore we find that property 2 holds in this case.

These results are somewhat disturbing. They suggest that in the very simple case of choosing between two different models with the same number of parameters, choosing that model with the largest maximized likelihood can result in the correct choice with a probability as low as 0.076 when $n = 20$. When both models are true ($\rho = 0$ and $\gamma = 0$), the probabilities of selection do not split evenly but clearly favour the MA(1) model particularly when n is small. Thus property 1 obviously does not hold. It does appear that the functional form of the likelihood of the MA(1) model gives it a much better than even chance of having a larger maximized likelihood than that of the AR(1) model. This suggests that as well as correcting for the number of parameters we should also consider correcting for the functional form of the likelihood particularly in small samples. Of less concern is the fact that property 2 does not hold when $n = 20$ in the regression. It appears to hold for $n = 50$ and also for the pure time-series models.

We also see clear evidence of nuisance parameters adversely affecting the probabilities of correct selection. The best results occur for the time-series models in which there are no β parameters. The next best results occur when the only regressor is the constant dummy. We therefore conjecture that better handling of nuisance parameters may be the key to improving the small-sample properties of IC based selection procedures.

3. An alternative approach

Procedures of the form of (4) can be easily modified to have property 1 in our case of selecting between the two disturbance models. Let $\hat{L}_\rho$ and $\hat{L}_\gamma$ denote the maximized log-likelihood functions of the two models and let p_ρ and p_γ denote respective penalty functions. Our problem is to find p_ρ and p_γ such that

$$\Pr\left[\hat{L}_\gamma - p_\gamma > \hat{L}_\rho - p_\rho \mid u_t \sim IN(0,\sigma^2)\right] = 0.5$$

or equivalently

$$\Pr\left[(\hat{L}_\gamma - \hat{L}_\rho) > (p_\gamma - p_\rho) \mid u_t \sim IN(0,\sigma^2)\right] = 0.5 . \qquad (7)$$

Observe that (7) is in the familiar form that defines a critical value for a test statistic. The test statistic is $\hat{L}_\gamma - \hat{L}_\rho$ which is the log of a likelihood ratio, $(p_\gamma - p_\rho)$ is the unknown critical value and is chosen to set the probability of "rejection" equal to 0.5. From (7) it is clear that we can only determine $p_\gamma - p_\rho$, in which case it is sensible to set one of the penalty functions equal to zero, say p_ρ. Equation (7) also suggests that the value for p_γ can be estimated using the Monte Carlo method. This would involve generating y_t as IN(0,1) random variables, computing $\hat{L}_\gamma - \hat{L}_\rho$ and repeating this many times to build up the empirical distribution of $\hat{L}_\gamma - \hat{L}_\rho$. The required p_γ value is the median of this empirical distribution. Note that in the simulations, we are able to set $\sigma^2 = 1$ and $\beta = 0$ because of invariance.

While this new penalty function will result in property 1 holding (at least approximately) it does nothing to guarantee probabilities of making the correct selection increase as $|\rho|$ or $|\gamma|$ increases (property 2). In the previous section, we conjectured that better handling of nuisance parameters might result in selection procedures with better small-sample properties. A standard and very successful method of dealing with nuisance parameters in the linear regression when making inference concerning the disturbance vector is through the principle of invariance.

In our case, choosing between (2) and (3) in the context of (1) is invariant to transformations of the form

$$y \longrightarrow \eta_0 y + X\eta \tag{8}$$

where η_0 is a positive scalar and η is a k×1 vector. Let $m = n-k$, $M = I_n - X(X'X)^{-1}X'$, $z = My$ be the ordinary least squares residual vector from (1) and P be an m×n matrix such that $PP' = I_m$ and $P'P = M$. The m×1 normalized vector

$$v = Pz \,/\, (z'P'Pz)^{1/2}$$

is a maximal invariant under the group of transformations of the form of (8). The joint density function of v under (2) and (3) can be shown to be, respectively (see King (1980)),

$$f_1(v;\rho)dv = \tfrac{1}{2}\Gamma(m/2)\pi^{-m/2}|P\Sigma(\rho)P'|^{-1/2}\left\{v'(P\Sigma(\rho)P')^{-1}v\right\}^{-m/2}dv \tag{9}$$

and

$$f_2(v;\gamma)dv = \tfrac{1}{2}\Gamma(m/2)\pi^{-m/2}|P\Omega(\gamma)P'|^{-1/2}\left\{v'(P\Omega(\gamma)P')^{-1}v\right\}^{-m/2}dv \tag{10}$$

where dv denotes the uniform measure on the surface of the unit m-sphere. Also note that for any n×n positive definite matrix Λ

$$v'(P\Lambda P')^{-1}v = \hat{u}'\Lambda^{-1}\hat{u} \,/\, z'z ,$$

where $\hat{u}$ is the generalized least squares (GLS) residual vector from (1) assuming covariance matrix $\sigma^2\Lambda$.

Because our selection problem is invariant to transformations of the form of (8), we need only consider selection procedures that are invariant to such transformations. The principle of invariance implies that invariant selection procedures can be constructed by treating v as the observed data and (9) and (10) as its likelihood function for AR(1) and MA(1) errors respectively. This suggests the IC procedures should be applied to v rather than the original y. In our case the AR(1) model is chosen if the log of (9) maximized with respect to ρ achieves a higher value than the log of (10) maximized with respect to γ. Otherwise the MA(1) model is chosen.

Both AR(1) and MA(1) disturbances are special cases of regression disturbances distributed as $u \sim N(0,\sigma^2\Lambda(\lambda))$ where $\Lambda(\cdot)$ is an nxn positive definite matrix function and λ is a px1 vector of unknown parameters. The problem of choosing a particular form of $\Lambda(\lambda)$ from a given range of possible $\Lambda(\lambda)$ matrix functions in the context of (1) is invariant to transformations of the form of (8). The above arguments for treating v as the observed data therefore apply for this more general problem. Ara and King (1993) have shown that for $u \sim N(0,\sigma^2\Lambda(\lambda))$, the likelihood of λ constructed as the joint density of v is equivalent to the marginal likelihood for λ which from Tunnicliffe Wilson (1989) is given by

$$f_m(\lambda|y) = |\Lambda(\lambda)|^{-1/2}|X'\Lambda^{-1}(\lambda)X|^{-1/2}(\hat{u}'\Lambda^{-1}(\lambda)\hat{u})^{-m/2} \qquad (11)$$

where $\hat{u}$ is the GLS residual vector from (1) assuming covariance matrix $\sigma^2\Lambda(\lambda)$.

In the case of choosing different $\Lambda(\lambda)$ specifications for the distribution of u in (1), this implies that our suggestion of applying IC procedures to v is equivalent to applying them to the marginal likelihoods (11). We therefore call these marginal likelihood based IC (MIC) procedures. In our case of choosing between AR(1) and MA(1) disturbances, all IC procedures of the form of (4) result in the same MIC procedure, namely choose that model which results in the largest maximized value of (11). Of course we can use the MIC procedure in conjunction with the Monte Carlo method for determining a penalty function that ensures property 1 holds (at least approximately).

4. Monte Carlo Experiment

The Monte Carlo experiment in the context of the regression model (1), outlined in Section 2, was repeated for each of the new procedures discussed in the previous section. These procedures are:

(i) IC with empirically calculated penalty functions (ICE),

(ii) MIC,

(iii) MIC with empirically calculated penalty functions (MICE).

Where required, logged marginal likelihoods were maximized using the IMSL routine DUVMIF. Selected estimated probabilities of correctly selecting the AR(1) disturbance model and the MA(1) disturbance model for $n = 20$ and $n = 50$ are presented in Tables 3 - 6. Plots of these probabilities, together with those for traditional IC procedures, are given in Figures 2 and 3 for X9 with $n = 20$ and $n = 50$, respectively.

It is clear from these results that the faults with standard IC procedures identified in section 2 cannot be rectified purely by finding penalty functions that make property 1 hold. For both sample sizes, the ICE procedure has probabilities of correctly selecting the AR(1) model that decline from 0.5 at $\rho = 0$ as ρ decreases (increases in the case of X4) from zero. These probabilities of correct selection reach as low as 0.257 when $n = 20$ and 0.391 when $n = 50$. For some X matrices such as X4 and X8, the probability of the ICE procedure correctly selecting the AR(1) model falls below 0.5 for ρ values on both sides of $\rho = 0$. This suggests that the ICE procedure tends to be biased against the AR(1) model. It has probabilities of correctly selecting the MA(1) model significantly below 0.5 only when $n = 20$ and for X2, X5, X6, X7 and X9 - the lowest probability being 0.374 at $\rho = 0.4$ for X9.

In contrast, property 2 always holds for the MIC procedures, at least for probabilities calculated at intervals of 0.1 with respect to ρ and γ. Again we see that the calculated probabilities at $\rho = \gamma = 0$ favour the MA(1) model confirming that property 1 does not hold. Probabilities at $\rho = 0$ of MIC procedures selecting the AR(1) model range from 0.429 to 0.441 when $n = 20$ and from 0.464 to 0.473 when $n = 50$. These probabilities are promising in the sense that they seem to be rapidly approaching 0.5 as n increases. It is interesting to note their lack of variability and their similarity to the probabilities of 0.445 ($n = 20$) and 0.472 ($n = 50$) we calculated for the IC procedures in the case of the pure time-series model with no regressors.

Finally we observe that the MICE procedures appear to obey both properties 1 and 2. In the case of $n = 20$, there are some calculated MICE probabilities at $|\rho| = 0.1$ in Table 3 which are below 0.5 but they are far from being significantly different than 0.5.

5. Concluding Remarks

When choosing between models with the same number of parameters, all IC model selection procedures of the form of (4) agree - the model with the largest maximized log-likelihood should be chosen. By studying this selection rule in the context of choosing between AR(1) and MA(1) errors in the linear regression model, we identified two problems with it in small samples. We found clear evidence that shows a particular model can be unfairly favoured purely because of the shape or functional form of its log-likelihood function. We also found evidence of nuisance parameters adversely affecting the probabilities of correct selection.

There is clearly a case for using penalty functions that take into account the shape or functional forms of the log-likelihood functions of the models under consideration. This can be done empirically using Monte Carlo methods. When selecting from two models which effectively overlap at one point in the parameter space, finding appropriate penalties is analogous to finding a critical value for a test statistic. This approach can be generalized to cases of selecting from more than two models - a topic we hope to report on in a future paper.

In the context of choosing between different forms of the covariance matrix in the linear regression model, we conclude that IC procedures should be applied to marginal likelihoods rather than conventional likelihoods. This is because the former approach results in better small-sample properties. The best results in our study were obtained when marginal likelihoods are used with penalty functions that take account of the functional form of the log-likelihood. Of the two issues identified in this paper, the more serious would seem to be the better treatment of nuisance parameters. In our view, it is a subject worthy of further research.

REFERENCES

Ara, I., and King, M.L. (1993), "Marginal Likelihood Based Tests of Regression Disturbances," Mimeo, Monash University, Department of Econometrics.

Beach, C.M., and MacKinnon, J.G. (1978), "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," *Econometrica*, 46, 51-58.

Breusch, T.S. (1980), "Useful Invariance Results for Generalized Regression Models," *Journal of Econometrics*, 13, 327-340.

Burke, S.P., Godfrey, L.G., and Tremayne, A.R. (1990), "Testing AR(1) Against MA(1) Disturbances in the Linear Regression Model: An Alternative Approach," *Review of Economic Studies*, 57, 135-145.

Cochrane, D., and Orcutt, G.H. (1949), "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms," *Journal of the American Statistical Association*, 44, 32-61.

Durbin, J., and Watson, G.S. (1950), "Testing for Serial Correlation in Least Squares Regression I," *Biometrika*, 37, 409-428.

Durbin, J., and Watson, G.S. (1951), "Testing for Serial Correlation in Least Squares Regression II," *Biometrika*, 38, 159-178.

Granger, C.W.J., King, M.L., and White, H. (1993), "Testing Economic Theories and the Use of Model Selection Criteria," *Journal of Econometrics*, forthcoming.

Hurvich, C.M., and Tsai, C.-L. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76, 297-307.

King, M.L. (1980), "Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression," *The Annals of Statistics*, 8, 1265-1271.

King, M.L. (1983), "Testing for Autoregressive Against Moving Average Errors in the Linear Regression Model," *Journal of Econometrics*, 21, 35-51.

King, M.L. (1987), "Towards a Theory of Point-Optimal Testing," *Econometric Reviews*, 6, 169-218.

King, M.L., and Giles, D.E.A. (1984), "Autocorrelation Pre-Testing in the Linear Model: Estimation, Testing and Prediction," *Journal of Econometrics*, 25, 35-48.

King, M.L., and McAleer, M. (1987), "Further Results on Testing AR(1) Against MA(1) Disturbances in the Linear Regression Model," *Review of Economic Studies*, 54, 649-663.

Nicholls, D.F., Pagan, A.R., and Terrell, R.D. (1975), "The Estimation and Use of Models with Moving Average Disturbance Terms: A Survey," *International Economic Review*, 16, 113-134.

Pesaran, M.H. (1973), "Exact Maximum Likelihood Estimation of a Regression Equation with a First-Order Moving Average Error," *Review of Economic Studies*, 40, 529-535.

Rowley, J.C.R., and Wilton, P.A. (1973), "Quarterly Models of Wage Determination: Some New Efficient Estimates," *American Economic Review*, 63, 380-389.

Silvapulle, P., and King, M.L. (1991), "Testing Moving Average Against Autoregressive Disturbances in the Linear Regression Model," *Journal of Business and Economic Statistics*, 9, 329-335.

Sims, C.A. (1974), "Distributed Lags," in eds. M.D. Intriligator and D.A. Kendrick, *Frontiers of Quantitative Economics 2*, Amsterdam: North-Holland, pp.289-338.

Tunnicliffe Wilson, G. (1989), "On the Use of Marginal Likelihood in Time Series Model Estimation," *Journal of the Royal Statistical Society B*, 51, 15-27.

Table 1: Estimated probabilities of correctly choosing AR(1) disturbances based on choosing the model with the largest maximized likelihood.

		ρ = -.8	-.6	-.4	-.2	0	.2	.4	.6	.8
						The regression model $y = X\beta + u$				
X1	20	.788	.546	.409	.351	.333	.363	.436	.575	.745
	50	.974	.819	.617	.461	.415	.463	.616	.820	.965
X2	20	.670	.346	.232	.215	.227	.258	.317	.417	.541
	50	.954	.752	.541	.398	.359	.399	.538	.752	.933
X3	20	.573	.192	.076	.078	.105	.140	.180	.221	.281
	50	.878	.611	.489	.325	.294	.334	.455	.658	.864
X4	20	.405	.267	.204	.166	.149	.142	.143	.191	.385
	50	.889	.685	.463	.332	.293	.326	.450	.665	.886
X5	20	.640	.302	.185	.179	.205	.245	.309	.407	.535
	50	.919	.705	.505	.379	.338	.381	.519	.731	.918
X6	20	.641	.290	.175	.151	.165	.193	.253	.347	.482
	50	.922	.694	.497	.365	.328	.375	.509	.729	.916
X7	20	.607	.240	.126	.142	.182	.235	.295	.390	.567
	50	.929	.737	.557	.418	.381	.437	.582	.784	.949
X8	20	.534	.330	.274	.282	.305	.331	.357	.399	.477
	50	.895	.700	.535	.421	.393	.435	.557	.751	.914
X9	20	.624	.237	.099	.092	.122	.172	.223	.302	.403
	50	.909	.671	.509	.375	.341	.380	.511	.713	.907
						The pure time-series model $y = u$				
	20	.884	.712	.560	.470	.445	.469	.557	.717	.890
	50	.987	.870	.682	.523	.472	.524	.683	.873	.987

Table 2: Estimated probabilities of correctly choosing MA(1) disturbances based on choosing the model with the largest maximized likelihood.

		γ = -.8	-.6	-.4	-.2	0	.2	.4	.6	.8
				The regression model $y = X\beta + u$						
X1	20	.942	.877	.780	.700	.667	.687	.762	.861	.938
	50	.986	.917	.769	.640	.585	.637	.771	.924	.989
X2	20	.969	.944	.890	.818	.773	.783	.837	.913	.968
	50	.988	.915	.796	.692	.641	.685	.815	.945	.993
X3	20	.987	.983	.968	.936	.895	.882	.917	.963	.989
	50	.994	.940	.844	.747	.706	.743	.858	.960	.996
X4	20	.979	.951	.903	.859	.851	.882	.933	.970	.985
	50	.993	.955	.859	.753	.707	.755	.868	.957	.994
X5	20	.969	.952	.910	.845	.795	.792	.842	.907	.963
	50	.991	.926	.817	.712	.662	.707	.824	.942	.993
X6	20	.975	.960	.917	.873	.835	.831	.865	.921	.965
	50	.991	.933	.822	.722	.672	.708	.824	.947	.992
X7	20	.976	.964	.936	.877	.818	.796	.836	.903	.946
	50	.987	.890	.774	.672	.619	.659	.780	.922	.986
X8	20	.938	.899	.831	.749	.695	.709	.778	.868	.931
	50	.988	.918	.778	.659	.607	.650	.777	.921	.986
X9	20	.971	.971	.958	.922	.878	.854	.876	.916	.957
	50	.989	.910	.802	.707	.659	.705	.821	.938	.990
				The pure time-series model $y = u$						
	20	.909	.803	.679	.590	.555	.588	.678	.807	.903
	50	.982	.903	.725	.585	.528	.584	.728	.897	.985

Table 3: Estimated probabilities of correctly choosing AR(1) disturbances based on ICE, MIC and MICE when n = 20.

	ρ =	-.8	-.6	-.4	-.2	-.1	0	.1	.2	.4	.6	.8
X1	ICE	.792	.558	.453	.454	.475	.500	.511	.504	.518	.607	.755
	MIC	.892	.720	.558	.471	.448	.441	.445	.463	.545	.672	.822
	MICE	.893	.724	.572	.512	.497	.500	.492	.501	.562	.677	.823
X2	ICE	.680	.372	.303	.374	.442	.500	.539	.559	.532	.544	.618
	MIC	.900	.721	.547	.463	.445	.438	.443	.463	.527	.628	.740
	MICE	.902	.726	.564	.504	.501	.500	.499	.509	.547	.637	.744
X3	ICE	.737	.396	.269	.333	.405	.500	.589	.675	.782	.821	.834
	MIC	.916	.745	.554	.460	.445	.435	.439	.449	.500	.577	.651
	MICE	.917	.752	.580	.506	.505	.500	.499	.503	.534	.592	.661
X4	ICE	.495	.426	.453	.507	.515	.500	.452	.383	.264	.247	.417
	MIC	.753	.617	.513	.453	.434	.429	.433	.452	.534	.678	.837
	MICE	.760	.630	.547	.506	.497	.500	.499	.511	.561	.687	.840
X5	ICE	.661	.337	.257	.344	.429	.500	.559	.594	.600	.603	.661
	MIC	.903	.723	.553	.463	.445	.440	.441	.458	.518	.619	.727
	MICE	.904	.729	.575	.508	.506	.500	.503	.506	.543	.632	.731
X6	ICE	.688	.361	.281	.355	.428	.500	.576	.632	.677	.679	.707
	MIC	.907	.724	.539	.457	.439	.430	.437	.452	.506	.594	.719
	MICE	.907	.732	.565	.507	.507	.500	.499	.505	.537	.609	.726
X7	ICE	.698	.359	.259	.349	.428	.500	.569	.616	.639	.652	.734
	MIC	.911	.724	.549	.459	.439	.429	.434	.451	.515	.642	.800
	MICE	.912	.731	.576	.506	.503	.500	.504	.508	.542	.654	.802
X8	ICE	.546	.363	.349	.427	.471	.500	.516	.519	.483	.473	.518
	MIC	.810	.656	.521	.460	.444	.441	.449	.463	.511	.607	.716
	MICE	.812	.664	.545	.506	.495	.500	.503	.502	.533	.619	.722
X9	ICE	.818	.546	.373	.383	.434	.500	.563	.622	.708	.753	.794
	MIC	.906	.725	.547	.463	.441	.433	.437	.418	.501	.595	.722
	MICE	.908	.734	.574	.514	.498	.500	.500	.502	.529	.611	.730

Table 4: Estimated probabilities of correctly choosing MA(1) disturbances based on ICE, MIC and MICE when n = 20.

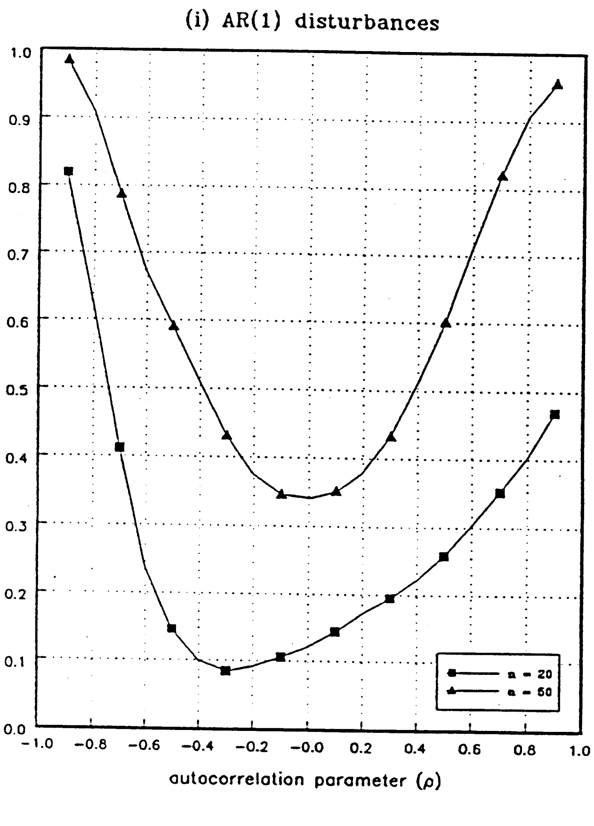
	γ =	-.8	-.6	-.4	-.2	-.1	0	.1	.2	.4	.6	.8
X1	ICE	.938	.864	.739	.596	.534	.500	.501	.542	.683	.838	.929
	MIC	.842	.768	.675	.593	.565	.559	.570	.596	.686	.803	.907
	MICE	.838	.763	.653	.549	.516	.500	.521	.558	.668	.797	.906
X2	ICE	.957	.922	.821	.654	.564	.500	.466	.484	.639	.826	.937
	MIC	.787	.738	.661	.591	.571	.562	.572	.595	.690	.805	.908
	MICE	.780	.728	.632	.549	.512	.500	.515	.544	.671	.798	.905
X3	ICE	.931	.904	.832	.686	.598	.500	.412	.334	.277	.415	.677
	MIC	.719	.686	.634	.585	.568	.565	.576	.596	.681	.806	.910
	MICE	.707	.666	.599	.532	.505	.500	.513	.541	.651	.799	.908
X4	ICE	.936	.852	.687	.526	.492	.500	.554	.642	.818	.928	.966
	MIC	.852	.774	.675	.600	.576	.571	.579	.599	.661	.729	.780
	MICE	.847	.763	.648	.541	.514	.500	.512	.537	.626	.711	.770
X5	ICE	.955	.925	.836	.681	.577	.500	.448	.436	.555	.760	.898
	MIC	.764	.722	.656	.588	.568	.560	.571	.592	.685	.803	.902
	MICE	.756	.702	.621	.537	.504	.500	.508	.542	.659	.795	.899
X6	ICE	.954	.915	.810	.659	.575	.500	.428	.391	.450	.658	.852
	MIC	.759	.711	.649	.586	.571	.570	.575	.595	.686	.797	.896
	MICE	.748	.689	.613	.530	.502	.500	.510	.539	.657	.789	.893
X7	ICE	.938	.908	.817	.665	.576	.500	.437	.399	.471	.665	.818
	MIC	.724	.690	.638	.590	.574	.571	.578	.597	.669	.760	.827
	MICE	.710	.672	.600	.536	.507	.500	.508	.539	.639	.750	.822
X8	ICE	.917	.862	.752	.602	.537	.500	.491	.516	.646	.798	.896
	MIC	.754	.708	.643	.584	.568	.559	.564	.583	.655	.740	.810
	MICE	.745	.694	.614	.538	.515	.500	.512	.544	.628	.726	.804
X9	ICE	.850	.816	.747	.640	.570	.500	.442	.397	.374	.470	.643
	MIC	.697	.667	.622	.579	.570	.567	.573	.594	.662	.751	.820
	MICE	.678	.643	.587	.523	.511	.500	.510	.539	.631	.731	.812

Table 5: Estimated probabilities of correctly choosing AR(1) disturbances based on ICE, MIC and MICE when n = 50.

		ρ = -.8	-.6	-.4	-.2	-.1	0	.1	.2	.4	.6	.8
X1	ICE	.974	.820	.620	.488	.484	.500	.494	.502	.621	.821	.965
	MIC	.988	.868	.677	.521	.485	.472	.484	.521	.672	.854	.976
	MICE	.988	.868	.678	.532	.504	.500	.504	.532	.674	.854	.976
X2	ICE	.954	.754	.544	.444	.461	.500	.499	.484	.553	.754	.934
	MIC	.986	.867	.668	.517	.483	.470	.481	.513	.656	.834	.963
	MICE	.986	.867	.669	.527	.506	.500	.503	.524	.658	.835	.963
X3	ICE	.878	.613	.458	.391	.436	.500	.517	.490	.498	.669	.867
	MIC	.987	.864	.658	.518	.479	.464	.475	.514	.639	.815	.938
	MICE	.987	.864	.660	.532	.506	.500	.505	.531	.641	.815	.938
X4	ICE	.890	.688	.487	.446	.490	.500	.439	.407	.463	.668	.886
	MIC	.961	.831	.653	.517	.483	.466	.480	.514	.662	.849	.974
	MICE	.961	.831	.653	.530	.508	.500	.506	.529	.664	.849	.974
X5	ICE	.929	.706	.511	.425	.453	.500	.506	.491	.546	.733	.919
	MIC	.986	.862	.662	.514	.480	.465	.479	.513	.649	.822	.959
	MICE	.986	.862	.663	.526	.512	.500	.509	.530	.652	.822	.959
X6	ICE	.922	.695	.503	.419	.443	.500	.507	.488	.538	.732	.917
	MIC	.985	.862	.663	.518	.479	.470	.483	.513	.650	.824	.959
	MICE	.998	.862	.664	.530	.508	.500	.506	.527	.652	.824	.959
X7	ICE	.929	.737	.561	.458	.466	.500	.513	.510	.593	.784	.949
	MIC	.986	.865	.666	.521	.485	.473	.488	.524	.653	.844	.969
	MICE	.986	.865	.667	.530	.503	.500	.508	.534	.654	.844	.969
X8	ICE	.895	.700	.540	.459	.477	.500	.506	.505	.572	.753	.915
	MIC	.974	.842	.653	.517	.480	.468	.482	.515	.640	.822	.953
	MICE	.974	.842	.655	.526	.508	.500	.510	.531	.642	.822	.953
X9	ICE	.909	.674	.517	.424	.453	.500	.525	.510	.540	.717	.908
	MIC	.985	.858	.658	.514	.480	.468	.478	.509	.640	.823	.959
	MICE	.985	.858	.660	.528	.507	.500	.507	.524	.6442	.823	.959

Table 6: Estimated probabilities of correctly choosing MA(1) disturbances based on ICE, MIC and MICE when n = 50.

	γ =	−.8	−.6	−.4	−.2	−.1	0	.1	.2	.4	.6	.8
X1	ICE	.986	.916	.766	.611	.537	.500	.532	.593	.766	.924	.989
	MIC	.973	.884	.719	.580	.543	.528	.541	.582	.725	.898	.985
	MICE	.973	.884	.718	.567	.521	.500	.521	.572	.724	.898	.985
X2	ICE	.987	.915	.791	.642	.557	.500	.523	.599	.801	.943	.993
	MIC	.963	.873	.709	.580	.542	.530	.543	.583	.723	.897	.983
	MICE	.963	.873	.708	.569	.518	.500	.521	.570	.722	.897	.983
X3	ICE	.994	.939	.837	.685	.580	.500	.500	.578	.826	.958	.995
	MIC	.944	.855	.701	.579	.547	.536	.547	.584	.721	.896	.984
	MICE	.944	.855	.699	.564	.517	.500	.515	.569	.721	.896	.984
X4	ICE	.992	.954	.839	.636	.533	.500	.578	.667	.857	.955	.994
	MIC	.977	.890	.728	.585	.546	.534	.548	.585	.719	.871	.964
	MICE	.977	.890	.727	.569	.520	.500	.521	.569	.716	.871	.964
X5	ICE	.991	.924	.811	.662	.569	.500	.510	.585	.800	.940	.993
	MIC	.956	.864	.707	.580	.545	.535	.546	.586	.727	.896	.983
	MICE	.956	.864	.705	.562	.511	.500	.517	.569	.726	.896	.983
X6	ICE	.991	.933	.817	.668	.581	.500	.511	.595	.802	.945	.992
	MIC	.955	.858	.707	.579	.543	.530	.543	.584	.723	.888	.979
	MICE	.955	.858	.705	.566	.514	.500	.520	.569	.722	.888	.979
X7	ICE	.987	.890	.768	.628	.554	.500	.511	.580	.768	.921	.986
	MIC	.953	.862	.705	.580	.541	.527	.541	.586	.723	.887	.9976
	MICE	.953	.862	.704	.570	.523	.500	.522	.574	.721	.886	.976
X8	ICE	.988	.917	.774	.619	.542	.500	.515	.582	.765	.920	.986
	MIC	.956	.860	.702	.575	.542	.532	.542	.581	.710	.878	.973
	MICE	.956	.860	.698	.562	.514	.500	.516	.563	.709	.878	.972
X9	ICE	.989	.908	.796	.655	.565	.500	.496	.577	.792	.935	.989
	MIC	.945	.853	.704	.581	.542	.532	.545	.584	.724	.886	.972
	MICE	.945	.853	.702	.564	.516	.500	.516	.567	.721	.885	.972



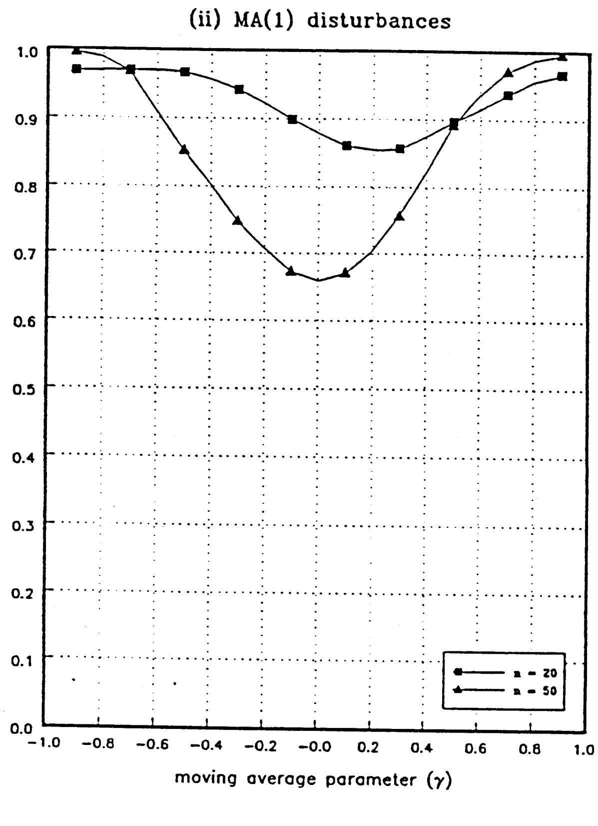


Figure 1: estimated probabilities of correctly selecting (i) AR(1) disturbances and (ii) MA(1) disturbances, using IC, for the regression model with design matrix X9

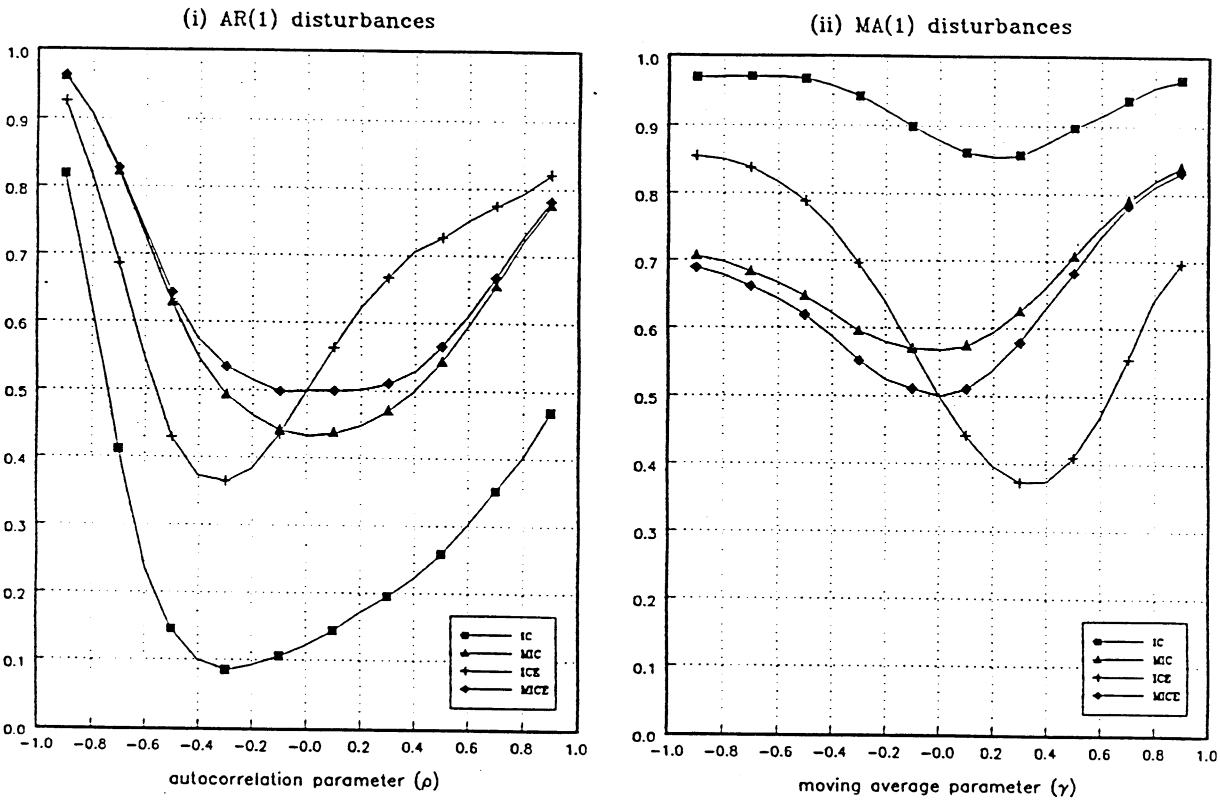


Figure 2: estimated probabilities of correctly selecting
(i) AR(1) disturbances and (ii) MA(1) disturbances,
for the regression model with design matrix X9, 20 observations.

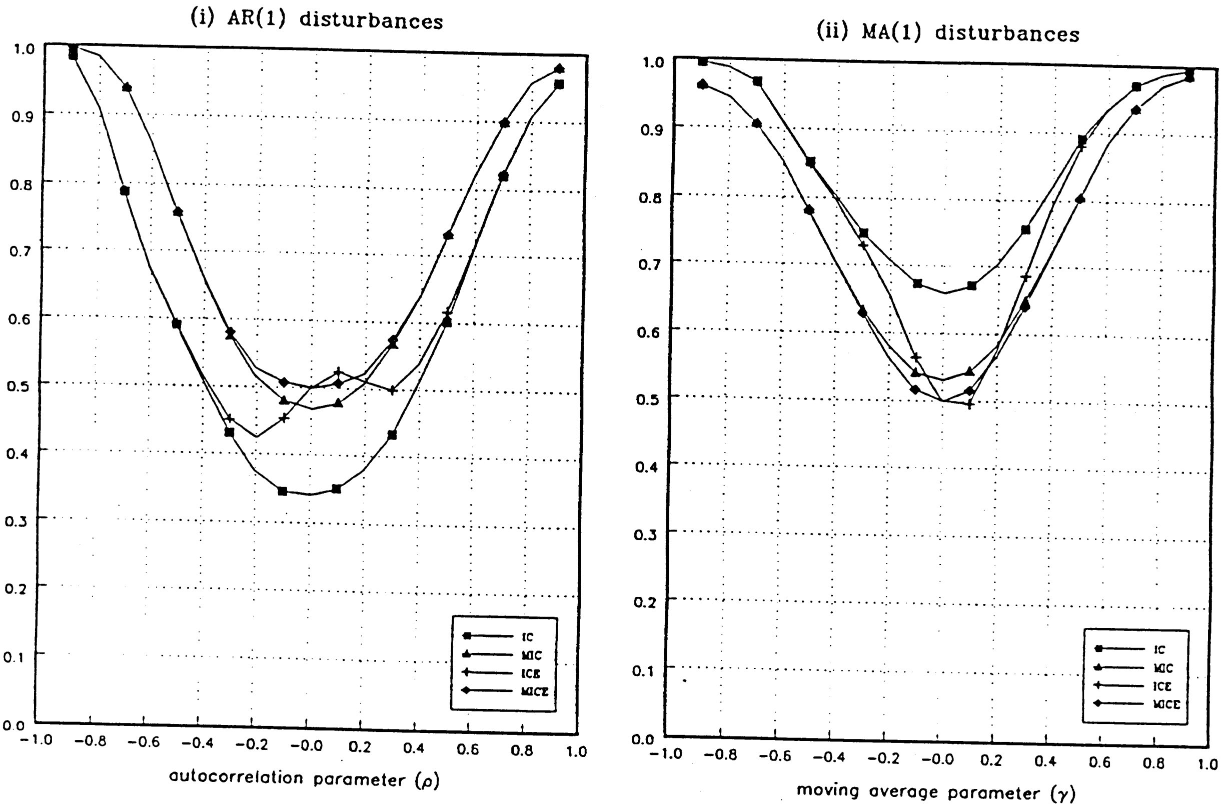


Figure 3: estimated probabilities of correctly selecting
(i) AR(1) disturbances and (ii) MA(1) disturbances,
for the regression model with design matrix X9, 50 observations.