ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B, 00120 Helsinki 12, Finland, tel. 601322

Keskusteluaiheita
Discussion papers

Timo Teräsvirta and Ritva Luukkonen

CHOOSING BETWEEN LINEAR AND THRESHOLD AUTOREGRESSIVE MODELS*

No. 141 8 December 1983

*An earlier version of this paper was presented to the 11th International Time Series Meeting, Toronto, 18-21 August 1983. The first author acknowledges financial support from Societas Scientiarum Fennica.

CHOOSING BETWEEN LINEAR AND THRESHOLD AUTOREGRESSIVE MODELS

Timo Teräsvirta
Research Institute of Finnish Economy, Lönnrotinkatu 4 B, SF-00120 Helsinki 12, Finland

Ritva Luukkonen
Institute of Occupational Health, Haartmaninkatu 1, SF-00290 Helsinki 29, Finland

Tong and Lim (1980) advocate the use of threshold models in certain situations where the linear model is claimed to be inadequate. While a procedure for specifying the dynamics of the threshold model is suggested, no method is given for distinguishing between the linear and threshold models. In this paper, the performances of two model selection criteria, AIC and SBIC when used for this purpose, are investigated. The Monte Carlo experiments indicate that SBIC is clearly the best of these two alternatives.

1. INTRODUCTION

In a recent paper, Tong and Lim (1980) state that "the new era of *practical* non-linear time series modelling is, without doubt, long-overdue" (their italics). To accompany this claim, the authors introduce a family of non-linear models, called threshold autoregressive (TAR) models, and demonstrate their applicability to practical problems by examples.

Even if we accepted the invitation of Tong and Lim to enter the world of non-linear models, we might still at least sometimes be tempted to look back. The theory of linear models is as yet much better developed than that of non-linear models, and many problems are easier to deal with in the linear framework. Weighty arguments in favour of these more complicated models are needed before a model builder might be willing to abandon the more familiar territory of linearity. Therefore, an interesting question is how it could be found out when the true model is linear and not threshold autoregressive. In many cases the theory of the phenomenon to be modelled is not very helpful in this respect. If the TAR models are regarded as an alternative in a model building situation, there should be ways of choosing the right family of models on the basis of the evidence contained in the data.

Tong and Lim (1980) employ Akaike's Information Criterion (AIC) (Akaike, 1974) in the specification of TAR models. A natural idea would then be to apply AIC also to choosing between linear and TAR models. This has been proposed by Tong in an unpublished report (Tong, 1979). The Monte Carlo results of this paper demonstrate, however, that AIC is not a suitable criterion for that purpose. They are rather in favour of another criterion considered here, the Schwarz' Bayesian Information Criterion (SBIC), cf. Schwarz (1978). SBIC seems to be much more reliable than AIC in selecting the linear alternative when the true model is linear. It also yields consistent estimates of the dimension of the model. There are also other criteria than SBIC with the last-mentioned property; see e.g. Hannan and Quinn (1979) and Geweke and Meese (1981). In other to keep the exposition brief they are not considered here.

In this paper we first consider TAR models and their specification (Sections 2 and 3). The simulation experiment is described in Section 4 and its results discussed in Section 5. The final remarks are in Section 6.

2. THRESHOLD AUTOREGRESSIVE AND RELATED MODELS

The single-equation TAR model with J regimes can be written as

$$\phi^{(j)}(z)y_t = \beta'^{(j)}(z)x_t + \varepsilon_t^{(j)},\ t=1,2,\dots,T,\ E\varepsilon_t^{(j)} = 0,\ \mathrm{var}(\varepsilon_t^{(j)}) = \sigma_j^2,\ j=1,2$$

$$\text{if } c_{j-1} \leq u_{t-d} < c_j;\ c_0 = -\infty,\ c_J = \infty,\ d>0,\ j=1,\dots,J. \tag{2.1}$$

In (2.1), u_{t-d} is a random variable, d is the delay parameter, c_j's are fixed threshold values, z is a lag operator, $zy_t = y_{t-1}$, and

$$\phi^{(j)}(z) = 1 - \sum_{k=1}^{p_j} \phi_k^{(j)} z^k \ , \ j = 1,2, \text{ and}$$

$$\beta^{(j)}(z) = \sum_{k=0}^{r_j} \beta_k^{(j)} z^k \ , \ j = 1,2 \, .$$

Furthermore , $\mathrm{cov}\ (\varepsilon_t^{(j)}, \varepsilon_s^{(k)}) = 0, \ t \neq s, \ \forall j, k$.

If $u_t = x_t$, then (y_t, x_t) following (2.1) is called an open-loop threshold autoregressive system (TARSO). Model (2.1) is called a TARSO $(J,(p_1,r_1),\ldots,(p_J,r_J))$ model. If $\beta^{(j)}(z) \equiv 0$, $j = 1,\ldots,J$, and $u_t = y_t$, we have a self-exciting threshold autoregressive or SETAR$(J,p_1,\ldots,p_J)$ model, see Tong and Lim (1980). In the applications published so far, the number of regimes $J = 2$.

Consider the following model

$$y_t = D_t \sum_{k=1}^{r_1} \beta_k^{(1)} x_{kt} + (1 - D_t) \sum_{k=1}^{r_2} \beta_k^{(2)} x_{kt} + D_t \varepsilon_t^{(1)} + (1-D_t)\varepsilon_t^{(2)}$$

$$E\varepsilon_t^{(j)} = 0, \ \mathrm{var}(\varepsilon_t^{(j)}) = \sigma_j^2, \ \mathrm{cov}(\varepsilon_t^{(j)}, \varepsilon_s^{(k)}) = 0, \ t \neq s, \ j,k = 1,2 \qquad (2.2)$$

where $D_t = 0$ if $\sum_{j=1}^{h} \pi_j u_{j,t-d_j} \leq 0$, $d_j > 0$, $j = 1,\ldots,h$, and $D_t = 1$ otherwise. The above model has been discussed by Goldfeld and Quandt (1973), see also Quandt (1982). In order to be able to estimate the $\beta^{(j)}$'s, σ_1^2, σ_2^2 and D_t, $t = 1,\ldots,T$, the authors suggested that D_t be approximated by

$$D_t = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\sum_{j=1}^{h} \pi_j u_{j,t-d_j}} \exp\ \{-\eta^2/2\sigma^2\}\, d\eta \, . \qquad (2.3)$$

Now, a TARSO$(2,(0,r_1),(0,r_2))$ model is a special case of (2.2) where the x_{kt}'s are lags of the same variable x_t. Furthermore $x_t = u_{1,t}$, $\pi_1 = 1$ and $u_{2,t-d_2} \equiv -1$. In TARSO models, D_t is supposed to have a degenerate distribution, and π_2 is specified together with the lag structure of the model. Now, the parameters of (2.2) completed with (2.3) can be estimated by the maximum likelihood method, and the linearity of the model can be tested by a LR test, see Goldfeld and Quandt (1973). That is not possible if the approach of Tong and Lim (1980) is applied as the likelihood function does not meet the necessary regularity conditions.

3. SPECIFICATION OF THRESHOLD AUTOREGRESSIVE MODELS

An inherent feature of a TARSO model is the assumption that the orders of the lag polynomials are unknown a priori. A similar assumption is made by Box and Jenkins (1970) in their treatment of ARMA and transfer function models. Thus they have to be specified from the data, together with the threshold values and the delay parameter. Goldfeld and Quandt (1973) do not have this problem; if their model contains lags, the lag structure is assumed known.

Tong and Lim (1980) have proposed the use of AIC as the main specification criterion of threshold models. Consider a TARSO$(2,(p_1,r_1),(p_2,r_2))$ model and define a general model selection criterion as

$$MSC(p_j,r_j) = \ln \hat{\sigma}^2(p_j,r_j) + (p_j + r_j + 1)g(T_j) \tag{3.1}$$

where $\hat{\sigma}^2(p_j,r_j) = T_j^{-1} \sum_{t\varepsilon j} (\hat{\phi}^{(j)}(z)y_t - \hat{\beta}^{(j)}(z)x_t)^2$ and T_j is the efficient number of observations in regime j. Setting $g(T_j) = 2T_j^{-1}$ in (3.1) yields AIC whereas $g(T_j) = T_j^{-1} \ln T_j$ corresponds to SBIC. Let the threshold value c be fixed at the sample 100 q^{th} percentile ξ_q of x. Define

$$AIC(\hat{p}_j,\hat{r}_j) = \min_{0\le p_j \le P_j, 0\le r_j \le R_j} h_j AIC(p_j,r_j) \text{ , } j = 1,2 \text{ .}$$

where $h_j = T_j/T$, $j = 1,2$. Then

$$AIC(\xi_q) = AIC(\hat{p}_1,\hat{r}_1) + AIC(\hat{p}_2,\hat{r}_2) \tag{3.2}$$

is the AIC value of the whole model when the threshold lies at ξ_q. Tong and Lim choose $\Xi = \{\xi_{0.3}, \xi_{0.4}, \xi_{0.5}, \xi_{0.6}, \xi_{0.7}\}$ and compute (3.2) for all values of Ξ keeping the delay parameter d fixed. The model corresponding to the minimum of (3.2) over Ξ is selected among the models with the same delay parameter. The above exercise is repeated for other values of $d_i \varepsilon \mathcal{D} = \{d_1,\ldots,d_m\}$ and this way $d, c, p_1, r_1, p_2, r_2, \sigma_1^2$ and σ_2^2 will be estimated. In practice it seems that the delay parameter is only a minor problem in the specification. Quite often there is one obvious alternative suggesting itself so that the others are easily excluded.

The specification procedure outlined in Tong and Lim (1980) does not contain any proviso for the possibility that the true model is linear although the use of AIC was suggested in Tong (1979). Since this criterion is used for the specification of the threshold model it would be easy to compute its value also for a set of linear models and compare the results. The minimum value of this extended set of alternatives would then indicate the final model.

It is well-known that AIC is does not estimate the dimension of the model consistently when a sequence of nested models is considered. The asymptotic probability of choosing too large a model remains positive. This has been shown in connection with AR models (Shibata, 1976), ARMA models (Hannan, 1980) and finite distributed lag models (Geweke and Meese, 1981). The same is true for polynomial distributed lag models when the lag length is determined first and the degree of polynomial thereafter (Teräsvirta and Mellin, 1983). For a rather general treatment of the problem, see Kohn (1983). In TARSO models, the situation is more complicated; the alternatives do not necessarily form a sequence of nested hypotheses. We are able to show that if the alternatives are nested, then AIC has a tendency to overestimate the dimension of the model. In this context it means selecting a threshold model with a positive probability even asymptotically when the true model is linear. See also Section 4 and the appendix.

On the other hand, SBIC estimates the dimension consistently in all cases of nested models mentioned above. Nevertheless, the asymptotic properties may mean little in a customary application but the differences in the asymptotic behaviour of AIC and SBIC do motivate their small sample comparison. Besides, recent simulations with a finite distributed lag model and a polynomial distributed lag model indicate that SBIC can be superior to AIC also in small samples, see Geweke and Meese (1981) and Teräsvirta and Mellin (1983).

4. SIMULATION EXPERIMENT

In order to compare the performance of AIC and SBIC in testing the linearity of assumed threshold models, we have carried out a simulation study. TAR models are usually applied to situations where the output variable displays cyclical variation. This is of course a rather superficial observation; more discussion on the nature of this cyclical variation can be found in Tong and Lim (1980). However, a linear model with this property has been constructed here. It is

$$(1 - 0.8z)y_t = (1 + z + z^2)x_t + \varepsilon_t \tag{4.1}$$

where $\varepsilon_t \sim n(0,1)$, $\mathrm{cov}(\varepsilon_t,\varepsilon_s) = 0$, $s \neq t$. Furthermore

$$x_t = (1 - 0.5z)(1 - 0.9z^4)\zeta_t \tag{4.2}$$

where $\zeta_t \sim n(0,1)$, $\mathrm{cov}(\zeta_t,\zeta_s) = 0$, $s \neq t$. From (4.1) and (4.2) it is seen that the output variable contains cyclical variation. Since this is a preliminary study, we economised the computations. The whole spectrum of threshold models were not scanned through in search of the best non-linear alternative. The choice was first limited to the family

$$(1 - \phi_1^{(j)}z)y_t = \mu^{(j)} + (\beta_0^{(j)} + \beta_1^{(j)}z + \beta_2^{(j)}z^2)x_t + \varepsilon_t^{(j)}, \; E\varepsilon_t^{(j)} = 0 ,$$

$$\mathrm{var}(\varepsilon_t^{(j)}) = \sigma_j^2 , \; \mathrm{cov}(\varepsilon_t^{(j)},\varepsilon_s^{(k)}) = 0 , \; t \neq s, \; j,k = 1,2,$$

$$j = 1 \text{ if } x_{t-4} \leq c \; ; \; j = 2 \text{ if } x_{t-4} > c . \tag{4.3}$$

Note that (4.1) is nested in (4.3). Thus the whole specification procedure described in the preceding section was not carried out since $p_1 = p_2 = 1$ and $r_1 = r_2 = 2$ were fixed. A limited experiment showed that if d was considered unknown, it received value four in more than nine cases out of ten. It was then permanently given that value. The threshold was specified by using the set Ξ of sample percentiles as in Tong and Lim (1980), and the model with the smallest AIC or SBIC value was selected. This obviously favours the selection of the linear model. If a larger set of combinations (p_1,r_1,p_2,r_2,Ξ) had been checked instead of mere $(1,2,1,2,\Xi)$, there would have been a positive probability to find a TARSO model with a still lower AIC or SBIC value. On the other hand, the degrees of the rational distributed lag in (4.1) are assumed known and are not varied either during the experiment.

Suppose the true linear model is nested in the threshold model. Assume furthermore that the threshold is given in advance. Then it can be shown that asymptotically SBIC chooses the linear model with probability one. When AIC is applied to the same problem, the probability of choosing the threshold model remains positive as the number of observations increases. This is shown in the appendix. The result indicates that SBIC has an edge over AIC at least in large samples.

The effective sample sizes in the Monte Carlo experiment were 50,100,150 and 500, respectively. The number of trials in each experiment was 400.

5. RESULTS

Restricting the family of threshold models in advance as in (4.3) has the effect that both AIC and SBIC often yield the same threshold value. This is because the number of parameters in the penalty function remains unchanged throughout. Minor differences arise from the fact that as the penalty function of AIC contains the factor 2, this is replaced by ln T_j in SBIC. The distributions of the threshold values are in Figures 1 and 2. An interesting observation is that either the 30th or 70th percentile is each time chosen in more than half of the trials. This tendency is more pronounced when SBIC is used than if AIC is the criterion.

The results of the performance of the two model selection criteria are in Table 1. They demonstrate that AIC is not a reliable criterion for detecting linearity when the alternative is a TARSO model. Even for the largest sample size ($T = 500$), the relative frequency of erroneously choosing the TARSO model is 0.4. SBIC does not perform well either when the number of observations is small but is fairly satisfactory and clearly better than AIC already at $T = 100$. A tentative conclusion is that the model builder would be well advised to prefer SBIC to AIC in the specification of threshold models.

A logical question to ask is what are the reasons for the mediocre performance of the model selection criteria AIC and SBIC in small samples. Why is the fit improved so dramatically when two regimes are assumed while in reality there is only one? One conspicuous detail in the applications of Tong and Lim (1980) is that the residual variances of the piecewise linear models are quite different. At least in ecological applications, it might be feasible to think that the white noise driving the system would have constant power without switches according to the regime. If changes in the variance of the disturbances were not assumed, would this affect the detection of linearity in small samples?

To investigate this possibility we constructed a restricted TARSO (RTARSO) model by setting $\sigma_1^2 = \sigma_2^2$ in (4.3). The simulations were repeated by using the RTARSO model as the alternative instead of (4.3). The results are in Table 2. AIC still has a tendency of choosing the threshold model rather frequently, even in large samples. On the other hand, SBIC performs very well already when T = 50. When $T \geq 100$, the observed frequency of erroneously choosing a RTARSO model is zero. We may thus conclude that the error variance is a crucial parameter in the specification of TARSO models. Note, however, that assuming $\sigma_1^2 = \sigma_2^2$ affects the model selection criteria but not the estimation of the parameters of the threshold model.

The models in this experiment were nested. Therefore, we also computed the values of the likelihood ratio (LR) statistic using threshold values from the specification procedure. Pretending that the LR statistic has a large sample χ^2 distribution under H_0 (linearity), a theoretical 0.05 significance level was used to test this hypothesis. The LR test fared somewhat better than AIC but was inferior to SBIC in every experiment with one exception. It rejected the linear model in 38 per cent of all cases, when the alternative was (4.3) and T = 50. On the other hand, Goldfeld and Quandt (1973) reported that the LR test works reasonably well in their experiments in which the likelihood function satisfies the regularity conditions.

6. FINAL REMARKS

This paper has investigated checking linearity in connection with the specification of TARSO models using model selection criteria. Of course, we are not suggesting that other possible techniques do not exist. An ultimate practical check would be post-sample prediction which has been applied in some examples in the paper of Tong and Lim (1980). The technique for obtaining these predictions in TAR models has been discussed for instance in Tong (1982).

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19, 716-723.

Box, G.E.P. and G.M. Jenkins (1970). *Time series analysis, forecasting and control.* San Francisco: Holden-Day.

Geweke, J.F. and R. Meese (1981). Estimating regression models of finite but unknown order. *International Economic Review* 22, 55-70.

Goldfeld, S.M. and R.E. Quandt (1973), The estimation of structural shifts by switching regressions. *Annals of Economic and Social Measurement* 2, 475-485.

Hannan, E.J. and B.G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society* B 41, 190-195.

Hannan, E.J. (1980). The estimation of the order of an ARMA process. *Annals of Statistics* 8, 1071-1081.

Harvey, A.C. (1981). *The economic analysis of time series.* Deddington: Philip Allan.

Kohn, R. (1983). Estimation of minimal subset dimension. *Econometrica* 51, 367-376.

Quandt, R.E. (1982). Econometric disequilibrium models. *Econometric Reviews* 1, 1-63.

Rao, C.R. (1965). *Linear statistical inference and its applications.* New York: Wiley.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461-464.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* 63, 117-126.

Teräsvirta, T. and I. Mellin (1983). Estimation of polynomial distributed lag models. University of Helsinki, Department of Statistics, Research Report No. 41.

Tong, H. (1979). An introduction to threshold autoregressive models, unpublished manuscript.

Tong, H. (1982). A note on using threshold autoregressive models for multi-step-ahead prediction of cyclical data. *Journal of Time Series Analysis* 3, 137-140.

Tong, H. and K.S. Lim (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society* B 42, 245-292 (with Discussion).

APPENDIX. Asymptotic properties of AIC and SBIC when these criteria are applied to choosing between the linear and TAR models

Consider a TAR model (2.1) with $J = 2$ and $\phi^{(j)}(z) \equiv 1$, $j = 1,2$. It can be written in matrix form as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \quad E\varepsilon = 0, \ \mathrm{cov}(\varepsilon) = \mathrm{diag}(\sigma_1^2 I_{T_1}, \sigma_2^2 I_{T_2}) \tag{A1}$$

where observation (y_t, x_t) belongs to the first submodel if $x_{t-d} \leq c$ and otherwise to the second submodel. Let X_j be a $T_j \times p$ matrix, $\mathrm{rank}(X_j) = p$ and assume c fixed and known. When model selection criteria are used to distinguish between the linear and the TAR model the latter will be chosen if

$$\mathrm{MSC}(p+1) - \mathrm{MSC}(2p+2) \geq 0 . \tag{A2}$$

From (3.1) it follows that (A2) is equivalent to

$$\ln\hat{\sigma}^2 + (p+1)g(T) - \{h_1 \ln\hat{\sigma}_1^2 + h_2 \ln\hat{\sigma}_2^2 + (p+1)(h_1 g(T_1) + h_2 g(T_2))\}$$

$$= \ \ln\frac{\hat{\sigma}^2}{\hat{\sigma}_1^{2h_1}\hat{\sigma}_2^{2h_2}} + (p+1)\{g(T) - h_1 g(T_1) - h_2 g(T_2)\} \geq 0 \tag{A3}$$

where $\hat{\sigma}^2 = T^{-1}(y - Xb)'(y - Xb)$, $\hat{\sigma}_j^2 = T_j^{-1}(y_j - X_j b_j)'(y - X_j b_j)$,

$b = (X'X)^{-1}X'y$, $y = (y_1', y_2')'$, $X = \mathrm{diag}(X_1, X_2)$, $b_j = (X_j'X_j)^{-1}X_j'y_j$,

$h_j = T_j/T$, $j = 1,2$.

The logarithm in (A3) is 2/T times the log-likelihood ratio when the null hypothesis $\beta_1 = \beta_2$, $\sigma_1^2 = \sigma_2^2$ is tested against (A1). Thus, asymptotically under the null hypothesis,

$$2\ell(\hat{\sigma}^2,\hat{\sigma}_1^2,\sigma_2^2) = T \ln \hat{\sigma}^2 \hat{\sigma}_1^{-2h_1} \hat{\sigma}_2^{-2h_2} \sim \chi^2(p+1).$$

The large sample probability of choosing the TAR model when the true model is linear is then

$$q = \Pr\{2\ell(\hat{\sigma}^2,\hat{\sigma}_1^2,\hat{\sigma}_2^2) \geq T\,(p+1)(h_1 g(T_1) + h_2 g(T_2) - g(T))\} \qquad \text{(A4)}$$

Using a result in Rao (1965, p. 78 (iv)), (A4) can be approximated from above and we obtain

$$q \leq T^{-1}(h_1 g(T_1) + h_2 g(T_2) - g(T))^{-1}. \qquad \text{(A5)}$$

Consider SBIC so that $g(k) = k^{-1} \ln k$. Then the r.h.s. of (A5) converges to zero as $T \to \infty$. When AIC is employed for the model selection, the r.h.s. of the probability inequality in (A4) converges to $2(p+1)$ as $T \to \infty$. Thus the probability q remains positive even asymptotically. These asymptotic results are also valid for TARSO models containing lags of the independent variable if the threshold value is assumed known because the disturbances of the model are white noise, see e.g. Harvey (1981, p. 48-49). In the numerical example of this paper, $\lim_{T\to\infty} q$ for AIC reaches the value .075.

A guess based on that example could be that the corresponding value is even higher if the threshold is not fixed in advance. In fact, we repeated our simulations assuming c = 0 throughout. Then, for T = 500, the relative frequency of erroneously choosing the TAR model by using AIC was 0.148 which is much closer to the asymptotic value than the corresponding value in Table 1.

Table 1. The observed relative frequencies of choosing the TARSO model when the true model is (4.1)

Number of observations	Model selection criterion	
	AIC	SBIC
50	0.573	0.438
100	0.460	0.090
150	0.423	0.040
500	0.395	0.008

Table 2. The observed relative frequencies of choosing the RTARSO model when the true model is (4.1)

Number of observations	Model selection criterion	
	AIC	SBIC
50	0.355	0.023
100	0.263	0
150	0.283	0
500	0.238	0

Figure 1. The distribution of threshold values in the Monte Carlo experiment (400 trials) when AIC was used in the specification

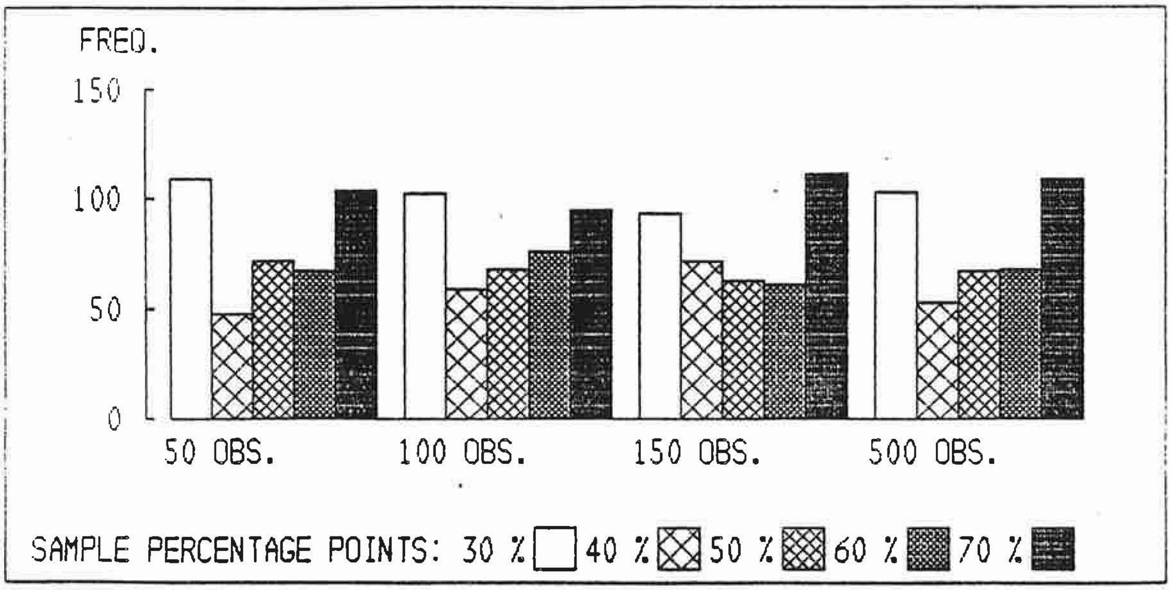

Figure 2. The distribution of threshold values in the Monte Carlo experiment (400 trials) when SBIC was used in the specification