

**ORDER DETERMINATION BY TABLE METHODS:
SOME MONTE CARLO EVIDENCE**

Robert KUNST

Forschungsbericht/
Research Memorandum No. 239
June 1987

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich des Autors und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden.

1. Introduction

As is known from time series theory, any stationary process (y_t) satisfying general regularity conditions may be approximated parsimoniously by ARMA models

$$\Phi(L)y_t = \theta(L)e_t \quad \text{or} \quad y_t = \Phi^{-1}(L)\theta(L)e_t$$

with (e_t) being a white noise process and $\Phi(\cdot)$ and $\theta(\cdot)$ polynomials in the lag operator L with all θ zeros outside or on and all Φ zeros outside the unit circle. Moreover, certain non-stationary processes, called "integrated processes", may be represented by the same formula, if $\Phi(\cdot)$ is allowed to have unit roots. These integrated processes describe economic time series quite well and are therefore often used for econometric purposes, including forecasting and seasonal adjustment.

The first step towards ARMA model estimation usually involves determination of the polynomial orders in Φ and θ , henceforth denoted p and q , respectively. Although it might be convenient to fix p and q ad hoc, this could prove extremely inefficient and should be avoided. Rather, maximal orders for an order determination procedure should be set, mostly depending on the number of observations. Popular suggestions are sublinear functions of the sample size, like e.g. the square root. Within this paper, the maximal orders were set to eight. This is a common selection in econometrics if quarterly data are used. Higher orders would have entailed excessively-sized and consequently expensive arrays in the mainframe.

Two classes of approaches for order determination play a role in applied research. The first class relies on information criteria, like the AIC of Akaike(1974). These information criteria, introducing a penalty term for increasing parameter dimension, have been the subject of many publications, so that Monte Carlo (small

sample) as well as theoretical results are available. The subclass containing AIC and FPE does quite well for medium-sized samples but can be proved to over-estimate orders asymptotically. The more stringent BIC family is consistent but tends to under-estimate the true order in finite samples. For smaller samples, the performance of all information criteria remains unsatisfactory (see e.g. Nickelsburg (1985)).

The second class contains what will be called "Table Methods" and extends the less mechanical and more optical analysis of Box and Jenkins(1976) to mixed models which are difficult to handle with their original procedures ACF and PACF¹. Of course, all table methods only use the information inherent in the original ACF, unless they are designed explicitly for the integrated case. Consequently, the advantage of table methods over ACF merely lies in a re-arrangement of information that makes the indicative patterns more telling. At the beginning, four different table methods were to form the basis of the present study: the GPAC of Woodward and Gray(1981) including the original S-array of Gray, Kelley, McIntire(1978); the corner method of Beguin, Gourieroux, Monfort (1980); the extended autocorrelation function EACF of Tsay and Tiao (1984a); the SCAN by Tsay and Tiao(1984b). Unfortunately the first candidate had to be dropped from the list, the reasons to be outlined in section 2. All of these procedures were proved to be weakly consistent by the respective authors.

¹ Within this paper, no distinction will be made between theoretical functions and their sample counterparts. Accordingly, the acronyms SACF etc. are not used and ESACF is simplified to EACF. In most cases, it is easy to see what is meant.

2. The table methods

Historically, the oldest of the methods stems from Gray, Kelley, McIntire(1978, GKM). It involves the calculation of two data-dependent tables, called the R- and S-array. The S-array is non-rectangular. For ARMA processes, certain parts of columns and rows of these arrays can be shown to be constant, zero, or infinite. Although the authors state that their procedure is better than information criteria they concede that this might not be true for small samples. To test for constancy and zero-ness of rows and columns a matrix of "D statistics" is provided, the maximal element indicating ARMA orders. A program listing is contained in GKM which makes it easy to check the performance. It evolved from experiments preliminary to this paper: that the D table is unable to identify patterns which are recognized by the eye; that both the D table and the eye tend to over-estimate true orders, mostly identifying complicated mixed models; that auxiliary arrays of much higher dimension than stated in the GKM program comments must be used to obtain sensible results; that only very large samples contain enough information to construct such arrays. Additionally, the procedure is unable to discriminate between MA and mixed models by construction. The refined GPAC by Woodward and Gray (1981) condenses the information of the S-array and is only reliable together with the original arrays, according to authors. In short, the methods of this type might be helpful if used with skill but cannot be integrated into this Monte Carlo simulation.

A different track was followed by Beguin, Gourieroux, Monfort (1980, BGM) who presented the corner method. Autocorrelation estimates r_i are arranged in matrices S_{ij} of dimension j

$$S_{ij} = \begin{bmatrix} r_i & r_{i-1} & \dots & r_{i-j+1} \\ r_{i+1} & r_i & \dots & r_{i-j+2} \\ \dots & \dots & \dots & \dots \\ r_{i+j-1} & \dots & \dots & r_i \end{bmatrix}$$

For ARMA(p,q) models, theoretical S_{ij} can be shown to be singular for $i \geq p$ and $j \geq q$. In a rectangular array of the determinants of S_{ij} this condition generates rectangular sub-matrices of zeros. When constructing statistical tests about the hypotheses of zero determinants, BGM run into troubles. The distribution of determinants depends on minor determinants which makes calculation time-consuming. Moreover, the suggestion in BGM involves inversion of (exactly) singular matrices. Replacing the inverses by pseudo-inverses is possible but did not result in satisfactory performance when tried for this study. The full BGM procedure, including calculation of co-factors and pseudo-inverses, more than doubled computer time of the competing routines.

For the sake of this paper's Monte Carlo, a simpler approximation to the determinantal distribution was pursued. Empirical r_i fractiles may be determined from the formula $2/\sqrt{n-i}$ which is also used for the EACF (see below). If the decay of the ACF is smooth and monotonous, the $(r_i)^j$ constitutes the primary part of the S_{ij} determinant. Thus, $(2/\sqrt{n-i})^j$ is a thumb rule tending to underestimate the comparable fractiles. An ARMA(2,1) process would produce a table not like in Table 1a, but rather like e.g. in Table 1b. ('X' at (i,j) denotes determinants greater than the thumb rule value, '0' denotes smaller determinants)

TABLE 1: Exemplary patterns generated by ARMA(2,1) processes

a) theoretical corner method output according to BGM

```
X X X X X
X X X X X
X 0 0 0 0
X 0 0 0 0
X 0 0 0 0
```

b) possible COR output

```
X X 0 X X
X X X X 0
X 0 0 0 0
X X X X 0
X 0 0 X 0
```

c) theoretical EACF output.

```
X X X X X
X X X X X
X 0 0 0 0
X X 0 0 0
X X X 0 0
```

This property was found by preliminary simulation. It follows that p and q were determined by looking for the rows with zeros to the right and X s to the left. The most parsimonious corner place (i,j) was then selected. This rule looks ad hoc but we shall see that it works rather well. The procedure set out here will be denoted by COR.

The most popular table method by now is the extended autocorrelation function (EACF) by Tsay and Tiao(1984a) which has been included into some econometric software packages (e.g. IAS-SYSTEM). It relies on the iterative ARMA estimation procedure set out in Tiao and Tsay (1983) which is also consistent for "integrated" unit root models. This model class extension necessitates basing the procedure on least squares estimation. This, in turn, entails heavy computational burden relative to Yule-Walker methods like COR (or the original ACF/PACF). Like COR, EACF produces a rectangular table, which should look like Table 1c in the case of ARMA(2,1) models. This coded EACF table is derived from a correlation table including the ACF in the zero row which is then subjected to the decision rule $2/\sqrt{n-i-j}$. For all details see Tsay and Tiao (1984a).

In Tsay and Tiao (1984b), the same authors presented an alleged improvement over EACF, labeled SCAN. The smallest eigenvalue of a canonical matrix (therefore SCAN) is used to detect whether the matrix is singular. From the eigenvalue and a corresponding eigenvector, a statistic may be calculated which is asymptotically distributed chi-square. Since SCAN is the most complicated of the methods, it has to prove whether additional precision justifies additional computer time. However, it is transparent that it is the only procedure using an asymptotically exact distribution which could help a lot to establish its leading position.

3. The simulation design

Monte Carlo experiments in time series analysis are quite expensive as the investigator has to cope with $3+p+q$ dimensions, not counting the number of replications. High numbers of replications make the outcomes more accurate. However, the philosophy of this paper rather is to survey the general performance and to keep the number of replications low. For most experiments, this was set to 50 or 100. More computer time might change the picture within certain experiments but certainly not the overall impression.

Simulations dimensions were:

- the sample size. This is a small to medium sample study. Sample sizes of 100 and 200 are common in econometrics. Lesser samples do not produce useful results. Larger samples were only investigated where asymptotics looked interesting by comparing 100 and 200 observations behaviour.
- the AR and MA order. The fundamental tables allowed for all specifications up to ARMA(7,7) to be identified. In certain rare cases, nothing was identified which means that the identified orders p or q should be greater than 7. Simulation experiments were restricted to ARMA(p,q) with p and q less or equal to 2. It is to be suspected that this covers all the interesting cases, including white noise, pure AR and MA, and top- as well as bottom-heavy ARMA.
- the corresponding roots, outside the unit circle for the stationary case. For AR(1), the first-order parameter was changed between 0.3 and 1.0 - the latter case being a random walk - to obtain insight into reaction to root size. For MA(1) and ARMA(1,1), roots were fixed to typical values. For the rest, the technique of "random roots" was pursued. Since generated roots values were added to the output, reaction to root size may be seen also in those cases.

The technique of random roots relies on subjecting the inverse ARMA roots to a random processor. First, the uniform distribution over the interval $(-1,1)$ was tried (RRu). This specification includes the area around zero, where the ARMA(p,q) process degenerates to lower orders in finite samples. Therefore, the β distribution was selected with parameters 2.0 and 3.0 and mirrored at 0 (RR β). This law describes a sharp decline at -1.0 and 1.0 and a parabolically flat area around 0. This technique still leaves the possibility of degenerating orders by cancelling factors and equal roots. This event is more difficult to exclude and will account for some misspecifications.

In some cases, a parallel experiment was conducted with the same β law but with restricting AR and MA roots to different signs. This makes canceling impossible but might reduce the model class artificially.

4. Simulation results

The results, given in detailed form in Table 2, will be commented one by one since each experiment has its own peculiarities. The expression "shadow model" will be used for the most frequent misspecification which, in some cases, is more frequent than the true model.

a) white noise

For data to be identified as white noise, EACF demands for an upper triangle of 36 zeros; SCAN needs a square of 64 zeros; and COR needs the first row of 8 elements to be zero. Obviously, this task is easiest for COR which has the highest power (76 % even for $T=100$) whereas one single outlying "X" in 64 destroys the judgment of SCAN. If this is taken into account as well as the existence of special tests for the hypothesis of serial non-correlation, the performance of all procedures is satisfactory and comparable.

b) the AR(1) process

Remember that COR has not been designed for the ARIMA case. Therefore, COR power is spuriously high for the AR parameter approaching the border. EACF has disappointingly low power for low autocorrelation. Because of the special asymmetric shape of the criterion pattern - a triangle - EACF tends to see MA processes in AR ones. Even for $T=200$, power remains low but the estimated MA order is now increased to 2 as the significance of the second-order autocorrelation increases. On the other hand, SCAN performs equally well for all parametrizations including $RR\beta$ which accounts for negative roots. The gain by increasing sample size is low. COR performs surprisingly well but remains second to SCAN in summary.

Compare the plot in Figure 1 for this experiment. Whereas EACF

power increases almost linearly, SCAN and COR seem to have turning points and rather flat areas.

c) the MA(1) process

Since EACF is known to frequently detect MA processes spuriously, high power was expected. The results show, however, that all procedures generate similar exactitude, SCAN gaining pole position with increasing sample size. For $T=100$, only SCAN makes the reverse mistake of taking MA for AR(1) whereas for all other cases and procedures ARMA(1,1) is the most frequent alternative which, integrating the true model, is an innocuous misspecification. This means that, for $T=100$, EACF and COR are to be preferred in this case.

d) the ARMA(1,1) process

This process

$$X_t = 0.5 X_{t-1} + e_t + 0.5 e_{t-1}$$

is "close" to AR(2) and MA(2) models. Consequently, these are the most frequent mistakes, EACF and COR preferring MA. COR dominates SCAN for $T=100$, but the order is reversed for $T=200$. Interestingly, also the absolute power of COR decreases. EACF does not suggest itself as the procedure to be used.

e) the AR(2) model

The results confirm the tendency in EACF to identify MA models, in this case at least ARMA(1,1), so often that the procedure becomes useless. COR is second to SCAN and surprises by offering no distinct "shadow model". COR performs less well if two roots are approaching the unit circle where it identifies AR(1). This behaviour corresponds to that of all "classical" time series estimates.

f) the ARMA(2,1) model

With three parameters or roots, the situation becomes more complicated, especially because of canceling roots. As a means of control, the RR β d specification was introduced where AR roots and the MA root were restricted to have different signs.

In the general RR β design, all procedures have very low power. The obvious alternative for canceling roots, the AR(1) model, is selected frequently and, at least by SCAN, more so than the simulated ARMA(2,1). COR is unable to identify the additional MA part and prefers pure AR models so often that it becomes useless. COR also faces difficulties for near-unit roots. In the RR β d design, SCAN power increases considerably. However, SCAN generates a shadow model ARMA(1,1) that dominates the true model for smaller samples. On the other hand, EACF performance deteriorates in the RR β d case to a score quota of only about 10%. The remaining 90% are distributed equally among all kinds of alternatives. In summary, although SCAN remains the procedure most able to cope with the situation it is dangerously sensitive to canceling roots.

g) the ARMA(2,2) model

As far as the number of parameters is concerned, this is the most difficult setup investigated here. As with ARMA(2,1), there is a chance for canceling roots. Therefore, additional to the standard RR β design, the RR β d design was used. The results show that neither of the procedures is able to identify ARMA(2,2) correctly, even with RR β d and 200 observations. EACF hit the correct model most frequently, however showing the flattest distribution without any marked "shadow model". SCAN and COR, on the other hand, persistently identify (1,1) if canceling roots are allowed, or (2,1) and (2,0), respectively, in the RR β d case.

For the larger sample of 200, no improvement is visible. Merely, the sample distribution becomes more pronounced. The number of smaller models (p,q) with p and q less or equal to 2 decreases in favor of over-parametrized specifications. It may be deduced that

SCAN and COR face difficulties with identifying MA roots in the mixed models. It would take much larger samples to obtain satisfactory results.

h) the ARMA(1,2) process

Even though there is one parameter less in the model than in the preceding experiment, again the procedures fail to convince the investigator. Expectedly, the reduced model (0,1) shows up quite frequently and mirrors the cases of nearly canceling roots. Its frequency slowly but monotonously decreases if sample size is increased. As in the preceding experiment, the procedures do not gain from excluding the possibility of canceling. The shadow model changes to ARMA(1,1) with SCAN and COR but not with EACF which again suffers from flat sample distributions.

Again, the gain from enlarging the sample is small. The probability of smaller models (p,q) with p equal to 0 or 1 and q less or equal to 2 decreases with EACF and COR from about 70% to less than 50% in $RR\beta$. However, this does not result in marked improvements of correctly identified models. SCAN is different from its competitors in giving more weight to ARMA(1,1) even in $RR\beta$ (about 20%). This means that its low performance is due to the failure to recognize a second MA root whereas, for EACF and COR, the trouble is rather over-parametrization due to spurious "X" in the tables.

Since numerical values of the roots were saved, it might be interesting to see which models are most likely to be recognized. For $T=200$ and $RR\beta$, these are given in Table 3. Surprisingly, none of them belongs to the $RR\beta d$ case. The vast majority of AR roots is negative, and most of the MA roots pairs have alternating signs. Obviously, this design where the first order MA correlation is lower than the 2nd order correlation makes pattern recognition easy whereas large 1st correlations in the residuals from AR processes produce spuriously high values in the tables.

In preliminary summary, the performance with regard to top- and bottom-heavy ARMA is far from symmetric leaving the top-heavy to

balanced case as the most difficult one. To put it the other way round, top-heavy models are likely to be much more frequent in practice than it would seem by applying table methods.

i) the MA(2) model

This should be an easier task for the procedures, the principal mistake to be expected being inability to detect the second root, i.e. an ARMA(0,1) process. The power of EACF is higher than that of the competing methods, all three of them gaining from extension of the sample. The vast majority of correctly identified models have roots with alternating signs. The effect of equal signs can be anything: sometimes MA(1) is selected, but sometimes spuriously over-parametrized models result. The probability of smaller models (including white noise, MA(1) and MA(2)) decreases with extended sample size for EACF and COR but increases for SCAN. In summary, EACF seems to be the best here, but compare the outcomes for AR(2) that will explain the obvious bias of this procedure.

4. Summary and Conclusions

50 replications are not enough to use the results of Table 2 as statistical significance bounds. However, some important conclusions are evident.

First, all table procedures show a notorious lack of power in the top-heavy to balanced case. Pure AR and MA models are the easiest ones but bottom-heavy models may also be treated safely if enough observations are available, canceling roots are excluded and SCAN is used. No solution can be offered to identify the correct number of MA roots in mixed models. The specifications resulting from table methods should always be checked for the possibility of additional MA roots.

Second, SCAN shows the best overall performance. It loses some ground with MA models or mixed models where the MA part dominates but it well justifies the additional computer time. On the other hand, EACF shows remarkably low power for AR or bottom-heavy models. If a quick routine is needed, COR has its merits which are surprising the more that its design was rather ad hoc. Of course, it is invalid for ARIMA cases². Compare the "total sum" given in Table 2b which ranks SCAN first with COR second and EACF third.

Tsay and Tiao (1984a) surmise that the "thumb rule" might underestimate true standard deviation and suggest experiments to fix optimal values. From this paper's simulations, it can be deduced that simple modifications of these numbers will not suffice for considerable improvements, at least not with EACF. In many cases, EACF tables reproduced the asymptotic pattern quite well, with the exception of some few scattered 'X' or vertical bars of significant values. If this overall pattern shows up, the obvious lower-order models should always be tried and compared.

² For example, COR identifies the Box/Jenkins series C as AR(1) whereas EACF and SCAN safely detect a second AR root.

Significance bounds of all procedures were put at around 5%. This means that the asymptotic probability of a spurious 'X' should be 5%. Of course, since separate matrix elements are not independent, the "size" of the procedures is lower than $1-.95^n$ with n denoting the number of zero elements. The results indicate that it rather amounts to $1-.95^m$ with m denoting the number of zeros in the (p,q) row. For example, in the white noise design, SCAN size is rather $1-.95^9$ than $1-.95^{81}$. If these properties are taken into account, significance bounds may be lowered to reach e.g. $.95=(1-\alpha)^m$ for the overall procedure.

Acknowledgements

The author wants to thank Professor R.S. Tsay for providing test data for the SCAN procedure.

TABLE 2a: Approximate percentage of true model identification³

p	q	design	T	#r	EACF	SCAN	COR	EACF SCAN COR (shadow models)		
0	0		100	50	72	60	76	0,2	1,1	0,2
1	0	.2	100	100	0	18	10	0,0	0,0	0,0
1	0	.3	100	100	5	37	18	0,1	0,1	0,1
1	0	.3	200	100	9	51	17	0,2	1,1	1,1
1	0	.4	100	100	8	37	17	0,1	0,1	0,1
1	0	.5	100	100	25	53	24	0,2	0,1	0,1
1	0	.6	100	100	30	55	25	0,2	1,1	1,1
1	0	.7	100	100	38	54	37	1,1	1,1	1,1
1	0	.8	100	100	54	68	58	1,1	1,1	1,1
1	0	.9	100	100	63	66	81	1,1	1,1	2,0
1	0	1.0	100	100	70	70	98	1,2	1,1	
1	0	RR β	100	50	26	48	34	0,2	1,1	1,1
1	0	RRu	100	50	28	38	34	0,0	0,0	0,0
0	1	.5	100	50	56	56	56	1,1	1,0	1,1
0	1	.5	200	50	60	68	64	1,1	1,1	1,1
1	1	.5/.5	100	50	24	46	56	0,2	0,1	0,1
1	1	.5/.5	200	50	36	60	44	0,2	2,0	0,2
1	1	RR β	100	50	12	28	20	0,0	0,0	0,0
1	1	RR β	200	50	16	50	28	0,0	0,0	0,0
2	0	RR β	100	50	10	42	20	1,1	1,1	1,0
2	0	RR β	200	50	14	52	34	1,1	1,1	0,2
2	1	RR β	100	50	8	10	10	1,1	1,1	1,0
2	1	RR β	200	50	18	18	10	1,0	1,0	1,0
2	1	RR β d	100	50	10	24	2	1,2	1,1	2,0
2	1	RR β d	200	50	8	50	14	1,2	1,1	2,0
2	2	RR β	100	50	8	4	8	1,1	1,1	1,1
2	2	RR β	200	50	6	2	4	0,2	1,1	1,1
2	2	RR β d	100	50	6	2	0	2,1	2,1	2,0
2	2	RR β d	200	64	6	5	0	1,2	2,1	2,0

³ See note on next page.

TABLE 2b: Approximate percentage of true model identification (continued)⁴

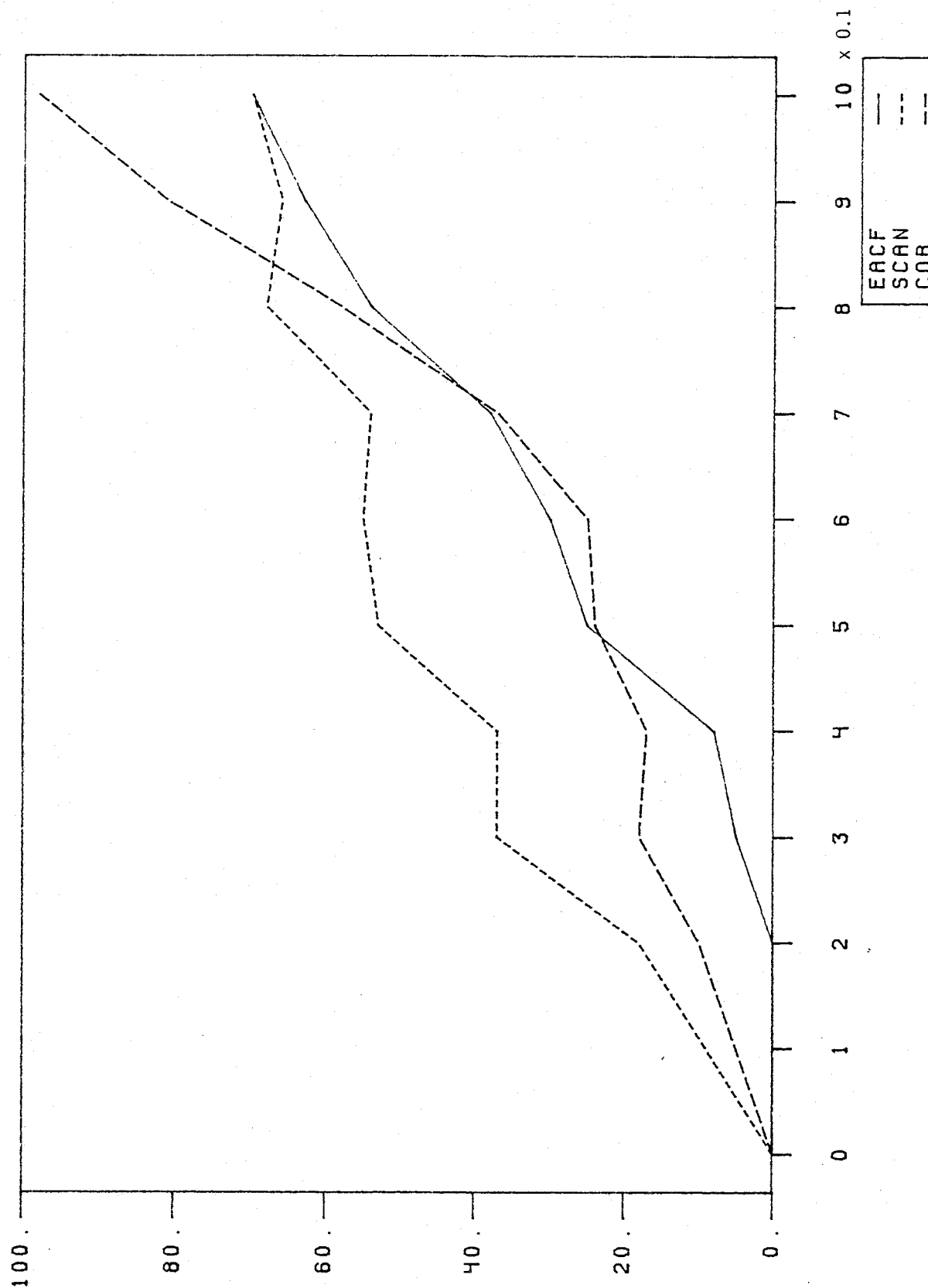
p	q	design	T	#r	EACF	SCAN	COR	EACF SCAN COR (shadow models)		
1	2	RR β	100	50	14	2	10	0,1	0,1	0,1
1	2	RR β	200	50	12	10	6	0,1	0,1	0,1
1	2	RR β	300	50	18	12	14	0,1	0,1	0,1
1	2	RR β d	100	50	14	2	4	0,4	1,1	1,1
1	2	RR β d	200	50	18	8	10	1,3	1,1	1,1
0	2	RR β	100	50	42	28	30	0,1	0,1	0,1
0	2	RR β	200	50	46	34	34	0,1	0,1	0,1
total sum				2314	25.8	38.4	34.0			

⁴ Note: p and q denote the true orders of the simulated ARMA(p,q) process; "design" gives the parameters in the cases of only one parameter or the RR designs described in the text otherwise; T is the sample size; #r is the number of replications; the next three columns give the percentages corresponding to the three table methods and the last three columns the most frequent incorrect model.

TABLE 3: ARMA(1,2) designs which led to recognition of the true model (AR root Φ and MA roots θ_1 and θ_2)

Φ	θ_1	θ_2	successful methods
-.65	.35	-.76	EACF
-.50	.44	.64	EACF
-.34	.19	-.88	SCAN
-.71	.66	-.49	EACF
-.61	-.55	.84	EACF
-.82	-.40	.58	EACF, COR
-.45	-.67	.79	SCAN
.87	-.34	-.42	SCAN
.75	.88	-.69	EACF, COR
.67	-.84	.70	SCAN
-.39	-.80	.67	SCAN
-.16	.70	-.40	COR

FIGURE 1: Percentage of true model identification for the AR(1) model for a sample size of 100 depending on the AR parameter



References

- Akaike, H. (1974): A new look at the statistical model identification; IEEE Transactions on Automatic Control AC-19,716-723.
- Beguín, J.M., C.S. Gouriéroux, A. Monfort (1980): Identification of a mixed autoregressive-moving average process: the corner method; in: O.D. Anderson (ed.): Time Series, pp. 423-436, North-Holland.
- Box, G.E.P. and G.M. Jenkins (1976): Time Series Analysis: Forecasting and Control, Holden-Day.
- Gray, H.L., G.D. Kelly, D.D. McIntire (1978): A new approach to ARMA modeling; Communications in Statistics B7,1-77.
- Nickelsburg, (1985): Small Sample Properties of Dimensionality Statistics for Fitting VAR Models to Aggregate Economic Data: A Monte Carlo Study; Journal of Econometrics 28,183-192.
- Tiao, G.C. and R.S. Tsay (1983): Consistency properties of least squares estimates of autoregressive parameters in ARMA models; Annals of Statistics 11,856-871.
- Tsay, R.S. and G.C. Tiao (1984a): Consistent estimates of autoregressive parameters and extended sample autocorrelation functions for stationary and nonstationary ARMA models; Journal of the American Statistical Association 79,84-96.
- Tsay, R.S. and G.C. Tiao (1984b): Use of Canonical Analysis in Time Series Model Identification; Mimeo.
- Woodward, W.A. and H.L. Gray (1981): On the Relationship between the S Array and the Box-Jenkins Method of ARMA Model Identification; Journal of the American Statistical Association 76, 579-587.