

64935

**COMPUTER INTENSIVE TECHNIQUES FOR MODEL
SELECTION**

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ECONOMICS

AND THE THE INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Sıdıka Başçı

MAY, 1998

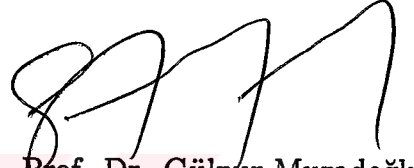
T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph.D.



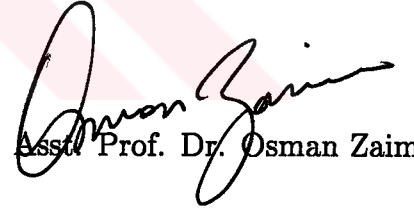
Prof. Dr. Asad Zaman

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph.D.



Assoc. Prof. Dr. Gülnur Muradoğlu

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph.D.



Asst. Prof. Dr. Osman Zaim

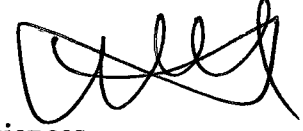
I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph.D.



Asst. Prof. Dr. Kivılcım Metin

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph.D.

Asst. Prof. Dr. Turan Erol



Approved by the Institute of Economics and Social Sciences

Director: Metin Heper



ABSTRACT

COMPUTER INTENSIVE TECHNIQUES FOR MODEL SELECTION

SIDIKA BAŞÇI

Ph.D Thesis in Economics

Supervisor: Prof. Dr. Asad Zaman

May 1998

There are three essays in this dissertation. In the first one, which appears in Chapter 2, a comparison of finite sample performances of six model selection criteria for Autoregressive (AR) processes exists. Simulation results report the effects of being parsimonious while selecting the model on forecasting. Moreover, in the chapter the assumption of normality, which can be seen in all of the previous theoretical and empirical studies, is relaxed and performances of the criteria under non-normal distributions are investigated. The second essay is presented in Chapter 3. In this essay three new model selection criteria are suggested where cross-validated estimates of variances are used. In the chapter, a comparison of the finite sample performances of these new criteria with the already existing ones is presented. The main concern of the third essay, that appears in Chapter 4, is detecting structural change when the change point is unknown. In the chapter, we derive some Bayesian tests to detect structural change with unknown change point under the assumptions of different prior distributions.

Key Words: Model selection, autoregressive processes, lag order determination, forecasting, cross-validation, structural change, unknown change point, Bayesian approach.

ÖZET

MODEL BELİRLENMESİ AMACINDA KULLANILAN BİLGİSAYAR YOĞUNLUKLU TEKNİKLER

SIDIKA BAŞÇI

Ekonomi Doktora Tezi,

Tez Yöneticisi: Prof. Dr. Asad Zaman

Mayıs 1998

Bu araştırmada üç makale yer almaktadır. 2. bölümde yer alan ilk makalede otoregresiv süreçlerde model belirlenmesi amacıyla kullanılan altı kriterin sonlu gözlem performansları karşılaştırılmaktadır. Simulasyon sonuçları model seçiminde cimri olmanın tahminler üzerindeki etkilerini rapor etmektedir. Ayrıca bölümde daha önceki teorik ve ampirik çalışmaların tamamında görülen normalite varsayımı kaldırılmakta ve normal olmayan dağılımlar altında kriterlerin performansları incelenmektedir. 3. bölümde ikinci makale sunulmaktadır. Bu makalede çapraz onaylama varyans tahminlerinin kullanıldığı üç yeni model belirleme kriteri önerilmektedir. Bölümde bu yeni kriterler ve halen kullanılmakta olan kriterlerin sonlu gözlem performans karşılaştırmaları yer almaktadır. 4. bölümde yer alan üçüncü makale değişim noktasının bilinmediği durumda yapısal değişimin yakalanmasıyla ilgilenmektedir. Bölümde, değişik ön dağılımlar varsayımları altında değişim noktasının bilinmediği durumlarda yapısal değişimin yakalanması için Bayesyen testler çıkartılmaktadır.

Anahtar Kelimeler: Model belirlenmesi, otoregresiv süreçler, gecikme sırası belirlenmesi, tahmin, çapraz onaylama, yapısal değişim, bilinmeyen değişim noktası, Bayesyen yaklaşım.

Acknowledgements

I would like to express my gratitude to Professor Asad Zaman for his valuable supervision. I am indebted very much to him for his suggestions during my thesis work and for his supports without which I could hardly proceed. Special thanks go to professors Gülnur Muradoğlu, Kıvılcım Metin, Osman Zaim and Turan Erol for their careful reading and valuable comments. I would also like to thank to Mehmet Orhan, Süheyla Özyıldırım, İsmail Sağlam and Murat Yülek for their encouragement.

Parts of this work has been presented at the 1996 Econometric Society European Meeting in İstanbul, 1997 Computing in Economics and Finance Conference in Stanford, California, and at seminars at Bilkent University. I wish to thank to the participants in these meetings.

My husband Erdem provided me a constant support during all my studies and gave strength to me at the times that I feel exhausted. I am especially thankful to him.

Contents

Abstract	iii
Özet	iv
Acknowledgements	v
Contents	vi
1 Introduction	1
2 A Comparison of Model Selection Criteria	8
2.1 Effects of Being Parsimonious on the Forecasting Performance of Model Selection Criteria	9
2.1.1 Criteria for AR Lag Order Selection	9
2.1.2 The Overestimation Problem	11
2.2 Effects of Skewness and Kurtosis on Model Selection Criteria	13
3 Variance Estimates and Model Selection	19
3.1 A Cross-Validated Estimate of Variance	21
3.2 An Alternative Cross-Validated Estimate of Variance	26
4 Alternative Bayesian Tests from Different Priors for Detecting Struc- tural Change with an Unknown Changepoint	33
4.1 The Rolling Chow Test	34
4.2 The Bayesian Approach	36
4.3 An Application	42
5 Conclusion	45

Bibliography	50
Appendix A	55



1 Introduction

In this dissertation there are three essays. The first one presents a comparison of various model selection criteria. In applied work, model selection is a frequently occurring problem of great importance, as forecasts, conclusions, interpretations, etc. depend critically on the particular model selected from the range of models examined. Most often, model selection is done by mechanical application of one or several of the criteria that have been developed for this purpose. Among these Akaike Information Criterion (AIC) (Akaike, 1973, 1974), Schwarz Criterion (SC) (Schwarz, 1978; Rissanen 1978), Hannan-Quinn Criterion (HQC) (Hannan and Quinn, 1979; Quinn, 1980), final prediction error (FPE) (Akaike, 1969, 1971), CAT (Parzen, 1974, 1977) and Shibata (Shibata, 1980) are the ones that are frequently used.

Some authors, such as Amemiya (1980), or Judge *et. al.* (1985), have argued against such mechanical model selection. Hendry (1995) argues in favor of systematic model simplification starting from a model complicated enough to nest all possibilities. In situations where forecasts are of interest, it is also possible to use forecast combination and avoid selection; see Zaman (1984) and Diebold (1989) for discussion and further references.

The first essay, which is chapter 2 of the dissertation, ignores these alternatives which have had small impact on practice so far and concentrates on some of the above stated model selection criteria and a few more. Finite sample performances of the criteria are investigated by Monte Carlo simulations. Specifically, we deal with the problem of estimating the lag order of Autoregressive (AR) processes since this constitutes a major problem area in model selection. Each of the above stated criteria can be used for this aim and it is possible that each of them may support some different model. Then comes the question of which one to use. There are a few

studies in the literature comparing the finite sample performance of these criteria. Among these Lütkepohl (1985) compares 12 different identification approaches for vector AR models in a Monte-Carlo study. In his study, SC and HQC emerge as being the best among the compared criteria. AIC also performs well. Koreisha and Pukkila (1993) augment Lütkepohl (1985) and show that the performance of the above mentioned criteria depends on the number of nonzero elements of matrices of the AR parameters and the maximum possible lag order that is used. Hurvich and Tsai (1989) makes a bias correction to the AIC for regression and AR models and name the new criterion as AIC_C . Hurvich and Tsai (1993) extend this work to vector AR models.

Prediction Sum of Squares (PRESS) introduced by Allen (1974) depends on the idea of using Cross-validation where the method is based on sequentially deleting points and recomputing the statistic of interest. The method of cross-validation was firstly introduced by Quenouille (1949) and named as jackknife. Methods like bootstrap are also closely related. Although the idea of these methods are very attractive, they are not simple in terms of computation so they could not be used in practice for a long time. After the recent advancements in the computer technology which gave rise to very fast computers, use of these methods became possible. For example, Hurvich and Tsai (1989) could include PRESS criterion in their Monte-Carlo study for the regression model.

One common observation about the studies mentioned above and also all the theoretical and empirical studies on model selection criteria is that they all assume normality while building the model. This assumption is used in Section 2.1 of this dissertation as well. Frequency distributions of the estimated lag order for six different criteria are presented in the section. Moreover, forecasting performances of

these criteria are also investigated. Since SC, HQC and AIC turn out to be the best criteria among the ones existing in the literature in Lütkepohl (1985), we include these in our Monte Carlo study. Other than these we search the performance of three additional criteria which are not considered in Lütkepohl (1985), AIC_C of Hurvich and Tasai (1989), PRESS of Allen (1974) and sequential F test. Simulation results show that a criterion being parsimonious, that is, choosing smaller models more frequently, means that the model chosen by that criterion will most probably lead to very good forecasts. This is because of the fact that a model chosen by setting the coefficient zero rather than putting an OLS estimate usually ends up with a better forecast since addition of variables almost always results with a higher variance of forecast (Allen, 1971, Walls and Weeks, 1969). From the simulation results we see that SC and HQC are relatively more parsimonious model selection criteria but PRESS has a tendency to select larger models. For this reason SC and HQC have smaller mean squared forecast errors compared to PRESS.

As mentioned above, the assumption of normality is used in all of the previous studies about model selection criteria. However some important time series, like those from the financial markets, violate this assumption. Usually skewness and kurtosis is observed in financial markets data. For this reason in Section 2.2 of this dissertation we consider the behavior of the model selection criteria in AR models where the error terms are not normal but there exists skewness and kurtosis. We still examine the performance of the six criteria considered in Section 2.1. The simulation results show that for both small and large samples skewness does not effect the performance of criteria under consideration. On the other hand, kurtosis does effect the performance of some of the criteria considerably. Moreover simulation results show that there is also a sample size effect on the performances of the model

selection criteria.

The large majority of the criteria used in the literature assess models using a function of the usual estimate of error variance and the model dimension. Different criteria are based on different functions, but all use the standard estimate of the variance $\hat{\sigma}^2$. The usual estimate is valid only if the model is correctly specified, and this assumption is especially dangerous in model search situations where we will inevitably search over incorrectly specified models. The second essay of this dissertation, which appears in Chapter 3, deals with this problem. Efron (1983) states that a cross-validated estimate of the variance is more robust to specification error and depending on this observation we expect that replacing the usual variance estimate by a cross-validated estimate should improve the performance of the model selection criteria.

As stated above, cross-validation method is based on sequentially deleting points and recomputing the statistic of interest. For the calculation of the function of PRESS, which is studied in Chapter 2, the method of cross-validation is used. There, only one observation is deleted in the process. It is also possible to delete some percentage of the observations sequentially and recompute the statistic of interest. In Chapter 3, in addition to PRESS, we consider deleting 10 percent of the observations and name the function that we obtain by this way as CV10. By using PRESS and CV10, we obtain two different cross-validated estimates of variance. In the chapter, we replace the usual estimate of the variance by these two forms of cross-validated estimates in the functions of the criteria AIC and AIC_C and see the effects of these replacements on model selection criteria within a Monte Carlo simulation study.

Modifying criteria such as AIC and AIC_C by replacing the usual estimate of variance by a cross-validated estimate of variance not only avoids the dangers that

arise from using usual estimate of variance, stated above, but also automatically puts a penalizing factor to the cross-validated estimates. From the simulation results of Chapter 2, we have the conclusion that PRESS is not parsimonious while selecting the model. The addition of penalizing factor as a result of the replacement solves this overestimation problem of PRESS mentioned in Chapter 2.

In Chapter 3, with a Monte Carlo study, we try to observe the improvements that we can obtain from using the modified versions of criteria, which use the cross-validated estimates of variance, over the already existing criteria. As in Chapter 2, we still consider the problem of estimating the lag order of AR models both with normal and non-normal error terms. We also compare the forecasting performances of the criteria. Simulation results show that the modified versions perform better most of the time, especially for large sample sizes.

In the third essay, which is Chapter 4 of the dissertation, we turn our attention to detecting structural change when the change point is unknown. Structural change is a well-known problem for applied econometricians, with serious possible consequences for model performance and forecasting. Assume that a model is selected by using one of the criteria discussed above. If the data possesses a structural change at some point then the model chosen by the criteria may not have a good performance, for example in terms of forecasting. For this reason it is important to know from the start whether the data possesses a structural change or not.

Given a known and dated change in the economic environment (such as World War 2), one can use the Chow's (1960) F test or variants to assess its impact on regression models. Since the economic environment is constantly changing, and it is not always known for sure which of the changes, and with what timing, affect the performance of our regression models, it is important to have a test for structural

change that does not require the knowledge of the change point in advance. In response to this need, many tests for ‘structural stability’ have been developed. Prominent among these are the *CUSUM* and *CUSUM of squares* test of Brown, Durbin and Evans (1975) where recursive residuals are used, and the *fluctuation test* of Sen (1980) and Ploberger, Krämer and Kontrus (1989) where recursive estimates of parameters are used. Less fancy, but always popular among practitioners, has been the intuitive ‘rolling’ Chow test, which simply repeats the Chow test at all possible changepoints. While this test has intuitive appeal, it had been plagued by unknown asymptotic distribution, and lack of theoretical basis. Andrews (1990) used Wiener process methods to derive the asymptotic distribution of the statistic, and also showed that it has superior power to other tests considered in the literature.

For a simple changepoint problem Cobb (1978) showed the existence of an informative ancillary statistic. Conditioning on the ancillary improves traditional frequentist inference, sometimes substantially, as shown by Cobb. Finding and conditioning on the ancillary appears difficult in the complexity of the regression model changepoint problem. As an indirect method of utilizing the relevant ancillary information is using a Bayesian method for assessing and estimating structural change. There are various studies which use Bayesian approach for structural change problems, such as, Chin Choy and Broemeling (1980), Holbert(1973), Holbert and Broemeling (1977) and Ferreira (1975). Given some prior distributions for the parameters it may be possible to find the posterior distributions of these parameters and also posterior distribution of the change point.

What is lacking in the literature is a comparison of performance under different prior distributions. In Chapter 4 of this dissertation we try to see the effect of different prior distributions on the posterior distributions. Moreover using these posterior

distributions we suggest tests for detecting structural change. In the literature we can see the idea of using such tests (for example see Zellner (1987)) but the explicit form of the tests are not derived so it has not been possible to use Bayesian tests of structural change with an unknown change point in applied work. In Chapter 4, we present the tests explicitly after a series of calculations so that they are ready for use in applied work. Andrews, Lee and Ploberger (1996) defines the class of optimal tests for unknown structural change problems. Since the Bayesian tests belong to this class they outperform other tests of structural change.

The main reason of the unpopularity of the Bayesian tests in applied work is the difficulty associated with the tabulation of critical values since these tests do not have a known distribution. However, with new and advanced computers we can easily obtain Monte Carlo estimates of critical values. The applied econometricians, using the data in hand can find out the critical value specific to their data. As stated above, these tests belong to the optimal class of tests defined in Andrews, Lee and Ploberger (1996). For this reason the effort of calculating the critical value will add a lot to the quality of the results obtained from the study. To illustrate the validity of our claim we compare the Bayesian tests with the popular rolling Chow test for İstanbul Stock Exchange Market weekly data in Chapter 4. The results definitely support the use of Bayesian tests. Finally, in Chapter 5 we have some concluding remarks.

2 A Comparison of Model Selection Criteria

In this chapter we present finite sample performances of six different model selection criteria for estimating the lag order of the AR processes. In section 2.1, we assume normality for the model which is a common assumption in the literature. We present probability distributions of estimating the lag order for each criteria. These distributions give us an idea whether the criteria are parsimonious or not. A criterion being parsimonious means that the criterion chooses smaller models, that is, it estimates lower orders of lag as the true lag order most of the time. We also present the mean squared forecast errors of the models selected by each of the criteria. By this way, we can see the effects of being parsimonious on the forecasting performance of model selection criteria.

In section 2.2, we relax the assumption of normality. In the literature there is no study considering the finite sample performances of model selection criteria for non-normal error terms. This is a lacking because some of the data, especially financial market data, posses skewness and kurtosis. For this reason, for an applied econometrician, it is important to know how model selection criteria behave under different distributions. In section 2.2, we provide this information by studying the finite sample performance of six criteria for two different error distributions. Specifically, we consider the effects of skewness and kurtosis on model selection criteria for AR models. We present the simulation results on probability of estimating the true lag order and mean squared forecast errors of the models chosen by each criteria.

2.1 Effects of Being Parsimonious on the Forecasting Performance of Model Selection Criteria

2.1.1 Criteria for AR Lag Order Selection

Assume that a given set of data $\mathbf{X} = (X_0, X_1, \dots, X_T)'$ is generated by a stationary AR(p) process (allowing for nonzero mean a_0):

$$X_t = a_0 + a_1 X_{t-1} + \dots + a_p X_{t-p} + u_t$$

where the $a_i, i = 0, \dots, p$, are unknown parameters and u_t is an i.i.d error term with common distribution F . Our objective is to find p , the true unknown lag order. It is assumed that there is an a prior maximum order M , so that our estimate for p can be any integer between $0, 1, \dots, M$.

We first briefly review the model selection criteria that we study in this chapter. To describe the criteria, let $\hat{a}_0, \dots, \hat{a}_p$ be the OLS estimates of the parameters of the AR(p) model. Let $\hat{X}_t = \hat{a}_0 + \hat{a}_1 X_{t-1} + \dots + \hat{a}_p X_{t-p}$ be the OLS estimate of X_t . Define the usual variance estimate $\hat{\sigma}_p^2 = (T - (p + 1))^{-1} \sum_{t=p}^T (X_t - \hat{X}_t)^2$. The first criterion AIC was introduced by Akaike (1973), and has been a very popular criterion for model selection. Define $AIC(p) = \ln(\hat{\sigma}_p^2) + \frac{2p}{T}$, and estimate p to be that integer between $1, \dots, M$ which minimizes the criterion AIC . Shibata (1976) shows that AIC is not consistent but overestimate p asymptotically with positive probability if $M \geq p$. Zaman (1984) shows how to calculate the probability distribution of the estimate $\hat{p}$ produced by AIC. Shibata (1980) discusses an asymptotic efficiency property of AIC.

Hurvich and Tsai (1989) makes a bias correction to the AIC, defined by $AIC_C(p) = T \ln(\hat{\sigma}_p^2) + T[1 + (p/T)]/[1 - (p + 2)T]$. They show that it is asymptotically efficient if the true model is infinite dimensional. When the true model is finite dimensional

AIC_C chooses the true lag order most often compared to other asymptotically efficient criteria.

The Schwarz Criterion, $SC(p) = \ln(\hat{\sigma}_p^2) + p \ln(T)/T$ was introduced by Schwarz (1978) based on Bayesian reasoning. It has the advantage of being consistent over the AIC. This and the Hannan-Quinn criterion $HQC(p) = \ln(\hat{\sigma}_p^2) + 2p \ln(\ln(T))/T$ both perform well in the Lütkepohl (1985) study where finite sample performance of 12 different criteria are compared, and hence are included in our study.

In addition to the AIC_C which was not part of the Lütkepohl (1985) study, we introduce two other criteria which have not been studied in this context. The criterion PRESS based on the idea of Cross-validation was first introduced by Allen (1974). This is defined as $PRESS(p) = \sum_{t=1}^T (x_t - \tilde{x}_t)^2$ where $\tilde{x}_t$, $t = 1, \dots, T$, is the predicted value of x_t from an AR(p) model *after omitting the t -th observation in the process of estimation*.¹ The value of p minimizing PRESS is selected. Although, as an idea, it is very attractive to use PRESS as a model selection criterion, it has not been used for a long time in practice. The reason of this unpopularity is that computationally it is very hard to calculate PRESS. With the advancements in computer technology, it is now possible to use PRESS easily. For example, Hurvich and Tsai (1989) could be able to include PRESS to their Monte Carlo study where they compare finite sample properties of several different model selection criteria for regression models.

The last criterion included in our study is the often-used sequential F test. We start with the largest model AR(M), and examine the t statistic of $\hat{a}_M$. If this is insignificant, we drop it and re-estimate an AR(M-1) model. We keep dropping the

¹There are potential problems which arise from omitting middle observations in time series. We made adjustments to account for such problems but found that such adjustments made no difference in the outcome, and hence have chosen to report results without any adjustments.

last coefficient until we get a significant one. In the simulation study the significance level is chosen to be 5 percent.

2.1.2 The Overestimation Problem

In this section we present the simulation results where we compare the finite sample performance of the six criteria defined in the previous section. Our comparison is based on the frequency distribution of the estimated lag order for the criteria and mean squared forecast errors of the models chosen by the criteria. In order to be able to compare our results with the results of Lütkepohl we preserve his assumptions about the true lag order, maximum possible lag order and sample size. We consider two different processes where one assumes two lags and the other one assumes three lags as the true lag order. M is taken as six. The sample sizes are 30 and 200. Monte Carlo sample size is 1000. To have a stationary process we generate the regression coefficients from a uniform distribution in the region $(-1/2, 1/2)$ for the true model with two lags and $(0, 1/3)$ for the true model with three lags. Error terms are generated from a standard normal distribution.

Tables 1a and 1b report the frequency distribution of the estimated lag order and mean squared forecast errors of the models chosen by the criteria defined in the previous section where the true model assumes two and three lags respectively. The first outcome which is important to note from the tables is that the probabilities of estimating lag orders greater than the true lag order is considerably high for PRESS compared to other criteria. For example, in table 1a, while the percentage of estimating the lag order 3 for PRESS criterion is 14.9, this percentage for SC is only 5.7 when sample size is 30. We have similar results in table 1b as well. In table 2, we have summary results on the probabilities of overestimation of the six criteria.

As can be seen from the table, the percentages of estimating lag orders higher than the true lag order for PRESS are highest for each true lag order and sample size. They are written in bold form in the table. For example, while the percentage of overestimation is 48.8 for PRESS, it is only 11 for SC when the true lag order is 2 and sample size is 30. These results imply that PRESS does not penalize high orders as much as the other criteria and so it is not a parsimonious model selection criterion. Same problem is valid to some degree for AIC and AIC_C , more for AIC_C . SC is the best criterion in terms of penalization, it is a parsimonious criterion. Depending on this observation the second outcome we can note is that not being parsimonious while selecting the model results with high mean squared forecast errors. In tables 1a and b PRESS has the highest mean squared forecast error. A model chosen by setting the coefficient to zero rather than putting an OLS estimate gives a better forecast since addition of variables almost always results with a higher variance of forecast (Allen, 1971) and (Walls and Weeks, 1969).

Both in tables 1a and 1b AIC_C and HQC estimate the true lag order most often for sample sizes 30 and 200, respectively. For sample size 30 and for both of the models SC has the lowest mean squared forecast error rather than the AIC_C . This is due to the fact that AIC_C is not that parsimonious while selecting the model but SC is the most parsimonious criterion. For sample size 200 and for both of the models mean squared forecast errors are close to each other.

The consequences of the overestimation problem of PRESS can better be understood when we reduce M to the true lag order. In tables 3a and 3b we have results for this case. In table 3a M is 2 and in table 3b M is 3. PRESS estimates the true lag with the highest probability in both of the tables for both of the sample sizes. Its forecasting performance is better than the other criteria now, since it is not possible

to estimate higher lag orders under this situation. For the model where the true lag order is 2 PRESS has the lowest mean squared forecast error. Although for the model where the true lag order is 3 AIC_C has a smaller mean squared forecast error, the difference is minor. Performance of AIC_C is also good for this case because it also does not penalize high orders of lag enough. SC becomes the worst criterion.

An alternative way of depicting the overestimation problem of PRESS is to search the model where PRESS function takes the minimum value. Figure 1 presents this information. We have 11 different models where the number of lags varies from 0 to 10. On the vertical axis we see the value of PRESS function. Number of lags for the true model is taken as 2 and once again error terms are assumed to be standard normal. Sample size is taken as 200. 10 different simulation results are presented in the figure. In 9 of the cases PRESS function takes its minimum value for a model where the lag order is higher than the true lag order.

The above simulation results show that parsimonious criteria are better in terms of forecasting performance. For example, PRESS is not that parsimonious while selecting the model and this affects its forecasting performance badly. PRESS chooses higher lag orders with a higher probability. If M is large compared to the true lag order this causes a low performance of PRESS in terms of estimating the true lag order. On the other hand, SC is an example of a parsimonious model selection criterion and its forecasting performance is very good when M is large compared to the true lag order.

2.2 Effects of Skewness and Kurtosis on Model Selection Criteria

All prior studies which consider the performance of model selection criteria, both theoretical and empirical, assume the errors to be normally distributed. This common assumption is also employed in the previous section of this chapter. In this section, we relax this assumption and try to assess the performance of some of the model selection criteria when error terms are nonnormal. From the theory of Edgeworth expansions, we know that the first two terms in asymptotic approximations arising from lack of normality correspond to skewness and kurtosis respectively. Thus a good approach to assess robustness is to vary skewness and kurtosis and study the behavior of the model selection criteria. Since most of the financial data possesses skewness and kurtosis, it is important for applied econometricians to know about the performance of the criteria under such cases.

In this section we have the same model presented in section 2.1.1. We generate the data for the simulation study as described in section 2.1.2 except for the normality of the error terms. Same criteria are under consideration. For our robustness studies, we use two classes of error distributions. Let $X \sim \mathcal{G}(k, 2) \equiv \chi_{2k}^2$. To study the effect of skewness, we considered $u_t \stackrel{iid}{\sim} F$, where F is the distribution of $(X/(2\sqrt{k})) - \sqrt{k}$. This has mean 0, variance 1, and $\mathbf{E}u_t^3 = 2\sqrt{k}$ so that skewness increases with k . To study the effect of kurtosis, let $u_t \stackrel{iid}{\sim} F_k$ where F is a Student's t distribution scaled to have variance 1. As k increases, the kurtosis decrease, converging to that of the normal distribution asymptotically. Note that skewness is 0 for this class of distributions. When k is one we have the Cauchy distribution. In the previous section we presented the frequency distribution of the estimated lag order. In this section rather than reporting the distribution we will only consider the probability of

correctly estimating the lag order p since providing the distribution for each degrees of freedom complicates the analysis of the results.²

In figures 2a,b we can see the simulation results where we have a skewed distribution for the error terms when the true lag order is 2. On the horizontal axis we can see the degrees of freedom, k . As we move to right on the horizontal axis skewness increases. Figure 2a and 2b plot the simulation results performed over a sample size of 30 and 200, respectively. From the figures we can see that skewness does not have much effect on the performance of criteria while estimating the true lag order. Although AIC_C is the best criterion for sample size 30, it becomes one of the worst criteria for sample size 200. Performance of PRESS is similar to AIC_C . Its performance also declines considerably as we increase sample size. Asymptotically consistent criteria, SC and HQC, are not doing that well for sample size 30 but they estimate the true lag most often when we increase sample size to 200. Performance of AIC and sequential F test are close to each other.

In figures 3a,b we have the same setting but this time we have error terms generated from a t distribution. Again on the horizontal axis we have degrees of freedom, k , and as k increases we get more closer to normality and so kurtosis decrease. We see from the figures that kurtosis has a considerable effect on the criteria under consideration, especially on AIC_C , sequential F and PRESS. When k is small, that is, for heavy-tailed error distributions, performance of PRESS and sequential F test are good but as k increase, that is, when we get closer to normal distribution, we see a decline in their performance. The decline is much more for PRESS. Converse is true for AIC_C . The result for AIC_C is what is to be expected because Hurvich and Tsai (1989) makes the bias correction under the assumption of normality. These results

²These distributions are available from the author upon request.

are more obvious for a sample size of 200. Performance of other criteria slightly increase as we move towards normal distribution when sample size is 30. For sample size 200, it seems that kurtosis does not effect these criteria. Only the performance of AIC decrease slightly. Once again SC and HQC are performing badly when sample size is 30 but they are the best ones when we increase sample size to 200. PRESS and AIC_C are the worst criteria for sample size 200.

In figures 4 and 5 we have the probabilities for correctly estimating p when the true lag order is 3. Figures 4a,b present the results for the skewed distribution case for sample sizes 30 and 200 respectively. Once again we see that skewness does not have much effect on the performance of criteria while estimating the true lag order. Again AIC_C and PRESS perform very well when sample size is 30 but they are among the worst when sample size is 200. HQC criterion performs badly when sample size is 30 but for sample size 200 it is among the best ones. The results for SC differ from the ones that we obtained for the case where we have the true lag order as 2. This time performance of SC does not increase as we increase the sample size from 30 to 200. Again performance of AIC and Sequential F are similar to each other.

Figures 5a,b present the results for the t distribution case for sample sizes 30 and 200 respectively when the true lag order is 3. Results again are similar to the case when we have true lag order as 2. AIC_C and PRESS criteria are influenced very much from kurtosis. PRESS and Sequential F criteria perform well for heavy-tailed error distributions but as we move to normality their performance decline. The decline is much more for PRESS. Converse is true for AIC_C . The effect of kurtosis on other criteria is more when sample size is 30, it is an increasing effect, but when sample size is 200 kurtosis does not effect the performance of these criteria. HQC

performs badly when sample size is 30 but its performance increase when sample size is 200. Differing from the case of true lag order of 2 performance of SC does not increase very much as we increase sample size from 30 to 200. Again PRESS and AIC_C are the worst criteria for sample size 200.

To conclude we can say that skewness does not effect the probability of estimating the true lag order but kurtosis effects it especially for the criteria PRESS and AIC_C . When the sample size is small and error distributions are not heavy-tailed (corresponding to large values of degrees of freedom, k) we suggest the use of AIC_C for lag order selection. On the other hand, for heavy-tailed error distributions PRESS and sequential F test are sucessfull in depicting the true lag order. For large samples HQC is definitely the best criteria to be used.

We also studied forecasting performance based on models selected by the various criteria. Simulation results are provided in figures 6 to 9. In figures 6 and 7 the true lag order for the model is 2 and in figures 8 and 9 it is 3. In figures 6a and b the error terms are generated from a skewed distribution. From the figures we see that skewness does not effect the forecasting performance of the criteria under consideration. This correlates with the results of probability of estimating the true lag order. In figure 6a where the sample size is 30, the best criteria for forecasting performance is SC. This does not correlate with our previous result since SC is not a good criterion while estimating the true lag order. On the other hand AIC_C which is the best criterion in terms of estimating the true lag order for this case has less forecasting performance than SC. This shows that estimating the true lag order most often does not imply that the forecasting performance will also be good. From the previous chapter we know that SC is more parsimonious while selecting the model. This implies that assigning a zero coefficient rather than the OLS estimate results

in better forecasts. In figure 6b we increase the sample size to 200. Still in this case SC performs very well but this time HQC is the best criterion. For sample size 200 the criteria which estimates the true lag order most often also seems to perform well in terms of forecasting.

In figures 7a and b we plot the mean squared forecast errors for a model where the error terms are generated from a t distribution for sample sizes 30 and 200, respectively. We see that kurtosis does effect the performance of the criteria. As we move to normality, mean squared forecast errors decrease. For both of the sample sizes forecasting performance of all the criteria are more or the less the same.

In figures 8 and 9 we plot the mean squared forecast errors for the model where the true lag order is 3. Results are similar to the case where the true lag order is 2. The only difference is when sample size is 200, SC is not performing well for a skewed distribution case. We know from the results presented in figure 4b that it is also not good in terms of estimating the true lag order for this case.

As a conclusion we can say that estimating the true lag order most often does not imply that the model chosen by that criteria will make the best forecast. If our aim is to estimate the true lag order and if we are sure that we do not have a heavy-tailed distribution, it is best to use AIC_C for small sample sizes For heavy-tailed distributions and small sample sizes PRESS or SEQF test should be used. For large sample sizes HQC is the best criteria to use. If our aim is to make a good forecast, for small sample sizes we suggest the use of SC but for large sample sizes it is better to use HQC.

3 Variance Estimates and Model Selection

In this chapter, we suggest some new model selection criteria where we use the cross-validation technique. The large majority of the criteria for model selection are functions of the $\hat{\sigma}^2$, the usual variance estimate for a regression model. Among the six criteria that we consider in the previous chapter four of them, AIC, AIC_C , SC and HQC also contain $\hat{\sigma}^2$ in their functions. The validity of the usual variance estimate depends on some assumptions, most critically the validity of the model being estimated. This is often violated in model selection contexts, where model search takes place over invalid models. A cross validated estimate of variance is more robust to specification errors (see, for example, Efron(1983)). In this chapter we consider the effects of replacing the usual estimate of variance by two different cross-validated estimates in the functions of already existing model selection criteria so by this way we suggest some new model selection criteria. The cross-validation technique is computationally time consuming but with the advancements in the computer technology, it is now possible to make such replacements.

Efron (1983) shows that the error rate of a predictive rule is underestimated if the same data used to both construct and to evaluate the rule. The residual $r_t = y_t - \hat{y}_t$ underestimates the true error at t since the t -th observation has been used in fitting the equation³. One way to reduce the problem is to use $r'_t = y_t - \tilde{y}_t$, where $\tilde{y}_t$ is the forecast of y_t based on a regression which excludes the t -th observation. This is actually PRESS of Allen (1974) and we studied its performance as a model selection criterion for AR models in the previous chapter. There, we have shown that PRESS did not perform well neither for lag order selection nor for forecasting for AR

³One way to see this is to note that $RSS(\hat{\beta}) < RSS(\beta)$ – the residual sum of squares is minimized by $\hat{\beta}$ so that it must be smaller than the true residual sum squares based on the true parameter β .

models. Only when error terms have a heavy-tailed distribution PRESS performs well. Simulation results show that this bad performance of PRESS is due to the fact that it does not penalize high orders of lag enough so there is an overestimation problem.

Since $\text{PRESS}/(T-K)$, where $K = p + 1$, is a cross-validated estimate of the variance, using PRESS for model selection is analogous to the use of the plain $\hat{\sigma}^2$. The performance of the usual variance estimate as a model selection criterion is much improved by penalizing high order lags – most existing model selection criteria do precisely this. The four criteria which are used in the simulation studies of previous chapter also do contain penalty factors in their functions. This leads to the natural idea that the performance of PRESS can similarly be improved. In section 3.1, we present simulation results which show the effects of substituting $\text{PRESS}/(T-K)$ as a cross-validated estimate of variance for the usual estimate of variance in the functions of some of the already existing criteria. As in chapter 2, we again consider the AR models and try to estimate the true lag order of the model. This, we believe, will avoid the over optimistic results mentioned in Efron (1983). In chapter 2, we mentioned that PRESS has an overestimation problem. By substituting PRESS in to the functions of conventional criteria we automatically add penalty factors to PRESS so we avoid the problem of overestimation.

As mentioned PRESS sequentially deletes one of the observations while making the prediction. It is also possible to delete two, three or more observations. One other approach is to delete some percentage of the data. For example Weiss and Indurkha (1996) apply this approach to learning methods where they try to select the right-size model. They exclude 10 percent of the data in their calculations. In section 3.2 we compare these two techniques of cross-validation, deleting one

observation and deleting 10 percent of the observations, in a simulation study again for AR models. One advantage of excluding more observations is on computer time. The time considerably decreases as more observations are excluded. More than this simulation results show that as sample size increase deleting 10 percent of the observations gives better results. In section 3.2 we also present an alternative cross-validated estimate of variance where we delete 10 percent of the observations. We substitute this estimate in to the function of AIC and compare the performance of this modified criteria and AIC in terms of estimating the true lag order of the AR models.

3.1 A Cross-Validated Estimate of Variance

In this section, we suggest two new criteria for model selection. We define $\tilde{\sigma}^2 = PRESS/(T - K)$ which is a cross-validated estimate of variance. One can replace $\hat{\sigma}^2$ by $\tilde{\sigma}^2$ in the functions of conventional model selection criteria. SC, HQC and AIC have similar functional forms. They all include the logarithm of the usual estimate of variance but they add to it different linear penalty factors. In a few Monte-Carlo simulations we observed that the improvements obtained by the replacement of usual estimate of variance by cross-validated estimate are similar for each of the criteria. For this reason we only include AIC, among the three, in our simulation study. On the other hand AIC_C have a nonlinear penalty factor. To see the effect of this nonlinearity a replacement for AIC_C also takes place. Then we can define the suggested two criteria as follows:

$$AICPRESS(k) = \ln \tilde{\sigma}_k^2 + (2k)/T.$$

$$AIC_CPRESS(k) = T \ln \tilde{\sigma}_k^2 + T \frac{1 + k/T}{1 - (k + 2)/T}$$

Note that they are exactly same as AIC and AIC_C except that instead of $\hat{\sigma}_k^2$ we have here $\tilde{\sigma}_k^2$.

The first simulation results presented in tables 4a and 4b are the frequency distributions of the estimated lag order and mean squared forecast errors for PRESS criterion and the newly suggested two criteria. Here we use the model presented in section 2.1 where error terms are generated from a normal distribution. In table 4a we provide the results for the true lag order of 2 and in table 4b for the true lag order of 3. These results suggest that the probabilities of estimating lag orders higher than the true lag order for the newly suggested criteria are considerably lower than the original PRESS. In other words, these criteria penalize high orders of lag more than the original PRESS. For example in table 4a, where the true lag is 2, the percentage of time PRESS estimates the lag order as 3 is 14.9 for sample size 30. On the other hand, this percentage is only 3.1 for AICPRESS and 4.7 for AIC_CPRESS . We have similar results in table 4b as well. In table 5, we have summary results on the probabilities of overestimation of the three criteria. As can be seen from the table, the percentages of overestimation for PRESS criterion are considerably higher than the two newly suggested model selection criteria. They are written in bold form in the table. This implies that by adding penalty factor to PRESS we overcome the problem of overestimating the true lag order. Moreover, from tables 4a and b, we can see that mean squared forecast errors of the newly suggested criteria are lower than the original PRESS. This is an indication that being parsimonious while selecting the model results in a better forecasting performance. Setting the coefficient to zero rather than estimating it with OLS improves the forecast.

In table 4a, AIC_CPRESS estimates the true lag order most often when the sample size is 30. When we increase sample size to 200, AICPRESS is slightly

better than the AIC_CPRESS . Results of mean squared forecast error correlate with the results of estimating true lag order. In table 4b, PRESS estimates the true lag order most often for sample size 30. This good performance of PRESS can be explained by the fact that the true lag order, p , in for this case is closer to the maximum possible lag order, M . From tables 3a and 3b, we know that PRESS has very high performance when $p = M$. For sample size 200, AIC_CPRESS estimates the true lag order most often and its forecasting performance is the best one for both of the sample sizes.

In figures 10 to 13 we present the simulation results for the probability of estimating the true lag order for the eight criteria introduced so far. In figures 10 and 11 the true lag order is 2. In figures 12 and 13 it is 3. As in chapter 2 we relax the assumption of normality in order to study the effects of skewness and kurtosis on model selection criteria. Since as we increase degrees of freedom of the t distribution we converge to normal distribution, it is also possible to have an idea how the criteria perform under normal distribution when degrees of freedom is 100 in the figures.

From the simulation results of chapter 2 we know that skewness does not effect the performance of the six criteria considered in that chapter. On the other hand kurtosis does have some effect on the criteria, especially on AIC_C , sequential F and PRESS. Depending on the results presented in figures 10 to 13 we can say that skewness also does not effect the performance of the two newly suggested criteria. On the other hand, it seems that AIC_CPRESS is sensitive to kurtosis. In figure 11a where the sample size is 30, we see a slight decrease in the performance of the criterion as we move towards normality. When we increase sample size to 200 in figure 11b, we see a converse result. For heavy-tailed distributions AIC_CPRESS is performing badly

but as we move to normality its performance increases. This implies that depending on the sample size the effect of kurtosis on AIC_CPRESS differ. The behaviour of AIC_CPRESS for sample size 200 is similar to the behaviour of AIC_C but for sample size 30 it is just the reverse. The same results for AIC_CPRESS can be seen more clearly in figures 13a and b where the true lag order is 3.

In figure 10a we present the case of skewed distribution for sample size is 30. AIC_C is the best criterion for this case. Performance of AIC, PRESS and HQC are close to eachother and they are the second best. Newly suggested criteria can only catch the performance of SC and sequential F test. Among these AICPRESS is the worst one. So for this case modifications do not lead to improvements over any of the criteria, AIC, PRESS or AIC_C . When we increase sample size to 200 in figure 10b, results change a lot. Performance of newly suggested criteria increase together with the asymptotically consistent criteria SC and HQC. Their performance are close to eachother and they are the best ones. On the other hand for this case AIC_C and PRESS becomes the worst criteria. So modifications for sample size 200 causes a considerable improvement over all of the criteria, AIC, AIC_C and PRESS.

In figures 11a and b we have t distribution for the error terms. Sample sizes are 30 and 200, respectively. The performance of AICPRESS is not good for sample size 30 but it approaches to the performance of SC and HQC when we increase sample size to 200. For this sample size it performs better than both AIC and PRESS so modifications caused an improvement. AIC_CPRESS performs well when sample size is 30 but still it performs worse than AIC_C for $k \geq 3$. Its performance is close to PRESS. For sample size 200 if we have heavy-tailed distribution performance of AIC_CPRESS is bad but still it performs better than either AIC_C and $PRESS$. So we can say that for t distribution also we ended up with an improvement when we

have sample size 200. In figures 12 and 13 we have the results for the model where the true lag is 3. They are similar to the results for the model where the true lag is 2 but the changes can be seen more clearly for this case.

In figures 14 to 17 we have the simulation results of the mean squared forecast errors for the 8 criteria. Simulation results of chapter 2 showed that skewness does not effect the performance of the criteria. In this chapter we see that it does not effect the performance of the additional two criteria also. These results are similar to the results for probability of estimating the true lag order. On the other hand in line with the results obtained in chapter 2, the mean squared forecast error of the additional two criteria decrease as we move to normality. This is an indicator that kurtosis does effect the forecasting performance of these two criteria.

In figure 14a true lag order is 2, sample size is 30 and error terms are generated from a skewed distribution. We see that the two newly suggested criteria have the best performance. SC has the closest performance. In figure 14b where sample size increases to 200 we still see that the two new criteria perform very well. In addition to SC, HQC also performs well in this case. So for skewed distribution case modifications lead an improvement in terms of forecasting both for small and large sample sizes. In figures 15a and b error terms are generated from a t distribution for sample sizes 30 and 200 respectively. For this case we do not see a significant difference in mean squared forecast errors of the criteria. They are almost the same. In figures 16 and 17 we have results where the true model has 3 lags. Results are similar to the ones that we obtained for the model with 2 lags. The only difference is that SC is not performing well for a skewed distribution case when the sample size is 200.

As a conclusion we can say that replacing the usual estimate of variance with

a cross-validated estimate of variance in the functions of criteria AIC and AIC_C improves their performance in most of the cases. When the aim is estimating the lag order although for sample size 30 original AIC and AIC_C perform better, for sample size 200 the modified versions are much more better. When the aim is forecasting it is definitely better to use modified versions since they are forecasting considerably well when the error terms are generated from a skewed distribution for both small and large samples and their performance are same with others when the error terms are generated from a t distribution.

3.2 An Alternative Cross-Validated Estimate of Variance

In the previous section we suggested two new model selection criteria where cross-validation method is used. Specifically, we replaced the usual estimate of variance which exists in the functions of AIC and AIC_C by a cross-validated estimate of variance. This new estimate was calculated by sequentially deleting one of the observations in the process of cross-validation. As stated before, it is also possible to delete sequentially some percentage of the data rather than just deleting one observation. In this section we consider deleting 10 percent of the observations sequentially while calculating an alternative cross-validated estimate of variance. We define the sum of squares based on the residuals obtained by this way $CV10 = \sum_{j=1}^J (x_j - \tilde{x}_j)'(x_j - \tilde{x}_j)$. J is the total number of intervals where in each interval 10 percent of the data is contained. $\tilde{x}_j$ is the predicted value of the vector x_j from an AR(p) model after omitting the observations in the j^{th} interval for $j = 1, 2, \dots, J$ in the process of estimation. Then $\bar{\sigma} = CV10/(T - K)$ is another cross-validated estimate of variance. In the previous section we replaced the usual estimate of variance, $\hat{\sigma}$ with a cross-validated estimate of variance, $\bar{\sigma} = PRESS/(T - K)$ in the

functions of both the AIC and AIC_C . The aim of considering these two criteria was to see whether a linear or a nonlinear penalty factor works better. The results show that in some of the cases AICPRESS is better and yet in some other AIC_CPRESS is better. So one does not have a complete superiority over the other. For this reason in order not to complicate the outputs of the simulations, in this chapter we exclude the criterion AIC_C from our analysis. Then the only criterion newly suggested in this chapter is

$$AICCV10(k) = \ln \bar{\sigma}_k^2 + (2k)/T.$$

Note that the only difference from the function of AIC is that $\hat{\sigma}$ is replaced by $\bar{\sigma}$.

In this section, we have five different model selection criteria where we compare the finite sample performances. They are PRESS, CV10, AIC, AICPRESS and AICCV10. The comparison is done by considering the probability of estimating the true lag order, mean squared forecast errors and absolute errors made by the models chosen by each of the criteria. In the simulation study we consider the difference of the probabilities of estimating the true lag order. Specifically three differences are of interest, $AIC - AICPRESS$, $AIC - AICCV10$ and $PRESS - CV10$. Let's define the indicator function $I(\text{criterion})$ as taking value 1 if the criterion estimates the true lag order and taking value 0 otherwise. Now let's define the following differences:

$$\Delta_1 = I(AIC) - I(AICPRESS)$$

$$\Delta_2 = I(AIC) - I(AICCV10)$$

$$\Delta_3 = I(PRESS) - I(CV10)$$

Note that Δ_i , $i = 1, 2, 3$ can have only three values. If both of the criteria estimates the true lag order or both estimate some lag order different than the true one it is 0. If the first criterion estimates the true lag order and the second one estimates

some other lag order it is 1. For the converse case it is -1. Let $P_{i,j}$ be the probability that $\Delta_i = j$ for $i = 1, 2, 3$ and $j = -1, 0, 1$. Note that expected value of Δ_i is $P_{i,1} - P_{i,-1}$ for $i = 1, 2, 3$. If expected value of Δ_i is estimated to be the Monte Carlo average, $\hat{\Delta}_i = \sum_{m=1}^{MCSS} \Delta_i / MCSS$, standard deviation of $\hat{\Delta}_i$ is $\sqrt{\frac{\hat{P}_{i,1} + \hat{P}_{i,-1} - (\hat{P}_{i,1} - \hat{P}_{i,-1})^2}{MCSS}}$ for $i = 1, 2, 3$. MCSS is the Monte Carlo sample size and $\hat{P}_{i,j}$ is the Monte Carlo estimate of $P_{i,j}$ for $i = 1, 2, 3$ and $j = -1, 0, 1$. While reporting the results of the simulation study we concentrate on $\hat{\Delta}_i$. Note that if it takes a positive value the first criterion is better than the second one in terms of estimating the true lag order but if it takes a negative value just the reverse is true.

Secondly we base our comparison on the differences of the mean squared forecast errors obtained from the models chosen by the criteria. Let's define it as $MSFE(criterion) = E(Y_t - \hat{Y}_t)^2$ for $t = 1, 2, \dots, T$. $\hat{Y}_t$ is the forecast of the model. In the simulation study we will concentrate on the difference of two mean squared forecast errors obtained from two different criteria. We can define the differences under consideration as follows:

$$\delta_1 = MSFE(AIC) - MSFE(AICPRESS)$$

$$\delta_2 = MSFE(AIC) - MSFE(AICCV10)$$

$$\delta_3 = MSFE(PRESS) - MSFE(CV10)$$

The expected value of δ_i for $i = 1, 2, 3$ is $\hat{\delta}_i = \sum_{m=1}^{MCSS} \delta_i / MCSS$. The estimate of the variance of the Monte Carlo estimates can be defined as $V = \sum_{m=1}^{MCSS} (\delta_i - \hat{\delta}_i)^2 / MCSS$ for $i = 1, 2, 3$. Then the standard deviation of $\hat{\delta}_i$ for $i = 1, 2, 3$ can be defined as $\sqrt{V/MCSS}$. While reporting the results of the simulation study we concentrate on $\hat{\delta}_i$. Note that if it takes a positive value second criteria has a better forecasting performance but if it is negative first criteria has a better forecasting performance.

Finally, we consider the absolute errors while comparing the criteria. Let's define

the absolute error of the model chosen by some criterion as $ABS(criterion) = |Y_t - \hat{Y}_t|$ for $t = 1, 2, \dots, T$. We can define the difference under consideration as follows:

$$d_1 = ABS(AIC) - ABS(AICPRESS)$$

$$d_2 = ABS(AIC) - ABS(AICCV10)$$

$$d_3 = ABS(PRESS) - ABS(CV10)$$

The expected value of d_i for $i = 1, 2, 3$ is $\hat{d}_i = \sum_{m=1}^{MCSS} d_i / MCSS$. The estimate of the variance of the Monte Carlo estimates can be defined as $V = \sum_{m=1}^{MCSS} (d_i - \hat{d}_i)^2 / MCSS$ for $i = 1, 2, 3$. Then the standard deviation of $\hat{d}_i$ for $i = 1, 2, 3$ can be defined as $\sqrt{V/MCSS}$. While reporting the results of the simulation study we concentrate on $\hat{d}_i$. Note that if it takes a positive value second criteria has a better forecasting performance but if it is negative first criteria has a better forecasting performance.

For the simulation study of this section the AR model presented in chapter 2 is valid. We assume that the true lag order, p is 2 and the maximum possible lag order, M is 6. We consider two types distributions for the error terms while generating the data, normal distribution and skewed distribution. For the skewed distribution case we choose the degrees of freedom as 1. A few simulation results show that changing the degrees of freedom does not effect the relative performance of criteria under consideration. The results of the previous section show that sample size has effect on the performance of the newly introduced criteria especially if the aim is to estimate the true lag order. For sample size 30 we do not see an improvement but for sample size 200 there is a considerable improvement. For this reason in this chapter we vary sample size and try to find out the critical sample size at which we will observe an improvement from using the new criteria. We compare the performance

of criteria for sample sizes $T = 30, 40, \dots, 200$. The Monte Carlo sample size is 1000.

In figures 18 to 20 we have the expected value of the probability differences of estimating the true lag order, $\hat{\Delta}_i$ for $i = 1, 2, 3$ for each sample size. Both normal and skewed distribution cases are presented in the figure. In figure 18 we compare AIC with AICPRESS. We can say that for small sample sizes AIC is estimating the true lag order more precisely than AICPRESS but for large sample sizes the converse is true. For the normal distribution case the critical sample size is 70 and for the skewed distribution case it is 120. These results correlate with the results that we obtained in the previous section. There, at sample size 30, AIC is performing better than AICPRESS and at sample size 200 just the reverse is true. (Figure 10a and b and figure 11a and b, note that t_{100} is close to normal distribution.) In figure 19, we compare AIC with AICCV10. For small sample sizes we see some fluctuations but for large sample sizes AICCV10 estimates the true lag order more precisely than AIC. The critical sample size is 110 for both of the distributions. In figure 20, we compare the criteria PRESS and CV10. Again for small sample sizes we see fluctuations but for large sample sizes CV10 is better than PRESS. All of the differences in the figures are significantly different than zero.

Since we see the fluctuations only in figures 19 and 20 we can conclude that these are due to CV10. Note that in figure 18 we have AICPRESS and for its calculation we are deleting only one observation for each sample size. However while calculating either AICCV10 or CV10 we are deleting a different number of observations for each sample size and for small sample size this change in the number of deleted observations is larger as a percentage so this may cause the fluctuations for small sample sizes.

In figures 21 and 22 we have the expected value of the differences of the mean

squared forecast errors of models chosen by the criteria, $\hat{\delta}_i$ for $i = 1, 2, 3$ for each sample size. Figures 21a, b and c presents the results for the normal distribution case and figures 22a, b and c presents the results for the skewed distribution case. In figure 21a we have a comparison of AIC and AICPRESS. For sample sizes 30 and 40 the forecasting performance of AIC is worse than AICPRESS. This does not correlate with the results on probability of estimating the true lag order because for those sample sizes AIC estimates the true lag order more than AICPRESS. This may be due to overestimation problem mentioned in chapter 2, where forming a model by putting a zero coefficient rather than putting an OLS estimate may result with a better forecast. For sample sizes 50 to 130 the difference in the errors are not significant so both of the criteria perform in a similar way. After sample size 130, AICPRESS is significantly better than AIC. In figure 21b we have a comparison of AIC and AICCV10. Except three sample sizes, 40, 90 and 150, AICCV10 is significantly better than AIC. In figure 21c, we have a comparison of PRESS and CV10. Except for four sample sizes, 40, 50, 90 and 100, CV10 is significantly better than PRESS. In figures 22a, b and c we have the results for the skewed distribution case. In figure 22a we compare AIC with AICPRESS and we see that for most of the sample sizes the difference of the mean squared forecast errors are not significant implying that performance of the both of the criteria are similar. In figure 22b, we compare the performance of AIC with AICCV10. For most of the sample sizes, especially for large sample sizes, the difference is significantly positive implying that forecasting performance of AICCV10 is better than AIC. In figure 22c, we compare PRESS with CV10. Again for most of the sample sizes and especially for large sample sizes the difference is significantly positive implying better performance of CV10.

In figures 23 and 24 we have the simulation results of the expected value of the difference of the two absolute errors of models chosen by two different criteria, d_i for $i = 1, 2, 3$ and for each sample size. In figures 23a, b and c the error terms are generated from a normal distribution and in figures 24a, b and c they are generated from a skewed distribution while building the model. First note that in all of the figures we have positive values. This implies that performances of both AICPRESS and AICCV10 are better than AIC and also CV10 is better than PRESS. Except for the case in figure 24a where we compare AIC with AICPRESS under a skewed distribution all of the differences are significantly different than zero.

As a conclusion out of these simulation results we can say that both of the cross-validated techniques perform better than AIC in terms of both estimating the true lag order and forecasting when the sample size is large. By large we mean a sample size which is approximately above 100. Also CV10 is better than PRESS for large sample sizes. For small sample sizes performance of CV10 fluctuates as sample size change. For this reason also it may not be a good idea to use CV10 for sample sizes under 100.

4 Alternative Bayesian Tests from Different Priors for Detecting Structural Change with an Unknown Changepoint

In chapters 2 and 3 of this dissertation we studied the performance of different model selection criteria for AR processes. In this final chapter, we deviate from this topic and concentrate on structural change problems. One may think that it is not related to model selection problems but in fact existence of a structural change in a data and not detecting it is an hazard for applied econometricians, with serious consequences for model performance and forecasting. In such a case whatever criterion is chosen, the performance will be bad. For this reason it is important to know from the start whether the data possesses a structural change or not.

In this chapter of the dissertation we work on detecting structural change when the change point is unknown. There are various tests used for this aim. Among them the most popular ones are *CUSUM* and *CUSUM of squares* test of Brown, Durbin and Evans (1975) and the *fluctuation test* of Sen (1980) and Ploberger, Krämer and Kontrus (1989). These tests depends on frequentist approach. However, in this chapter we follow the Bayesian approach and suggest Bayesian tests formed under different prior distributions. By this way we are able to see how robust is the Bayesian approach to changes of assumptions about the priors of the model. In the literature we see the idea of using posterior odds as a test of structural change (for example see Zellner (1987)) but the explicit form of the tests are not given so it is not possible to use Bayesian tests in practice for the unknown change problems. Here we present the tests explicitly after a series of calculations so that they are ready for use in applied works.

Since Bayesian tests do not have a known distribution the critical values for these tests must be found by Monte Carlo estimate. The applied econometricians using the data in hand can find out the critical value specific to their data. The effort of calculating the critical value will worth it since these tests belong to the optimal class of tests for structural change with unknown change point. (See Andrews, Lee and Ploberger (1996))

4.1 The Rolling Chow Test

Consider the regression model $y_t = x_t\beta_t + e_t$, for $t = 1, 2, \dots, T$, where y_t is scalar, x_t is a $1 \times K$ row vector, β_t is a $K \times 1$ vector of regression coefficients and e_t is the scalar error term. We will use Y_t and Y^t to denote the $t \times 1$ vector $(y_1, \dots, y_t)'$ and the $(T - t) \times 1$ vector $(y_{t+1}, \dots, y_T)'$ respectively. Similarly, X_t and X^t are the $t \times K$ matrix with rows x_1 to x_t and x_{t+1} to x_T respectively. Similarly, $\epsilon_t = (e_1, \dots, e_t)'$ and $\epsilon^t = (e_{t+1}, \dots, e_T)'$. We will assume throughout that $\epsilon \equiv \epsilon_T \equiv \epsilon^0 \sim N(0, \sigma^2 \mathbf{I}_T)$

We wish to test the null hypothesis that $\beta_t = \beta_0$ for all t against the alternative that there exists a changepoint m such that $\beta_t = \beta_1$ for all $t \leq m$ and $\beta_t = \beta_2$ for all $t > m$. When m is known, the usual (Chow) F-test is UMP Invariant; see Zaman (1996). Define $SSR_t = \|Y_t - X_t\hat{\beta}_t\|^2$, where $\hat{\beta}_t = (X_t'X_t)^{-1}X_t'Y_t$ and $SSR^t = \|Y^t - X^t\hat{\beta}^t\|^2$ where $\hat{\beta}^t = (X^tX^t)^{-1}X^tY^t$. Then the F test for a changepoint at m is based on

$$\begin{aligned} F(m) &= \frac{\{SSR_T - (SSR_m + SSR^m)\}/k}{(SSR_m + SSR^m)/(T - 2k)} \\ &= \frac{(\hat{\sigma}_0^2 - \hat{\sigma}_m^2)/k}{\hat{\sigma}_m^2/(T - 2k)} \end{aligned} \quad (1)$$

In the second equality, $\hat{\sigma}_0^2$ is the ML estimate of σ under the null hypothesis, while $\hat{\sigma}_m^2$ is the ML estimate under the assumption that there is a changepoint at m .

When m is unknown, a natural intuitive approach would be to estimate m , and then apply the F-test at the estimated m . To find this m , note that the likelihood of the data under the alternative hypothesis that there is a changepoint at m can be written as:

$$l_1(\beta_1, \beta_2, \sigma^2, m) = (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{1}{2\sigma^2} \left\{ \|Y_m - X_m\beta_1\|^2 + \|Y^m - X^m\beta_2\|^2 \right\}\right)$$

Given m , l_1 is maximized by setting $\beta_1 = \hat{\beta}_m$, $\beta_2 = \hat{\beta}^m$ and $\sigma^2 = (SSR_m + SSR^m)/T$. Thus we have

$$\sup_{\beta_1, \beta_2, \sigma^2} l_1 = (2\pi\hat{\sigma}_m^2)^{-T/2} \exp(-T\hat{\sigma}_m^2/2)$$

Maximizing the likelihood with respect to m is equivalent to finding the m for which $\hat{\sigma}_m^2$ is minimum. From expression (1), it is clear that the $F(m)$ is monotone decreasing in $\hat{\sigma}_m^2$ and hence is maximized by minimizing $\hat{\sigma}_m^2$. It follows immediately that the test obtained by applying the Chow test at $\hat{m}$ is numerically equivalent to the one obtained by calculating the largest possible value of the F statistic over all choices of m :

$$F(\hat{m}) = \max_{1 \leq m \leq T} F(m)$$

Thus the Sup F test can be interpreted as a two stage test; at the first stage we estimate the most likely changepoint $\hat{m}$ and at the second we apply the Chow test at $\hat{m}$.

The problem with Sup F test and also the other tests considering the change point endogenously is that change point appears only under the alternative hypothesis but not under the null hypothesis as a parameter. Asymptotic analysis of such problems can be found in Davies (1977, 1987), Andrews and Ploberger (1991), Hansen(1991) and King and Shively (1993). They show that the asymptotic distributions differ from the standard ones. Andrews (1993) determines the asymptotic distributions of

Sup Wald (Sup W), Sup Lagrange Multiplier (Sup LM) and Sup Likelihood Ratio (Sup LR) test statistics under the null hypothesis of parameter stability and under the alternative hypothesis of parameter instability including one time structural change. Since Sup W, Sup LM and Sup LR test statistics are extensions of Sup F test statistic, same asymptotic distribution applies for Sup F test statistic also. Moreover, Andrews (1990) compares this test with tests such as Cusum and Cusum of squares of Brown, Durbin and Evans (1975) and fluctuation test of Sen (1980) and Ploberger, Krämer and Kontrus (1989) in terms of power. He concludes that Sup F test is more powerful than all of the above stated tests.

Seber and Wild (1989) show that under the null hypothesis of parameter stability, the finite sample null distribution of the Sup F statistic does not depend on the parameters β and σ^2 . The null distribution does depends on the matrix X of explanatory variables in finite samples (but not asymptotically). Therefore, it is possible to simulate the null distribution of Sup F for any particular data set and arbitrary β and σ^2 values. The critical value for α % significance level can be found from this simulation. The hypothesis can be rejected if Sup F is greater than this value.

4.2 The Bayesian Approach

As discussed earlier in part I, there is useful ancillary information in simpler forms of the problem (see Cobb(1978)). Finding and conditioning on a suitable ancillary appears complex. Instead, we plan to use a Bayesian approach to automatically incorporate the ancillary information. If we want the Bayesian approach to mimic a frequentist test, we need to use uninformative priors. The following lemma is basic to the Bayesian calculations with uninformative priors:

Lemma 1 Suppose the density $L(y|\beta, \sigma^2)$ is $y|\beta, \sigma^2 \sim N(X\beta, \sigma^2 \mathbf{I}_T)$ and the (improper) prior on β, σ^2 is $\pi(\beta, \sigma^2) = \sigma^{-d}$. Define $\hat{\beta} = (X'X)^{-1}X'y$ and $SSR = \|y - X\hat{\beta}\|^2$. Then the marginal density $m(y) = \int L(y|\beta, \sigma^2)\pi(\beta, \sigma^2) d\beta d\sigma$ is:

$$\begin{aligned} m(y) &= \int_0^\infty \int_{R^k} (2\pi\sigma^2)^{-T/2} \exp(-\|y - X\beta\|^2/2\sigma^2) \\ &= c SSR^{-(T-k+d-1)/2}, \end{aligned} \quad (2)$$

where $c = 2^{(d-3)/2}(\pi)^{-(T-k)/2} \det(X'X)^{1/2} \Gamma((T-k+d-1)/2)$.

Proof: Note that $\|y - X\beta\|^2 = SSR + \|X(\beta - \hat{\beta})\|^2$, and hence:

$$\int L(y|\beta, \sigma^2) d\beta = (2\pi\sigma^2)^{-T/2} \exp(-SSR/2\sigma^2) \int \exp(-\|X(\beta - \hat{\beta})\|^2/2\sigma^2) d\beta$$

The integral is now that of a k dimensional normal density for β , with mean $\hat{\beta}$ and covariance $\sigma^2(X'X)^{-1}$. Multiplying inside the integral by the required normalizing constant, and dividing by the same constant outside the integral yields the following result:

$$\int L d\beta = (2\pi\sigma^2)^{-(T-k)/2} \det(X'X)^{-1/2} \exp(-SSR/2\sigma^2)$$

Let $A = (2\pi)^{-(T-k)/2} \det(X'X)^{-1/2}$ and let $a = (T-k+d-1)/2$. From the previous equation, we get

$$\begin{aligned} \int \int L d\beta d\sigma / \sigma^d &= A \int (\sigma^2)^{-a} \exp(-SSR/2\sigma^2) d\sigma / \sigma \\ &= A 2^{a-1} SSR^{-a} \Gamma(a) \end{aligned}$$

The second line follows after making a change of variables to $u = SSR/2\sigma^2$. Collecting terms completes the proof of the Lemma.

To determine suitable priors for our Bayesian test, first consider the problem of discriminating between the null of no change, and the alternative $H(m)$ of a changepoint at a fixed point of time m . In order to be able to apply lemma 1

to the changepoint model, note that the hypothesis of a changepoint at m can be represented as a standard regression model $Y = X(m)\beta_m + \epsilon$ with $\beta'_m = (\beta'_1, \beta'_2)$ being the $2k \times 1$ vector of regression coefficients and $X(m)$ design matrix being defined as:

$$X(m) = \begin{pmatrix} X_m & 0 \\ 0 & X^m \end{pmatrix}$$

Define $SSR(m) = SSR_m + SSR^m$ to be the sum of squared residuals in this regression model, with a changepoint at m . Assume the prior density is $\pi_0(\beta_0, \sigma) = \sigma^{-d_0}$ under the null, and $\pi_1(\beta_1, \beta_2, \sigma) = \sigma^{-d_1}$ under the alternative. Let p_0 be the prior probability of the null being true. Let the alternative hypothesis be that change occurs at some m with $m_0 \leq m \leq m_1$, so that $H_1 = \cup_{m=m_0}^{m_1} H(m)$. Let p_m be the prior probabilities of $H(m)$. Moreover assume that π_m is the prior probability of a change at point m . The posterior probability of no change is given by Bayes formula:

$$\begin{aligned} P(H_0|y) &= \frac{P(y|H_0)p_0}{P(y|H_0)p_0 + \sum_{m=m_0}^{m_1} \pi_m P(y|H_m)p_m} \\ &= \frac{c_0 p_0 SSR^{-(T-k+d_0-1)/2}}{c_0 p_0 SSR^{-(T-k+d_0-1)/2} + \sum_{m=m_0}^{m_1} \pi_m c_1 p_m SSR(m)^{-(T-2k+d_1-1)/2}} \end{aligned} \quad (3)$$

The posterior probability of alternative hypothesis that a change occurs at some m is given by Bayes formula:

$$\begin{aligned} P(H_1|y) &= \frac{\sum_{m=m_0}^{m_1} \pi_m P(y|H_m)p_m}{P(y|H_0)p_0 + \sum_{m=m_0}^{m_1} \pi_m P(y|H_m)p_m} \\ &= \frac{\sum_{m=m_0}^{m_1} \pi_m c_1 p_m SSR(m)^{-(T-2k+d_1-1)/2}}{c_0 p_0 SSR^{-(T-k+d_0-1)/2} + \sum_{m=m_0}^{m_1} \pi_m c_1 p_m SSR(m)^{-(T-2k+d_1-1)/2}} \end{aligned} \quad (4)$$

The second equalities for both of the cases follow from the use of Lemma 1.

The usual F -test is monotonic in the ratio $SSR/SSR(m)$. For the Bayes test to have the same property, it is essential that $d_1 = d_0 + k$. In this case, both SSR and $SSR(m)$ occur with the same power, and the Bayes statistic is monotonic in the ratio. A natural way to achieve this is to let $d_0 = k$ and let $d_1 = 2k$, so that the Bayes

prior is flat on β and σ is raised to negative the power of the dimension of the space of coefficients. $\det(X(m)'X(m))$ which will appear in c_1 is $\det(X_m'X_m)\det(X^mX^m)$.

Note that the posterior probabilities of the hypotheses $H(m)$ receive a factor 2^k as opposed to $2^{k/2}$ for the null, so that this formulation of the prior favors the higher dimensional models of the alternative hypothesis. The term $\det(X(m)'X(m))$ is the product of the precision matrices for the OLS estimators of the two regimes in the changepoint model. The formula above shows that if posterior probabilities favor those models for which the OLS estimates are precise, relative to those which are not. The third factor is the Sum of Squared Residuals, which is the only factor taken into account by classical approaches to the changepoint problem.

In some of the applied econometric works there is some idea about the distribution of the parameters. For this reason it is important to obtain a test statistic where we have proper priors for the parameters. The following lemma is basic to Bayesian calculations with informative priors:

Lemma 2 Suppose the density $L(y|\beta, \tau)$ is $y|\beta, \tau \sim N(X\beta, \tau I_T)$, the prior on β is $\pi(\beta|\tau) \sim N(\mu, \tau\Sigma)$ and the prior on τ is $\pi(\tau) \sim IG(\lambda, \alpha)$. Define $\hat{\beta} = (X'X)^{-1}X'y$, $\Sigma_{(p)}^{-1} = X'X + \Sigma^{-1}$, $\mu_{(p)} = \Sigma_{(p)}(X'y + \Sigma^{-1}\mu)$ and $A = \|y - X\hat{\beta}\|^2 + \|X\hat{\beta}\|^2 + \mu'\Sigma^{-1}\mu - \mu_{(p)}'\Sigma_{(p)}^{-1}\mu_{(p)}$. Then the marginal density $m(y) = \int L(y|\beta, \tau)\pi(\beta|\tau)\pi(\tau)d\beta d\tau$ is:

$$\begin{aligned} m(y) &= \int_0^\infty \int_{R^k} L(y|\beta, \tau)\pi(\beta|\tau)\pi(\tau)d\beta d\tau \\ &= \frac{\det(\Sigma)^{-1/2}}{(2\pi)^{T/2}\det(\Sigma_{(p)})^{-1/2}} \frac{\lambda^\alpha \Gamma(T/2 + \alpha)}{(A/2 + \lambda)^{T/2 + \alpha} \Gamma(\alpha)} \end{aligned}$$

Proof:

$$L(y|\beta, \tau)\pi(\beta|\tau) = \frac{\det(\Sigma)^{-1/2}}{(2\pi\tau)^{(T+k)/2}} \exp \left\{ -\frac{1}{2\tau} [\|y - X\beta\|^2 + (\beta - \mu)'\Sigma^{-1}(\beta - \mu)] \right\}$$

Note that $\|y - X\beta\|^2 + (\beta - \mu)'\Sigma^{-1}(\beta - \mu) = A + (\beta - \mu_{(p)})'\Sigma_{(p)}^{-1}(\beta - \mu_{(p)})$, and hence

$$\int L(y|\beta, \tau)\pi(\beta|\tau)d\beta = \frac{\det(\Sigma)^{-1/2}}{(2\pi\tau)^{(T+k)/2}} \exp \left(-\frac{A}{2\tau} \right) \int \exp \left\{ -\frac{1}{2\tau} (\beta - \mu_{(p)})'\Sigma_{(p)}^{-1}(\beta - \mu_{(p)}) \right\} d\beta$$

The integral is now that of a k dimensional normal density for β , with mean $\mu_{(p)}$ and covariance $\tau\Sigma_{(p)}$. Multiplying inside the integral by the required normalizing constant, and dividing by the same constant outside the integral yields the following result:

$$\begin{aligned}\int L(y|\beta, \tau)\pi(\beta|\tau)d\beta &= \frac{\det(\Sigma)^{-1/2}}{(2\pi\tau)^{T/2}\det(\Sigma_{(p)})^{-1/2}} \exp\left(-\frac{A}{2\tau}\right) \\ \int L(y|\beta, \tau)\pi(\beta|\tau)d\beta\pi(\tau) &= \frac{\det(\Sigma)^{-1/2}}{(2\pi\tau)^{T/2}\det(\Sigma_{(p)})^{-1/2}} \frac{\lambda^\alpha}{\Gamma(\alpha)\tau^{\alpha+1}} \exp\left(-\frac{A}{2\tau} - \frac{\lambda}{\tau}\right) \\ \int \int L(y|\beta, \tau)\pi(\beta|\tau)\pi(\tau)d\beta d\tau &= \frac{\det(\Sigma)^{-1/2}}{(2\pi)^{T/2}\det(\Sigma_{(p)})^{-1/2}} \frac{\lambda^\alpha}{\Gamma(\alpha)} \int \frac{1}{\tau^{T/2+\alpha+1}} \exp\left(-\frac{A/2+\lambda}{\tau}\right) d\tau\end{aligned}$$

The integral is now that of an inverse gamma density for τ , with parameters $\frac{T}{2} + \alpha$ and $\frac{A}{2} + \lambda$. Multiplying inside the integral by the required constant and dividing by the same constant outside the integral yields the following result:

$$\int \int L(y|\beta, \tau)\pi(\beta|\tau)\pi(\tau)d\beta d\tau = \frac{\det(\Sigma)^{-1/2}}{(2\pi)^{T/2}\det(\Sigma_{(p)})^{-1/2}} \frac{\lambda^\alpha \Gamma(T/2 + \alpha)}{(A/2 + \lambda)^{T/2+\alpha} \Gamma(\alpha)}$$

This completes the proof.

Assume the following prior distributions: $\pi(\beta_0) \sim N_k(\mu_0, \sigma_0^2 \Sigma_0)$, $\pi(\beta_i) \sim N_k(\mu_i, \sigma_1^2 \Sigma_i)$, $i = 1, 2$, $\pi(\sigma_0^2|H_0) \sim IG(\lambda_0, \alpha_0)$ and $\pi(\sigma_1^2|H_1) \sim IG(\lambda_1, \alpha_1)$. In order to be able to apply lemma 2 to the change point model, some further definitions must be made. Define $A(m) = A_m + A^m$. The covariance matrix for β_m is defined as:

$$\Sigma_m = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$$

Following lemma 2 the marginal density of y under the null hypothesis will be

$$P(y|H_0) = \frac{\det(\Sigma_0)^{-1/2}}{(2\pi)^{T/2}\det(\Sigma_{0(p)})^{-1/2}} \frac{\lambda_0^{\alpha_0} \Gamma(T/2 + \alpha_0)}{(A/2 + \lambda_0)^{T/2+\alpha_0} \Gamma(\alpha_0)},$$

and the marginal density of y under the alternative hypothesis will be

$$P(y|H_1) = \frac{\lambda_1^{\alpha_1} \Gamma(T/2 + \alpha_1)}{(2\pi)^{T/2} \Gamma(\alpha_1)} \sum_{m=m_0}^{m=m_1} \frac{\det(\Sigma_m)^{-1/2}}{\det(\Sigma_m(p))^{-1/2}} \frac{1}{(A(m)/2 + \lambda_1)^{T/2+\alpha_1}}.$$

The posterior probability of no change is given by Bayes formula:

$$P(H_0|y) = \frac{P(y|H_0)p_0}{P(y|H_0)p_0 + \sum_{m=m_0}^{m_1} \pi_m P(y|H_m)p_m} \quad (5)$$

The posterior probability of alternative hypothesis that a change occurs at some m is given by Bayes formula:

$$P(H_1|y) = \frac{\sum_{m=m_0}^{m_1} \pi_m P(y|H_m)p_m}{P(y|H_0)p_0 + \sum_{m=m_0}^{m_1} \pi_m P(y|H_m)p_m} \quad (6)$$

Using posterior odds as a test of structural change is suggested before in the literature. (For example see Zellner (1987)) We also obtain our test for the unknown change point case by calculating the posterior odds in favour of the null hypothesis. Using equations (3) and (4) the posterior odds under the assumption of improper priors can be written as:

$$\begin{aligned} K_0^I &= \frac{P(H_0|y)}{P(H_1|y)} \\ &= \frac{P(y|H_0)p_0}{1/(m_1 - m_0 + 1) \sum_{m=m_0}^{m_1} P(y|H_m)p_m} \\ &= \frac{c_0 p_0 SSR^{-(T-1)/2}}{1/(m_1 - m_0 + 1) \sum_{m=m_0}^{m_1} c_1 p_m SSR(m)^{-(T-1)/2}} \end{aligned} \quad (7)$$

where

$$c_0 = 2^{(k-3)/2} (\pi)^{-(T-k)/2} \det(X'X)^{1/2} \Gamma((T-1)/2)$$

and

$$c_1 = 2^{(2k-3)/2} (\pi)^{-(T-2k)/2} \det(X'X)^{1/2} \Gamma((T-1)/2).$$

We assume that the prior distribution of the change point π_m is uniform over the possible range.

Using equations (5) and (6) the posterior odds under the assumption of proper priors can be written as:

$$\begin{aligned}
K_0^P &= \frac{P(H_0|y)}{P(H_1|y)} \\
&= \frac{P(y|H_0)p_0}{1/(m_1 - m_0 + 1) \sum_{m=m_0}^{m_1} P(y|H_m)p_m} \\
&= \frac{p_0 \lambda_0^{\alpha_0} \Gamma(T/2 + \alpha_0) \Gamma(\alpha_1) (m_1 - m_0 + 1)}{p_m \lambda_1^{\alpha_1} \Gamma(T/2 + \alpha_1) \Gamma(\alpha_0)} \\
&\quad \times \sum_{m=m_0}^{m_1} \frac{|\Sigma_0|^{-1/2} |\Sigma_m(p)|^{-1/2} (A(m)/2 + \lambda_1)^{T/2 + \alpha_1}}{|\Sigma(m)|^{-1/2} (A/2 + \lambda_0)^{T/2 + \alpha_0}}
\end{aligned} \tag{8}$$

K_0^I and K_0^P are two Bayesian test statistics for testing structural change with unknown change point where improper and proper priors are assumed respectively. As can be seen they are explicitly given and so they are ready to use in applied work. Although the formulas may seem complicated the constant terms can be omitted while doing the calculations since they would not effect the result. The applied econometrician should calculate the critical value specific to his data by a small computer program ⁴ and then easily make the test for structural change. Since these tests belong to the optimal class of tests of Andrews, Lee and Ploberger (1996) putting this little effort will improve the results. It is better to use these tests than using other tests existing in the literature such as the rolling Chow test.

4.3 An Application

In this section we provide an example that illustrates how one may test for the existence of structural change using the Bayesian tests derived above. The data that is used in the example is İstanbul Stock Exchange (ISE) composite index weekly returns for the period of January 7, 1993 to December 29, 1994. For this period there are 104 return observations. The returns are calculated as the first difference of the natural logarithms of the ISE index.

⁴The program is available from the author upon request.

The period selected includes the crash of January 1994 related to the macroeconomic instability in that period. Our aim is to detect whether the crisis gave rise to a change in the mean of the drift in log stock prices. To that end, ignoring heteroscedasticity and the slight non-normality in the data, we model the return series as an independently and normally distributed noise term with occational changes in mean. Formaly we can write

$$R_t = \beta_{0_t} + \epsilon_t$$

where R_t is the return at time t , β_{0_t} is the mean valid at time t and $\epsilon_t \sim_{iid} N(0, \sigma^2)$.

For this model we test the following null hypothesis against the alternative:

$$H_0 : \beta_{0_t} = \beta_0 \quad t \in \{1, \dots, T\}$$

$$H_1 : \beta_{0_t} = \beta_{0_1} \quad t \in \{1, \dots, m\}$$

$$\beta_{0_t} = \beta_{0_2} \quad t \in \{m+1 \dots T\}$$

where $m \in \{1, \dots, T-1\}$ is an unknown change point. For testing the hypothesis we use the three tests. The first one is the rolling Chow test defined in section 6.1. The second one is the Bayesian test with improper priors defined by the equation (7). The priors for this test are as in Lemma 1 where for this case $d_0 = 1$ and $d_1 = 2$. The last test is the Bayesian test with proper priors defined in equation (8). The priors of the test are as in Lemma 2. While choosing the parameters of the distributions we used our knowledge about the nature of the data. We chose $\mu = 0.015$ which is the mean of the returns for ISE. While choosing λ and α we tried to make the variance τ as small as possible. Due to numerical reasons there was a lower limit for our choice. We therefore took $\lambda = 8000$ and $\alpha = 0.01$. These values correspond to a mean of 0.0125 for the variance hyperparameter in the proper distribution of

the return mean, μ .

In order to be able to perform the tests we firstly need to find the critical values. For this aim we use bootstrap and have 10000 iterations. We find 5, 10, 20 and 40 percent significance level critical values. They are presented in table 6. The value of Max F test is 3.4685416. Comparing this value with the critical values in table 6 gives us the result that Max F test does not reject the null hypothesis. The value of Bayesian tests with improper and proper priors are 0.042069272 and 0.063012933 respectively. Comparing with the critical values we again can not reject the null hypothesis. Depending on the results of these tests one can conclude that there is no structural change in the data.

In table 7 we also provide the powers of the tests. To obtain those powers we again use bootstrap with 10000 iterations. While forming our alternative model we assume that there is a change in the return mean at observation 55. This corresponds to the 1994 crisis period. As can be seen from table 7, the powers of both of the Bayesian tests are higher than the power of the rolling Chow test. This is what is to be expected since Bayesian tests belong to the optimal class of tests defined in Andrews, Lee and Ploberger (1996). On the other hand, except for the 20 percent significance level case power of the Bayesian test with improper priors is higher than the Bayesian test with proper priors but the values are close to each other. Depending on these results we can not say that one is better than the other. In fact since the choice of parameters effect the performance of the Bayesian test with proper prior distributions, it is possible to increase its performance.

5 Conclusion

In this dissertation we dealt with some model selection problems. After the introductory chapter 1, in chapter 2 we compared finite sample properties of some model selection criteria for estimating true lag order of AR processes. As the basis for the comparison we took probability of correctly estimating the lag order and mean squared forecast errors of the models chosen by each of the criteria. In the chapter we tried to examine the effects of being parsimonious while selecting the model. Moreover, we also studied how different distributions effect the performance of the model selection criteria. In chapter 3, in addition to some of the widely used criteria in the literature we included some new criteria that we have developed, to our comparison. We used the idea of replacing the usual estimate of variance by different forms of cross-validated estimates of variance in the functions of conventional criteria. In chapter 4, we dealt with structural change problems for the unknown change point situation. We followed the Bayesian approach and developed Bayesian tests under different assumptions about the prior distributions. We also have included an application of these tests to İstanbul Stock Exchange weekly data where we compared the tests with the popular rolling Chow test.

In Chapter 2 we had six criteria under consideration, namely PRESS, AIC, SC, HQC, AIC_C and sequential F. In a Monte Carlo study, we calculated the frequency distribution of the estimated lag order for each criteria under the assumption of normality for the model. The simulation results have shown that PRESS and AIC_C have an overestimation problem implying that they are not parsimonious in model selection. Due to this character of these criteria, mean squared forecast errors obtained from the models chosen by them are very high. Since putting a zero coefficient rather than the OLS estimate in the model usually ends up with better forecasts,

performance of parsimonious criteria have been observed to be better.

In the chapter we also relaxed the assumption of normality for the model which has always been made in previous theoretical and empirical studies in this field. It is possible that some of the time series data, such as the ones from financial markets may possess skewness and kurtosis. For this reason, it is important to have an idea about the performance of model selection criteria under these situations. We provided this information in the Chapter for the criteria named above. Simulation results have shown that although skewness does not effect the performance of the criteria under consideration kurtosis has some effect on some of the criteria. While performance of AIC_C in correctly estimating the true lag order is observed to be very poor for heavy-tailed distributions, it quickly improved as we moved to normality. On the other hand, PRESS criterion and sequential F tests performed well for heavy-tailed distributions but we saw a decrease in their performance as we moved to normality. This decrease is more clear for PRESS criterion. Other criteria under consideration seemed not to be effected from kurtosis. Sample size had some effect on the criteria as well. Therefore if the aim is to estimate the lag order correctly for small sample sizes and heavy-tailed distributions we suggest the use of either PRESS or sequential F. On the other hand, if one is not sure on the presence of a heavy-tailed distribution, we definitely suggest the use of AIC_C for small sample sizes. For large sample sizes SC and HQC are the criteria with the best performance. If the aim is to make a good forecast, under skewed distribution case and for small sample sizes we suggest the use of SC but if the sample size is large it is better to use HQC while selecting the model. When the data possesses kurtosis any of the criteria under consideration can be used since their forecasting performances are almost the same.

In Chapter 3 we suggested two different criteria for model selection problems. These criteria are obtained by replacing the usual estimate of variance in the functions of AIC and AIC_C by a cross-validated estimate. We were motivated by Efron's 1983 paper where it is shown that a cross-validated estimate is more robust to specification errors. Since PRESS/(T-K) is the cross-validated estimate of variance putting it in the functions of AIC and AIC_C automatically adds a penalty factor to PRESS and solves the over estimation problem of PRESS. Simulation results of Chapter 3 clearly showed this effect. In the chapter we also compared the performance of these two criteria with the other six. Before the advancements in the computer technology it was not possible to make such comparisons since cross-validation technique is very much computer intensive. If the aim is to correctly estimate the lag order, for small sample sizes the new criteria does not have an improvement over AIC and AIC_C but for large sample sizes the new criteria are definitely much more better. If the aim is to make good forecasts we would suggest the use of our new criteria for both small and large sample sizes. If there is a skewed distribution their performance is better and if there is a t distribution they perform as good as the other criteria.

PRESS deletes one of the observations while making the calculations. There are other cross-validation techniques where more observations can be deleted in the process of calculations. In the chapter we also considered the case where ten percent of the data is deleted. Another cross-validated estimate of variance obtained by this way is substituted in place of the usual estimate of variance in the function of AIC. So in addition to the two suggested criteria we suggested one more criterion. Previous simulation results showed that there is sample size effect on the criteria so we varied sample size and tried to see the effects on the criteria. We considered normal distribution case and a skewed distribution case where degrees of freedom

was taken as 1. For sample sizes approximately greater than 100 both of the criteria obtained by cross-validated estimates of variance have performed better than AIC in terms of both correctly estimating the lag order and forecasting. Normally deleting ten percent of the data gave better results than deleting just one observation. Also in terms of computer time of calculations cross-validation based on omitting ten percent of the data is more efficient. However for sample sizes smaller than 100 deleting ten percent of the observations is observed to result in fluctuations in the performance. For this reason for small samples it is better to use PRESS which only deletes one observation. For sample sizes less than 60 AIC performs better than the modified versions of it by cross-validation.

Here we present two summary tables depending on the simulation results of chapters 2 and 3 where we report our suggestions about which criteria to use under different conditions. Specifically, we consider the three cases where the model is generated from a normal distribution, a heavy-tailed distribution and a skewed distribution. Moreover, we also consider the sample size effect. We report our suggestions for small sample sizes and large sample sizes. In table 8, one can find the suggested criteria if the aim is to estimate the lag order of an AR model. On the other hand, if the aim is to make a good forecast, one can find the suggested criteria in table 9.

In chapter 4, we considered the problem of testing structural change when the change point is unknown. We derived two different Bayesian tests for this aim. For the first one we used improper priors and for the second one we used proper priors. Since these tests belong to the optimal class of tests mentioned in Andrews, Lee and Ploberger (1996) it is expected that they will outperform other tests. Unfortunately, they do not have a known distribution but again thanks to advancements in computing technology, it became very easy to calculate critical values for the data in

hand by either Monte-Carlo or Bootstrap techniques and make the tests depending on that value.

In the chapter we also have included an application of these tests to İstanbul Stock Exchange weekly data. In addition to the two Bayesian tests we also applied the popular rolling Chow test. All of the three tests did not reject the null hypothesis of no structural change at both 5 percent and 10 percent significance level suggesting that there is no structural change in the data. In the chapter we also have a Monte Carlo study where we apply bootstrap, to obtain the powers of the tests. For both 5 percent and 10 percent significance levels powers of the Bayesian tests are higher than the power of the rolling Chow test. Among the Bayesian tests at 5 percent significance level the test with improper priors have higher power but for 10 percent case the test with proper priors have higher power.



Bibliography

1. Akaike, H., 1969, Fitting Autoregressive Models for Prediction, *Annals of the Institute of Statistical Mathematics*, 21, 243-247.
2. Akaike, H., 1971, Autoregressive Model Fitting for Control, *Annals of the Institute of Statistical Mathematics*, 23, 163-180.
3. Akaike, H., 1973, Information Theory and an Extension of the Maximum Likelihood Principle, 2nd International Symposium on Information Theory, edited by B. N. Petrov and F. Csàki, Budapest: Akadèmiai Kiadó, 267-281.
4. Akaike, H., 1974, A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control* AC-19, 716-723.
5. Allen, D. M., 1971, Mean Square Error of Prediction as a Criterion for Selecting Variables. *Technometrics* 13, 469-75.
6. Allen, D. M., 1974, The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction, *Technometrics* 16, 125-7.
7. Amemiya, T., 1980, Selection of Regressors, *International Economic Review*, 21, 331-345.
8. Andrews, D. W., 1993, Tests for parameter instability and structural change with unknown change point, *Econometrica*, 61, 4, 821-856.
9. Andrews, D. W. and W. Ploberger, 1994, Optimal tests when nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.
10. Andrews, D. W., I. Lee and W. Ploberger, 1996, Optimal Change Point Tests for Normal Linear Regression, *Journal of Econometrics*, 70, 9-38.

11. Brown, R. L., J. Durbin and J. M. Evans, 1975, Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society, Series B*, Vol. 37, pp.149-163.
12. Chin Choy, J. H. and L.D. Broemeling, 1980, Some Bayesian inferences for a changing linear model, *Technometrics*, 22, 1, 71-78.
13. Chow, G., 1960, Tests of the equality between two sets of coefficients in two linear regressions, *Econometrica*, 28, 561-605.
14. Cobb, G. W., 1978, The problem of the Nile: Conditional Solution to a change-point problem, *Biometrika*, 65, 2, 243-51.
15. Davies, R. B., 1977, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 64, 247-254.
16. Davies, R. B., 1987, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 74, 33-43.
17. Diebold, F. X., 1989, Forecast Combination and Encompassing: Reconciling Two Divergent Literatures, *International Journal of Forecasting*, 5, 4, 589-92.
18. Efron B., 1983, Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation, *Journal of the American Statistical Association*, 78, 382, 316-331.
19. Ferreira, P. E., 1975, A Bayesian analysis of a switching regression model: Known number of regimes, *Journal of American Statistical Association*, 70, 370-374.

20. 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.
21. Hansen, B. E., 1991, Inference when a nuisance parameter is not identified under the null hypothesis, Working Paper No. 296, Rochester Center for Economic Research, University of Rochester.
22. Hendry, D. F., 1995, *Dynamic Econometrics*, Advanced Texts in Econometrics, New York and Oxford: Oxford University Press.
23. Holbert, D., 1973, A Bayesian Analysis of Shifting Sequences with applications to Two-phase regression, Ph.D. dissertation, Oklahoma State University, Stillwater, Oklahoma.
24. Holbert, D. and L. D. Broemeling, 1977, Bayesian inference related to shifting sequences and two-phase regression, *Communications in Statistics*, A6, 3, 265-275.
25. Hurvich, C. M. and C. L. Tsai, 1989, Regression and Time Series Model Selection in Small Samples, *Biometrika* 76, 2, 297-307.
26. Hurvich, C. M. and C. L. Tsai, 1993, A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection, *Journal of Time Series Analysis*, 14, 3.
27. Judge, G. G., W. E. Griffiths, R. C. Hill and T. Lee, 1985, *The Theory and Practice of Econometrics* 2nd Edition, John Wiley & Sons, Inc., New York, Wiley Series in Probability and Mathematical Statistics.
28. King, M. L. and T. S. Shively, 1993, Locally optimal testing when a nuisance

parameter is present only under alternative, *Review of Economics and Statistics*, 75, 1-7.

29. Koreisha, S. G. and T. Pukkila, 1993, Determining the Order of a Vector Autoregression when the Number of Component Series is Large, *Journal of Time Series Analysis* 14, 1, 47-69.
30. Lütkepohl, H., 1985, Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process, *Journal of Time Series Analysis* 6, 1, 35-52.
31. Parzen, E., 1974, Some Recent Advances in Time Series Modelling, *IEEE Transactions on Automatic Control* AC-19, 723-730.
32. Parzen, E., 1977, Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes, in *Multivariate Analysis IV*, edited by P. R. Krishnaiah, Amsterdam: North-Holland, 283-295.
33. Ploberger, W., W. Krämer and K. Kontrus, 1989, A new test for structural stability in the linear regression model, *Journal of Econometrics*, 40, 307-318.
34. Quenouille, M., 1949, Approximate Tests of Correlation in Time Series, *Journal of the Royal Statistical Society Series B* 11, 18-84.
35. Quinn, B. G., 1980, Order Determination for a Multivariate Autoregression, *Journal of the Royal Statistical Society B* 42, 182-185.
36. Rissanen, J., 1978, Modeling by Shortest Data Description, *Automatica* 14, 465-471.
37. Schwarz, G., 1978, Estimating the Dimension of a Model, *The Annals of Statistics* 6, 461-464.

38. Seber, G.A.F. and C.J. Wild, 1989, Nonlinear Regression, Chapter 9, Wiley Series in Probability and Mathematical Statistic
39. Sen, P. K., 1980, Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52, 203-218.
40. Shibata, R., 1976, Selection of the Order of an Autoregressive Model by Akaike's Information Criterion, *Biometrika* 63, 117-126.
41. Shibata, R., 1980, Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process, *The Annals of Statistics* 8, 147-164.
42. Walls, R. C. and Weeks, D. L., 1969, A Note on the Variance of a Predicted Response in Regression, *American Statistician* 23 (3), 24-6.
43. Weiss, S. M and N. Indurkha, 1996, Selecting the Right-Sized Model for Prediction, *Applied Intelligence*, 6, 261-273.
44. Zaman. A., 1984, Avoiding Model Selection by the use of Shrinkage Techniques, *Journal of Econometrics* 25, 239-346.
45. Zellner, A., 1987, An introduction to Bayesian inference in econometrics, Robert E. Krieger Publishing Company Malabar, Florida.F

Appendix A

Table 1a: Frequency Distributions of the Estimated Lag Order and Mean Squared
Forecast Errors

LAG	PRESS	AIC	SCHWARZ	HQC	AICC	SEQF
0	15.2 (2.2)	27.6 (5.6)	47 (15)	32.8 (10.1)	21.3 (2.2)	31.8 (4.5)
1	17 (5.6)	20.4 (11.7)	23.1 (21.6)	22.4 (17.5)	22.7 (6.3)	19.9 (12.7)
2	19 (35.7)	19.8 (51.5)	18.9 (59.6)	20.6 (60.4)	23.9 (35.2)	20.4 (50)
3	14.9 (13.3)	11.5 (12.5)	5.7 (3.3)	9.3 (7.7)	14.2 (14.2)	5.8 (7.8)
4	13 (13.5)	7.7 (8.7)	2.5 (0.5)	5.3 (2.4)	9.3 (13.3)	7.3 (7.9)
5	9.7 (12.2)	4.7 (4.8)	1.3 (0)	3.7 (1)	4.6 (12.5)	6.8 (8.9)
6	11.2 (17.5)	8.3 (5.2)	1.5 (0)	5.9 (0.9)	4 (16.3)	8 (8.2)
MSFE	1.2964176 (1.0475164)	1.2805204 (1.0426257)	1.2508594 (1.0410533)	1.2658192 (1.0400394)	1.2639535 (1.0473658)	1.286568 (1.0452231)

The maximum possible lag order, $M = 6$ so each of the criterion can choose a lag order between 0 and

6. Each cell contains the percentage of times the criterion estimates the lag order defined for that row.

The percentages given without the bracket are for sample size 30 and with bracket are for sample size

200. The true lag order is 2.

Table 1b: Frequency Distributions of the Estimated Lag Order and Mean Squared
Forecast Errors

LAG	PRESS	AIC	SCHWARZ	HQC	AICC	SEQF
0	10.7 (1.8)	20.1 (5.8)	39.4 (20.2)	25.7 (11)	13.5 (2.2)	22.1 (3)
1	11.4 (3.8)	15.2 (7)	17.2 (14.9)	15.5 (10)	15.1 (3.9)	13.6 (6.4)
2	15.6 (10.9)	15.5 (17.9)	15 (23.3)	16.6 (21.5)	19.3 (11.1)	19.1 (19.4)
3	25.3 (30.9)	22.4 (40.1)	18.3 (38.4)	21.2 (44.1)	28.8 (32.4)	22.4 (43.6)
4	14.1 (17)	10.2 (13.2)	5.3 (2.8)	8.1 (8.4)	12.2 (16.4)	8.1 (8.5)
5	11 (15.9)	7.9 (8.3)	2.5 (0.2)	6.3 (2.6)	7.1 (14.6)	6.7 (8.9)
6	11.9 (19.7)	8.7 (7.7)	2.3 (0.2)	6.6 (2.4)	4 (19.4)	8 (10.2)
MSFE	2.7774073 (1.0508664)	2.703317 (1.0493574)	2.3900524 (1.058067)	2.6909891 (1.050398)	2.5821681 (1.0504079)	2.6393904 (1.0498533)

The maximum possible lag order, $M = 6$ so each of the criterion can choose a lag order between 0 and

6. Each cell contains the percentage of times the criterion estimates the lag order defined for that row.

The percentages given without the bracket are for sample size 30 and with bracket are for sample size

200. The true lag order is 3.

Table2: Probabilities of Overestimation

True Lag	PRESS	AIC	SCHWARZ	HQC	AICC	SEQF
2	48.8 (56.5)	32.2 (31.2)	11 (3.8)	24.2 (12)	32.1 (56.3)	27.9 (32.8)
3	37 (52.6)	26.8 (29.2)	10.1 (3.2)	21 (13.4)	23.3 (50.4)	22.8 (27.6)

Each cell contains the percentage of times the criterion estimates a lag order which is higher than the true lag order defined for that row. The percentages given without the bracket are for sample size 30 and with bracket are for sample size 200.

Table 3a: Frequency Distributions of the Estimated Lag Order and Mean Squared

Forecast Errors

LAG	PRESS	AIC	SCHWARZ	HQC	AICC	SEQF
0	33.7 (6.1)	50.7 (10.7)	70.6 (28.1)	57.2 (18.8)	40.2 (6.2)	65.1 (11.3)
1	32.9 (18.5)	28.6 (22.6)	20.5 (29.5)	27.2 (25.9)	32.4 (18.8)	22.3 (26.7)
2	33.4 (75.4)	20.7 (66.7)	8.9 (42.4)	15.6 (55.3)	27.4 (75)	12.6 (62)
MSFE	1.1544474 (1.0179495)	1.1622951 (1.0193658)	1.1688363 (1.0285697)	1.1625925 (1.0222913)	1.1555603 (1.0180726)	1.16524 (1.019797)

The maximum possible lag order, $M = 2$, which is also the true lag order. Each of the criterion can choose a lag order between 0 and 2. Each cell contains the percentage of times the criterion estimates the lag order defined for that row. The percentages given without the bracket are for sample size 30 and with bracket are for sample size 200.

Table 3b: Frequency Distributions of the Estimated Lag Order and Mean Squared
Forecast Errors

LAG	PRESS	AIC	SCHWARZ	HQC	AICC	SEQF
0	26.3 (2.7)	45.9 (6.2)	66.4 (20.7)	53.1 (11.5)	34.4 (2.9)	55.9 (5.3)
1	22.2 (4.3)	22.9 (7.8)	19.6 (14.2)	22.4 (10.6)	24 (4.4)	19 (8.8)
2	22.2 (17.3)	15.1 (21.5)	8.2 (24.1)	12 (23.5)	19.3 (17.8)	12 (24.5)
3	29.3 (75.7)	16.1 (64.5)	5.8 (41)	12.5 (54.4)	22.3 (74.9)	13.1 (61.4)
MSFE	1.2592056 (1.0242339)	1.2639075 (1.0265765)	1.27444 (1.0410616)	1.2669366 (1.0306172)	1.2583346 (1.024156)	1.2649815 (1.0271869)

The maximum possible lag order, $M = 3$, which is also the true lag order. Each of the criterion can choose a lag order between 0 and 3. Each cell contains the percentage of times the criterion estimates the lag order defined for that row. The percentages given without the bracket are for sample size 30 and with bracket are for sample size 200.

Table 4a: Frequency Distributions of the Estimated Lag Order and Mean Squared

Forecast Errors

LAG	PRESS	AICPRESS	$AIC_C PRESS$
0	15.2 (2.2)	55.6 (12.1)	47.8 (9)
1	17 (5.6)	23.7 (19.5)	25 (17)
2	19 (35.7)	16.1 (60.9)	20.6 (60.1)
3	14.9 (13.3)	3.1 (5.5)	4.7 (8.8)
4	13 (13.5)	1.4 (1.5)	1.8 (3.1)
5	9.7 (12.2)	0.1 (0.3)	0 (1.1)
6	11.2 (17.5)	0 (0.2)	0 (0.9)
MSFE	1.2964176 (1.0475164)	1.2396852 (1.0399644)	1.2278714 (1.0399849)

The maximum possible lag order, $M = 6$ so each of the criterion can choose a lag order between 0 and 6. Each cell contains the percentage of times the criterion estimates the lag order defined for that row. The percentages given without the bracket are for sample size 30 and with bracket are for sample size 200. The true lag order is 2.

Table 4b: Frequency Distributions of the Estimated Lag Order and Mean Squared Forecast Errors

LAG	PRESS	AICPRESS	$AIC_C PRESS$
0	10.7 (1.8)	48.8 (14.5)	41.9 (9.5)
1	11.4 (3.8)	18.5 (11.7)	19.6 (9.8)
2	15.6 (10.9)	15.7 (23.1)	17.4 (21.2)
3	25.3 (30.9)	14 (42.3)	17.4 (44.3)
4	14.1 (17)	2.4 (5.6)	3 (9.1)
5	11 (15.9)	0.6 (1.7)	0.7 (3.2)
6	11.9 (19.7)	0 (1.1)	0 (2.9)
MSFE	2.7774073 (1.0508664)	2.4721167 (1.0529307)	2.4643141 (1.0499285)

The maximum possible lag order, $M = 6$ so each of the criterion can choose a lag order between 0 and 6. Each cell contains the percentage of times the criterion estimates the lag order defined for that row. The percentages given without the bracket are for sample size 30 and with bracket are for sample size 200. The true lag order is 3.

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

Table5: Probabilities of Overestimation

True Lag	PRESS	AICPRESS	$AIC_C PRESS$
2	48.8 (56.5)	4.6 (7.5)	6.5 (3.9)
3	37 (52.6)	3 (8.4)	3.7 (15.2)

Each cell contains the percentage of times the criterion estimates a lag order which is higher than the true lag order defined for that row. The percentages given without the bracket are for sample size 30 and with bracket are for sample size 200.

Table 6: Critical Values of the Tests for Structural Change with Unknown Change

	Point			
	% 5	% 10	% 20	% 40
Max F	8.4851955	6.8230660	5.1841102	3.5391726
K_0^I	0.0077395627	0.013783390	0.022916494	0.035712655
K_0^P	0.063009112	0.063010387	0.063011571	0.063012625

Table 7: Powers of the Tests for Structural Change with Unknown Change Point

	% 5	% 10	% 20	% 40
Max F	0.3421	0.4651	0.6085	0.7733
K_0^I	0.3625	0.4929	0.6305	0.7884
K_0^P	0.3571	0.4838	0.6357	0.7881

Table 8: Suggested Criteria for Estimating the Lag Order of an AR Model

	Small Sample Size	Large Sample Size
Normal Distribution	AIC_C	AICPRESS, AIC_CPRESS , SC, HQC, AICCV10
Heavy-tailed Distribution	PRESS, SEQF, AIC_CPRESS ,	AICPRESS, SEQF, AIC, PRESS
Skewed Distribution	AIC_C PRESS	AICPRESS, AIC_CPRESS , AICCV10, HQC

Depending on the simulation results of the chapters 2 and 3, we can make the above suggestions. Each cell contains the criteria that perform best for the given case.

Table 9: Suggested Criteria for Forecasting

	Small Sample Size	Large Sample Size
Normal Distribution	AIC	AICPRESS, AICCV10
Heavy-tailed Distribution	ALL	ALL
Skewed Distribution	AICPRESS AIC_CPRESS	AICPRESS, AIC_CPRESS , AICCV10, HQC

Depending on the simulation results of the chapters 2 and 3, we can make the above suggestions. Each cell contains the criteria that perform best for the given case.

Figure 1: Value of PRESS for 10 Different Simulations (Sample Size = 200)

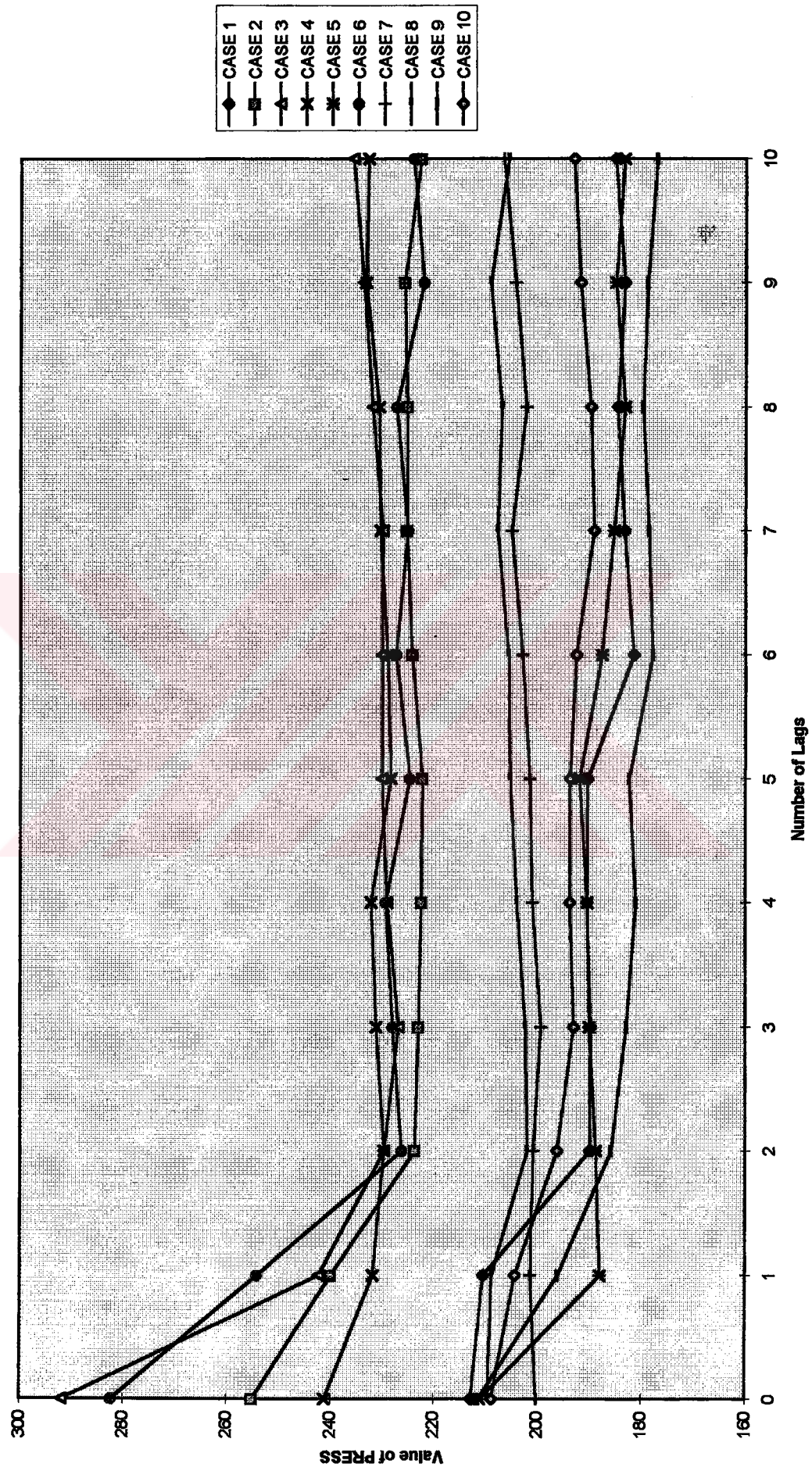


Figure 2a: Probability of Estimating the True Lag Order (True Lag Order = 2, Sample Size = 30, A Skewed Distribution)

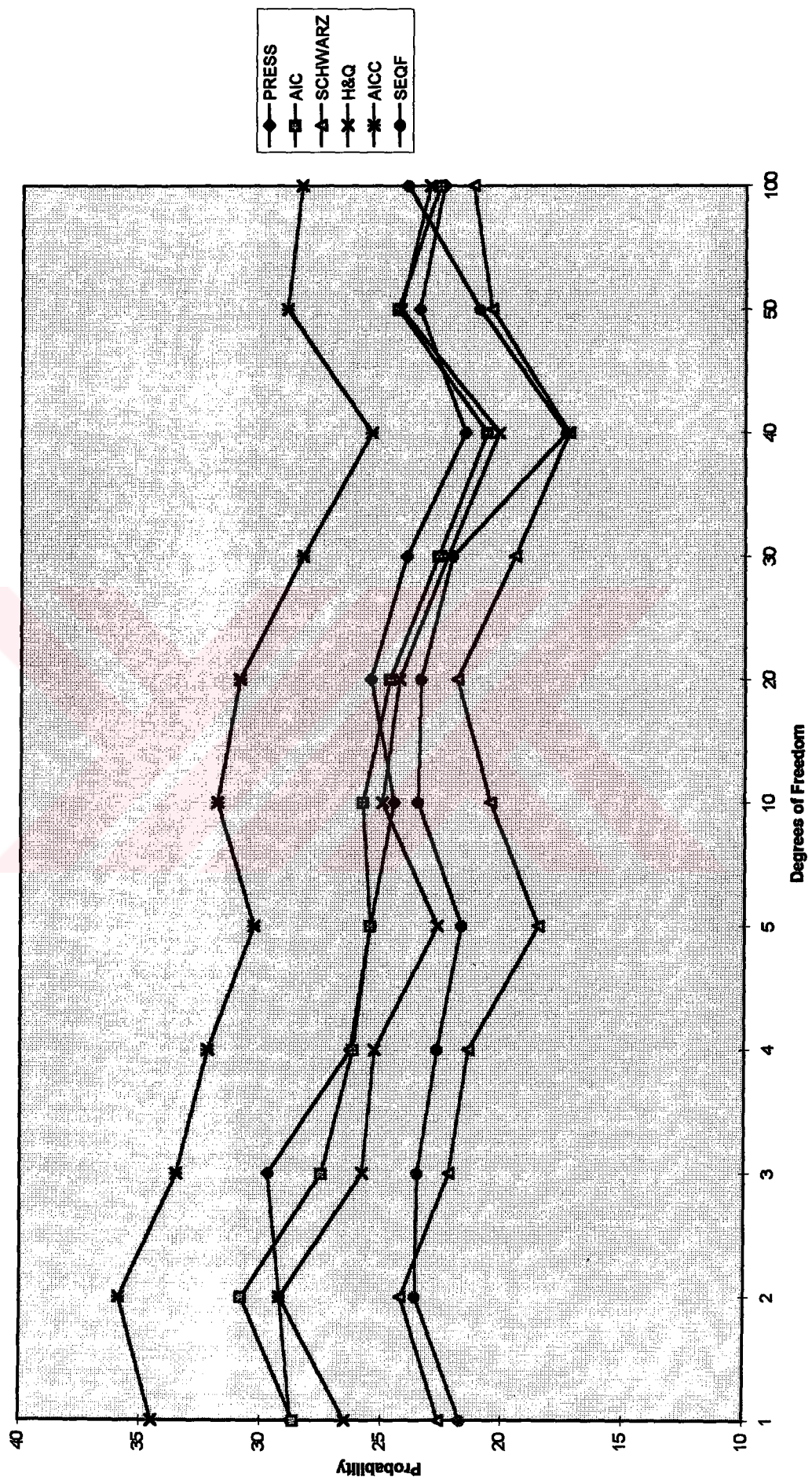


Figure 2b: Probability of Estimating the True Lag Order (True Lag Order = 2, Sample Size = 200, A Skewed Distribution)

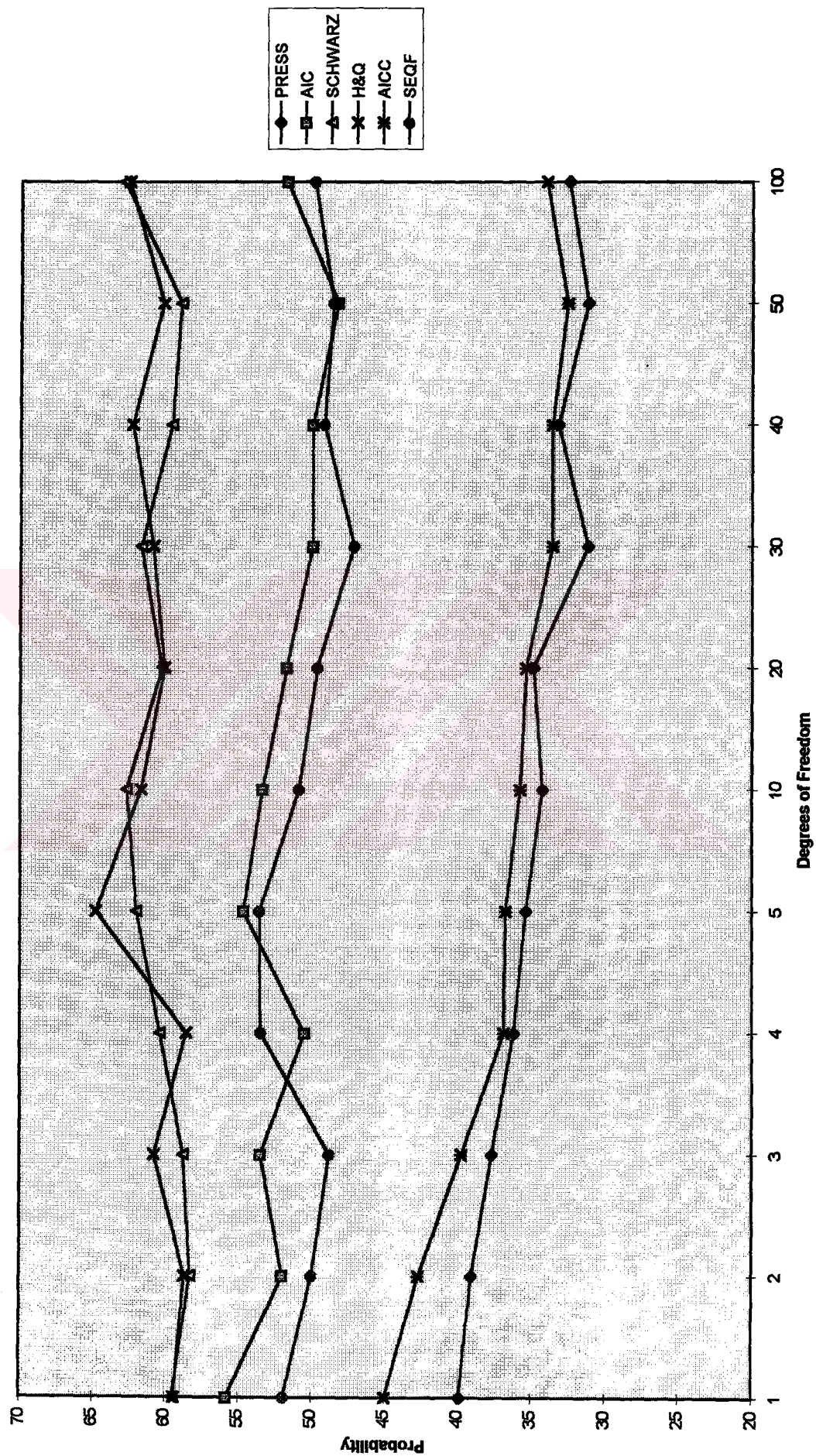


Figure 3a: Probability of Estimating the True Lag Order (True Lag Order = 2, Sample Size = 30, t Distribution)

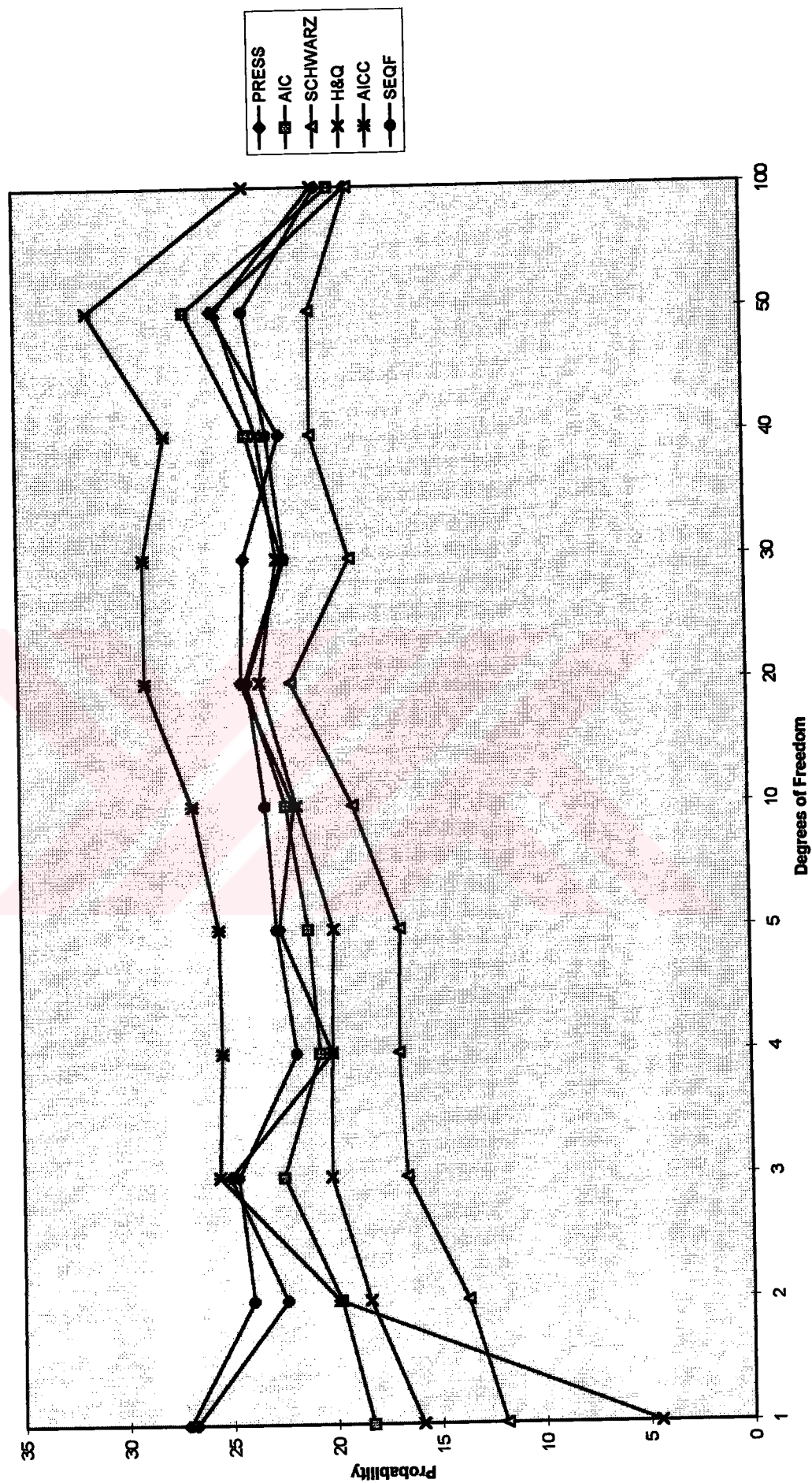


Figure 3b: Probability of Estimating the True Lag Order (True Lag Order = 2, Sample Size = 200, t Distribution)



Figure 4a: Probability of Estimating the True Lag Order (True Lag Order = 30, Sample Size = 30, A Skewed Distribution)

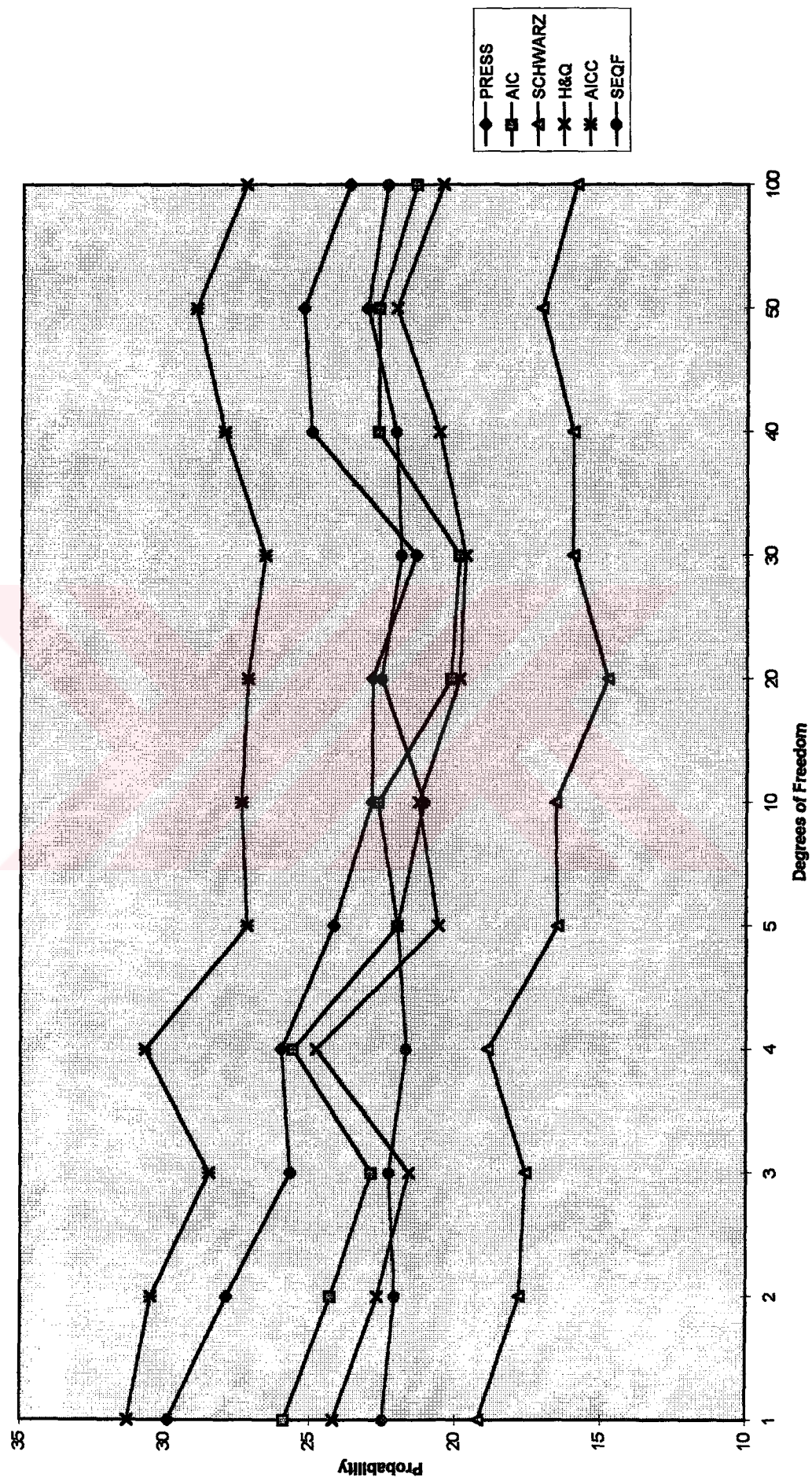


Figure 4b: Probability of Estimating the True Lag Order (Sample Size = 300, True Lag Order = 3, A Skewed Distribution)

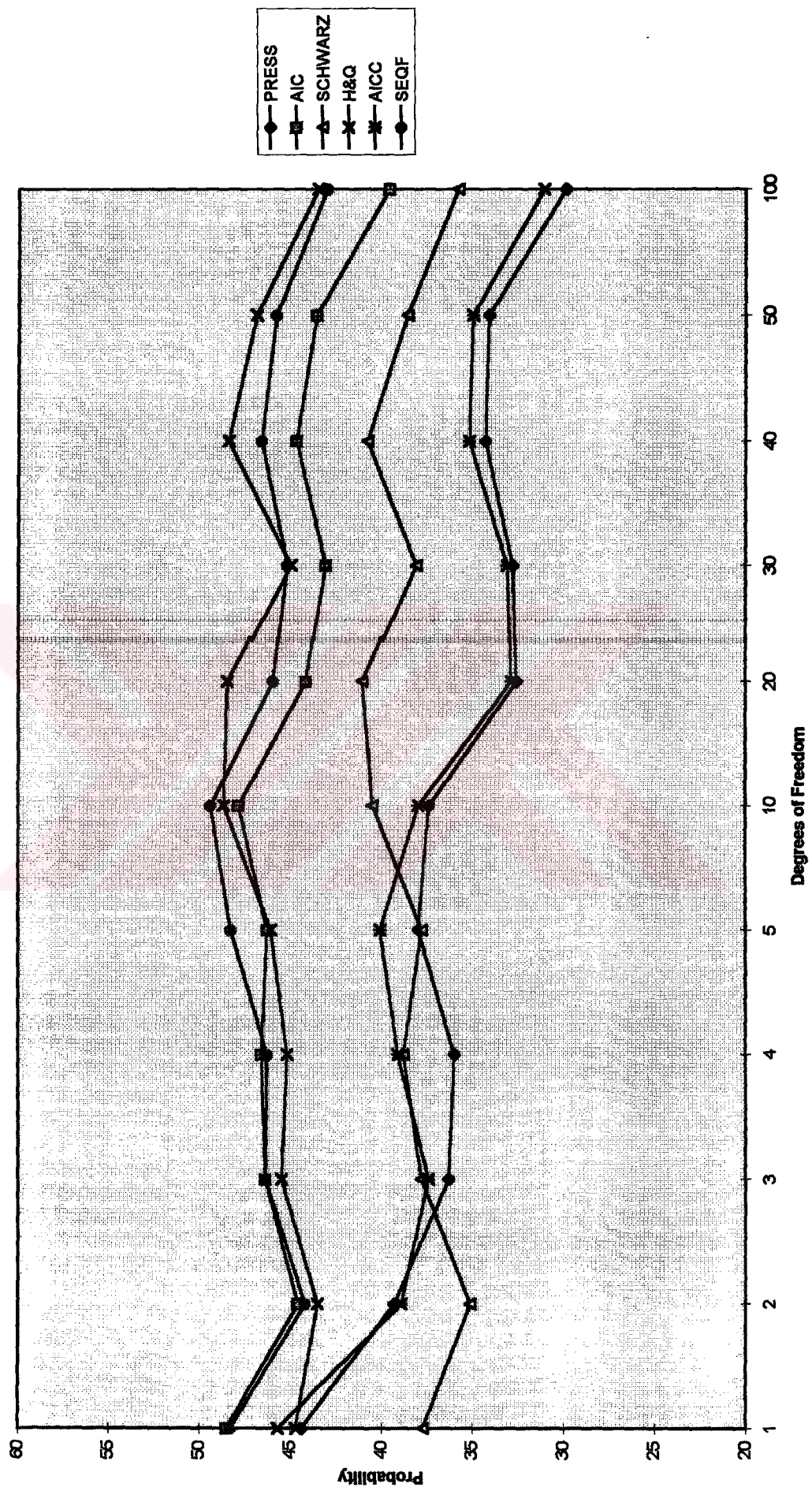


Figure 5a: Probability of Estimating the True Lag Order (True Lag Order = 3, Sample Size = 30, t Distribution)

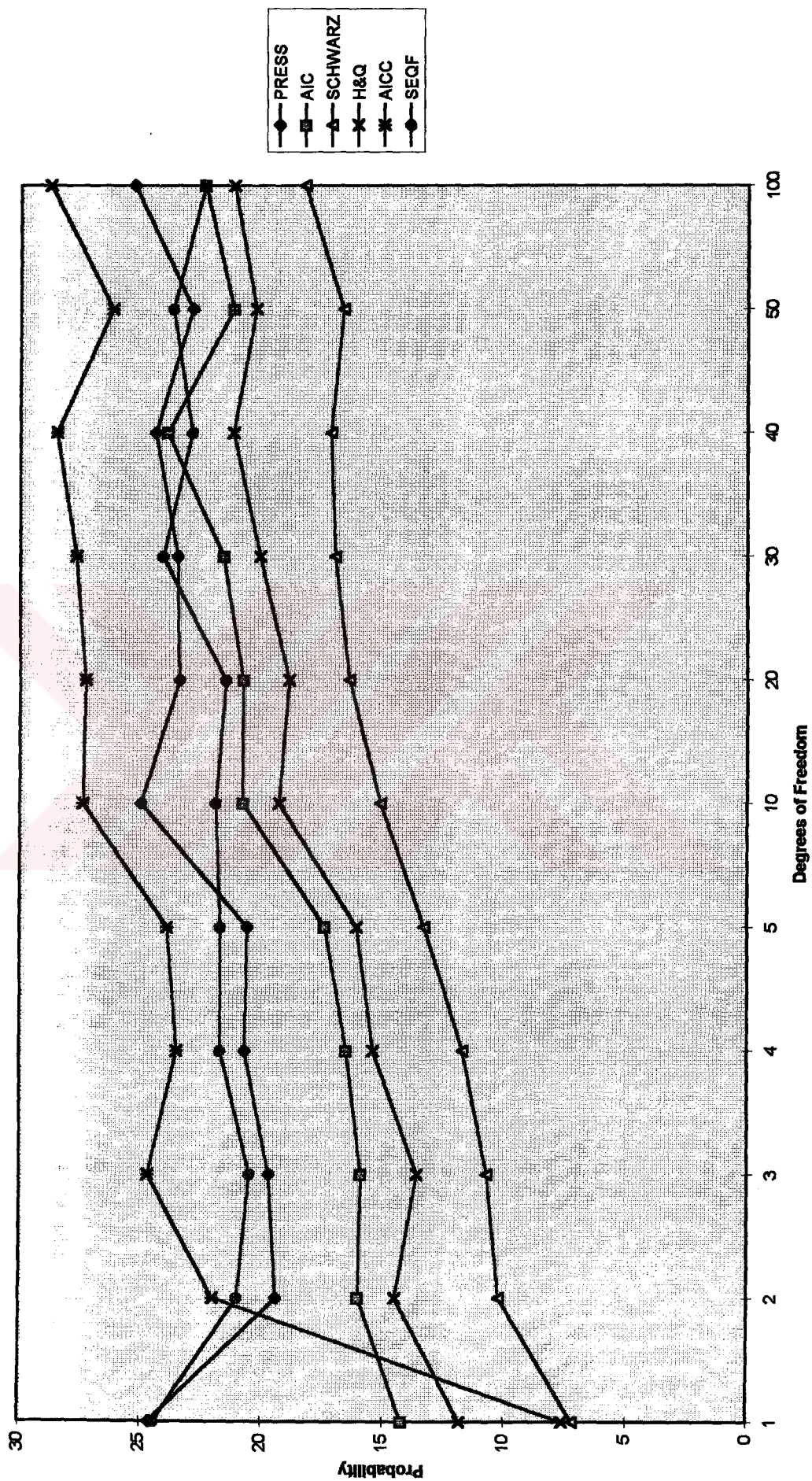


Figure 5b: Probability of Estimating the True Lag Order (True Lag Order = 3, Sample Size = 200, t Distribution)

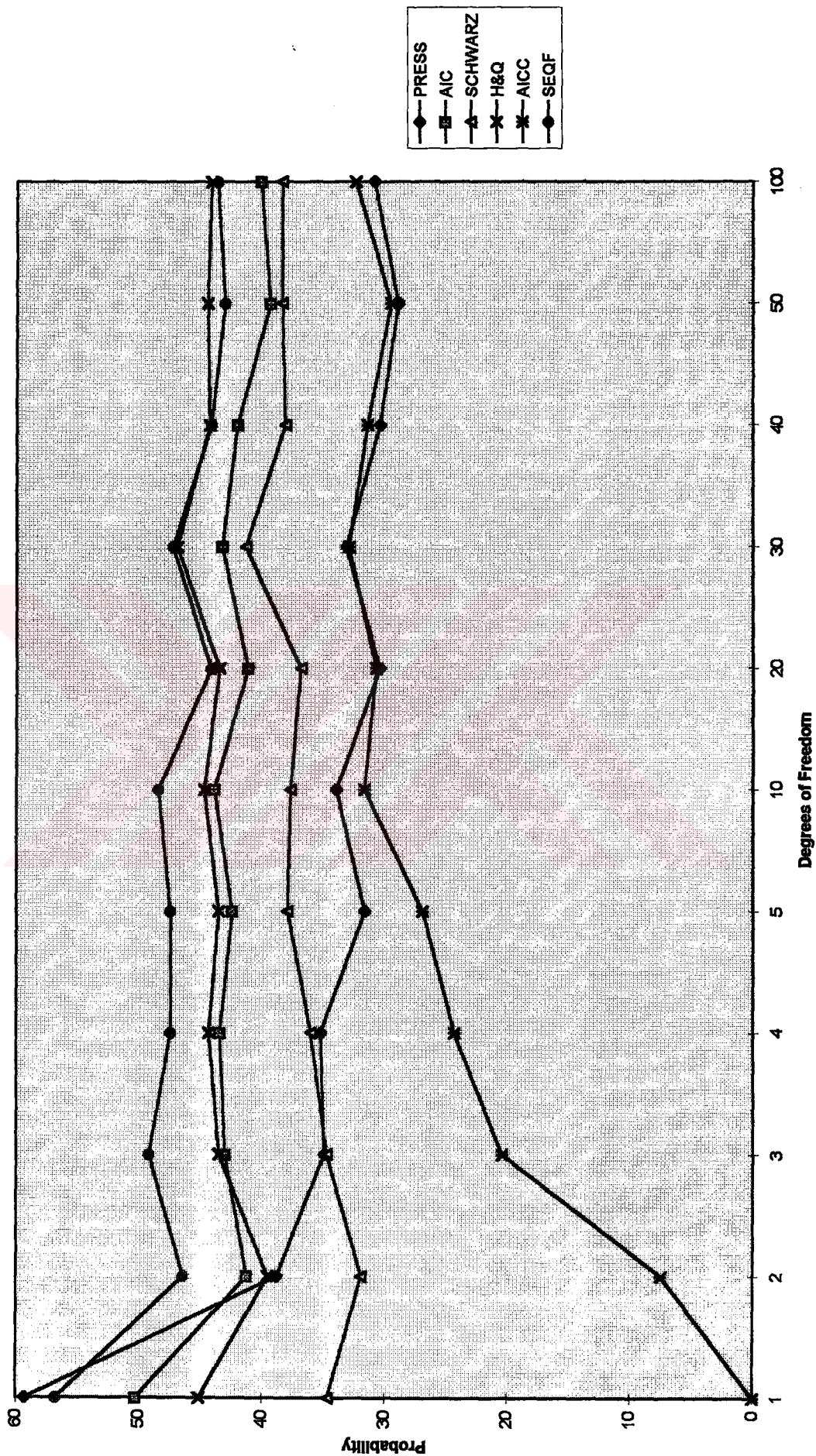


Figure 6a: Mean Squared Forecast Error (True Lag Order = 2, Sample Size = 30, A Skewed Distribution)

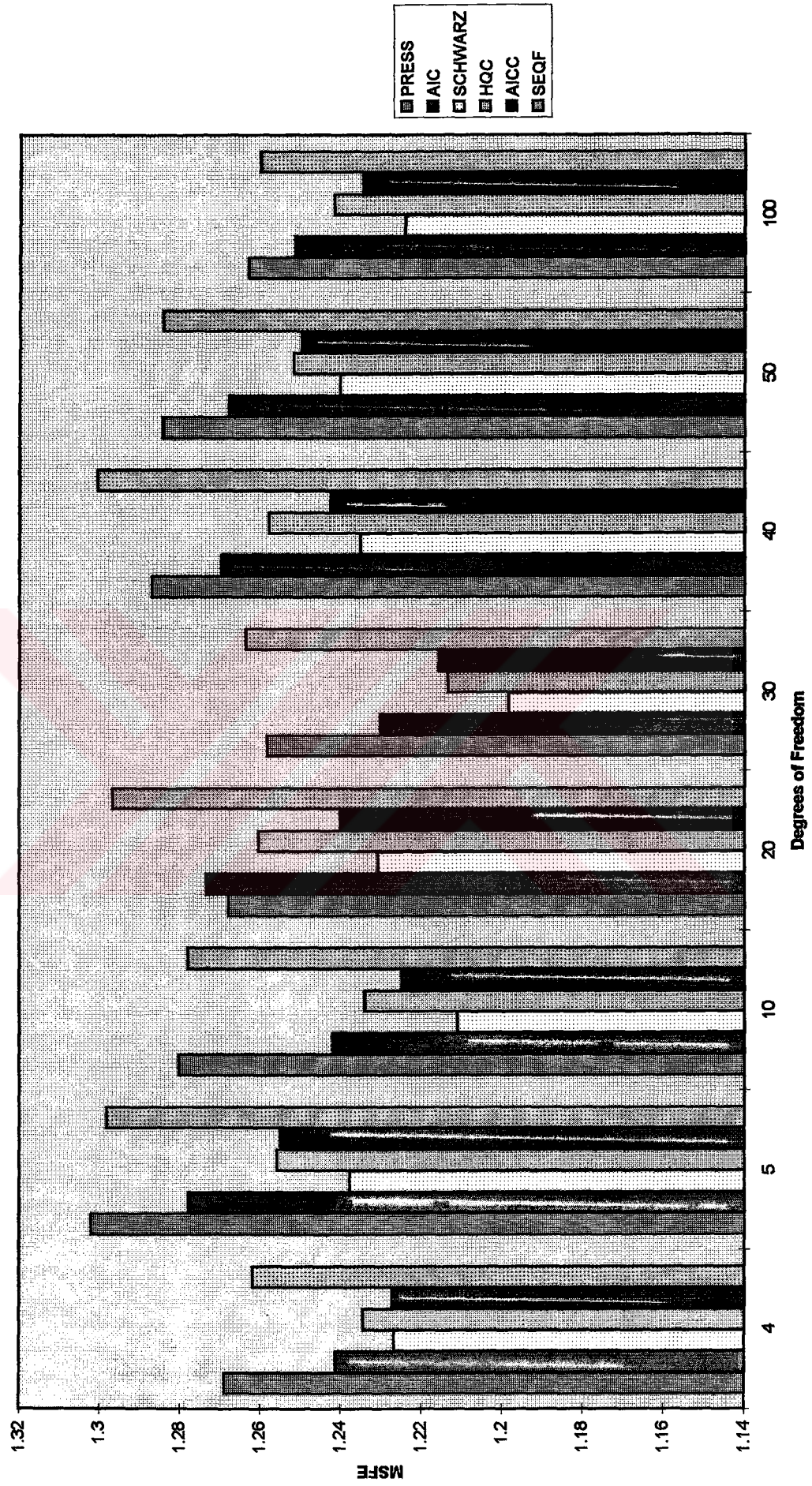


Figure 6b: Mean Squared Forecast Error (True Lag Order = 2, Sample Size = 200, A Skewed Distribution)

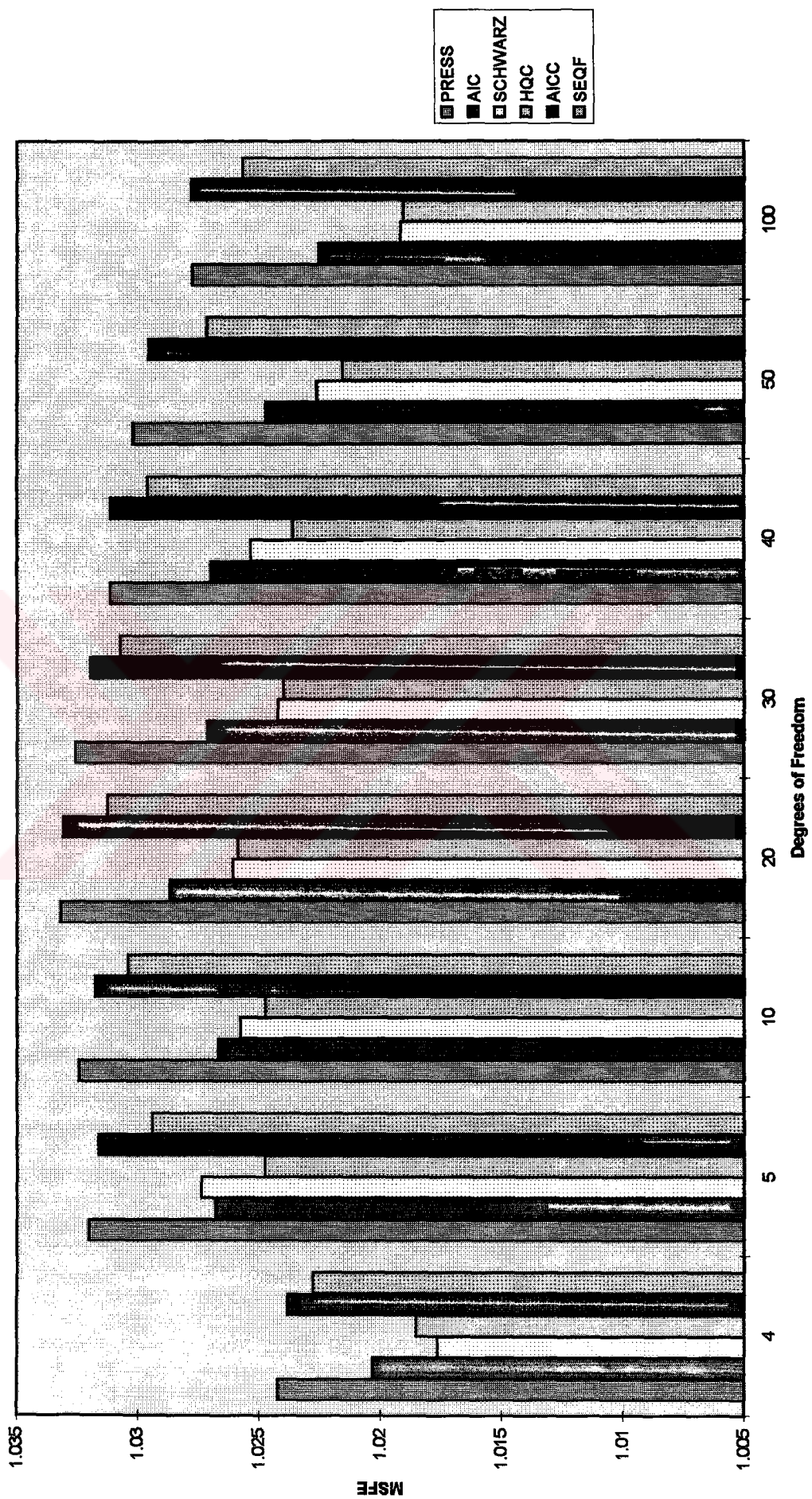


Figure 7a: Mean Squared Forecast Error (True Lag Order = 2, Sample Size = 30, t Distribution)

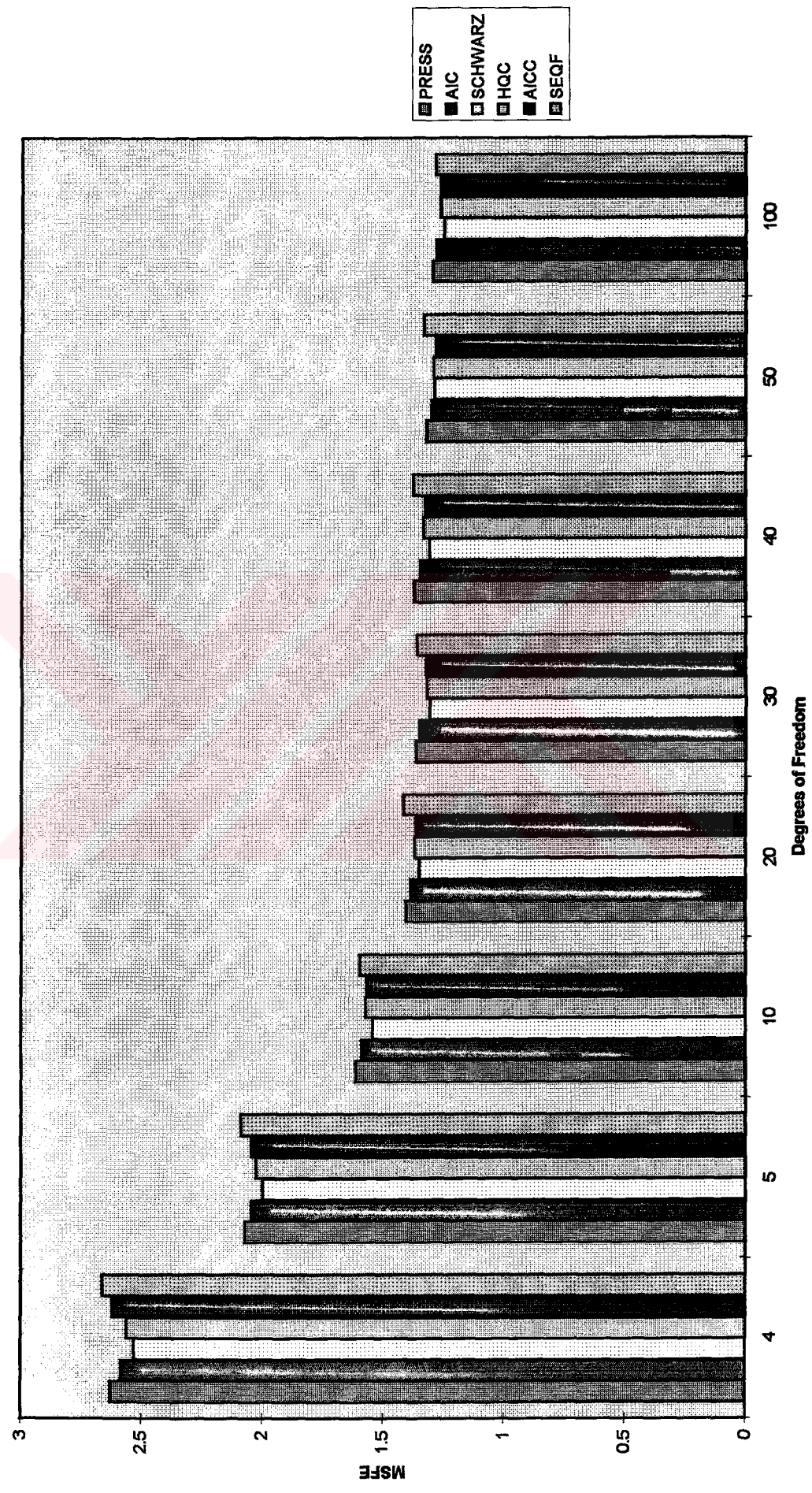


Figure 7b: Mean Squared Forecast Error (True Lag Order = 2, Sample Size = 200, t Distribution)

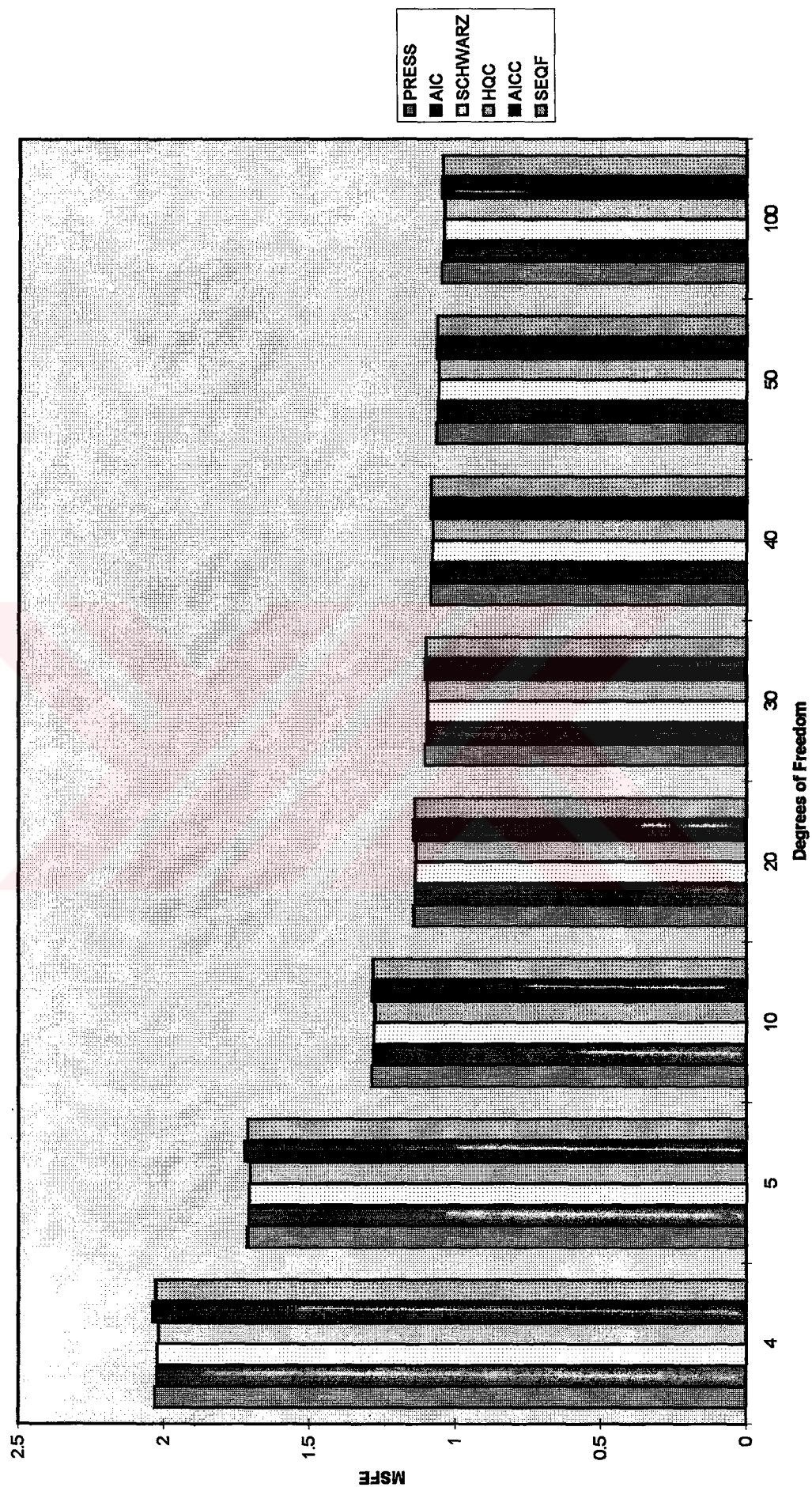


Figure 8a: Mean Squared Forecast Error (True Lag Order = 3, Sample Size = 30, A Skewed Distribution)

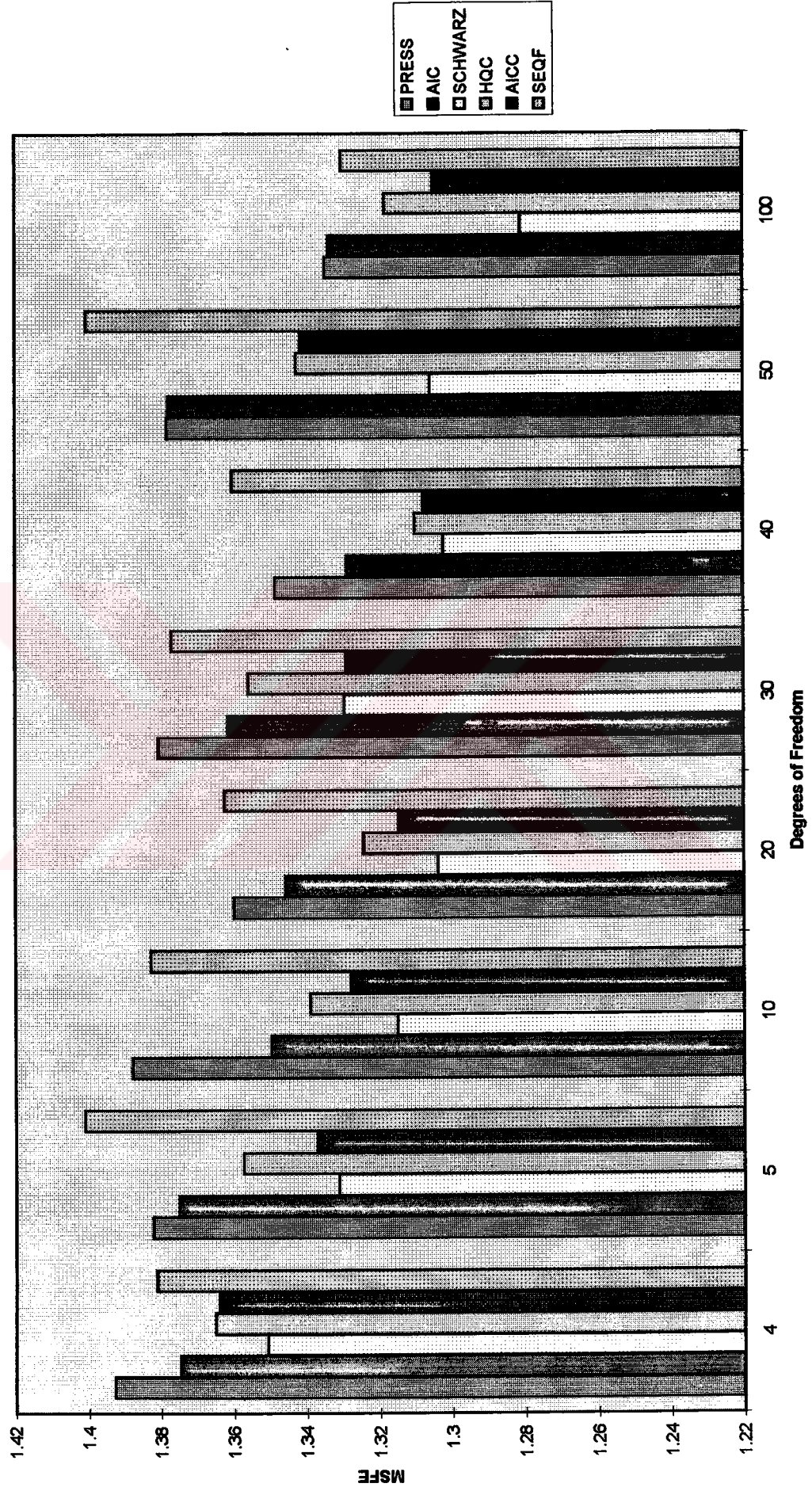


Figure 8b: Mean Squared Forecast Error (True Lag Order = 3, Sample Size = 200, A Skewed Distribution)

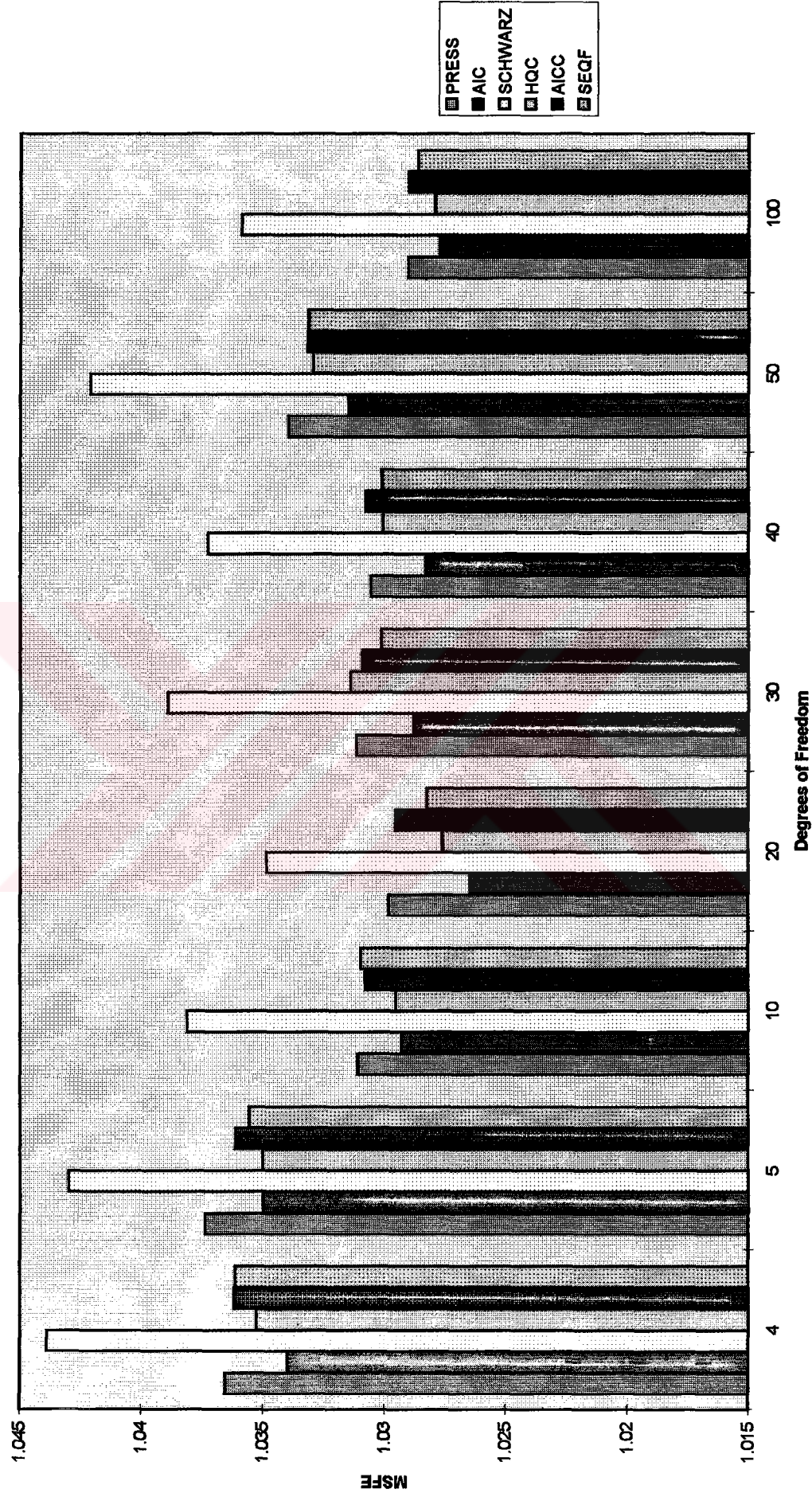


Figure 9a: Mean Squared Forecast Error (True Lag Order = 3, Sample Size = 30, t Distribution)

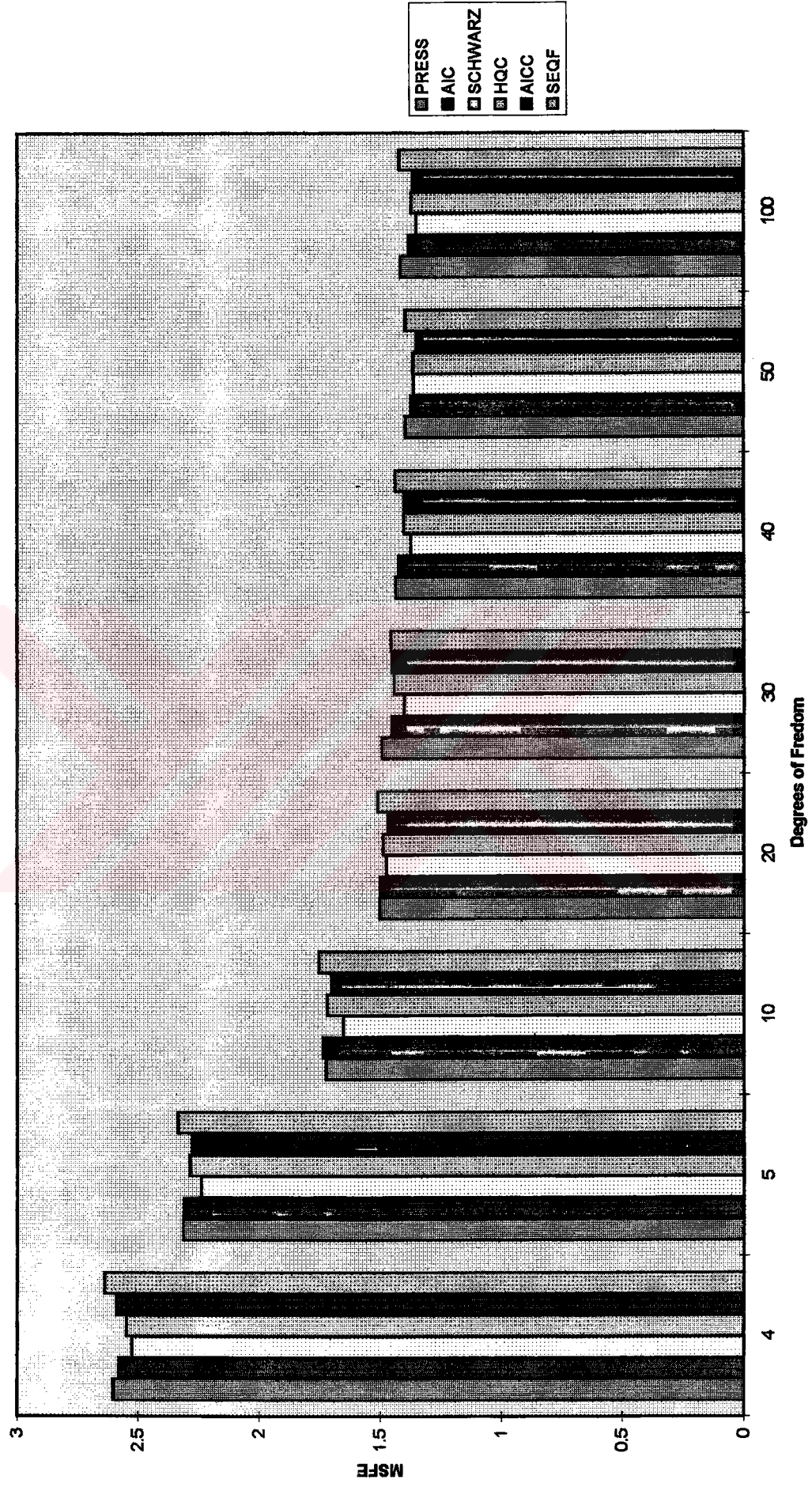


Figure 9b: Mean Squared Forecast Error (True Lag Order = 3, Sample Size = 200, t Distribution)

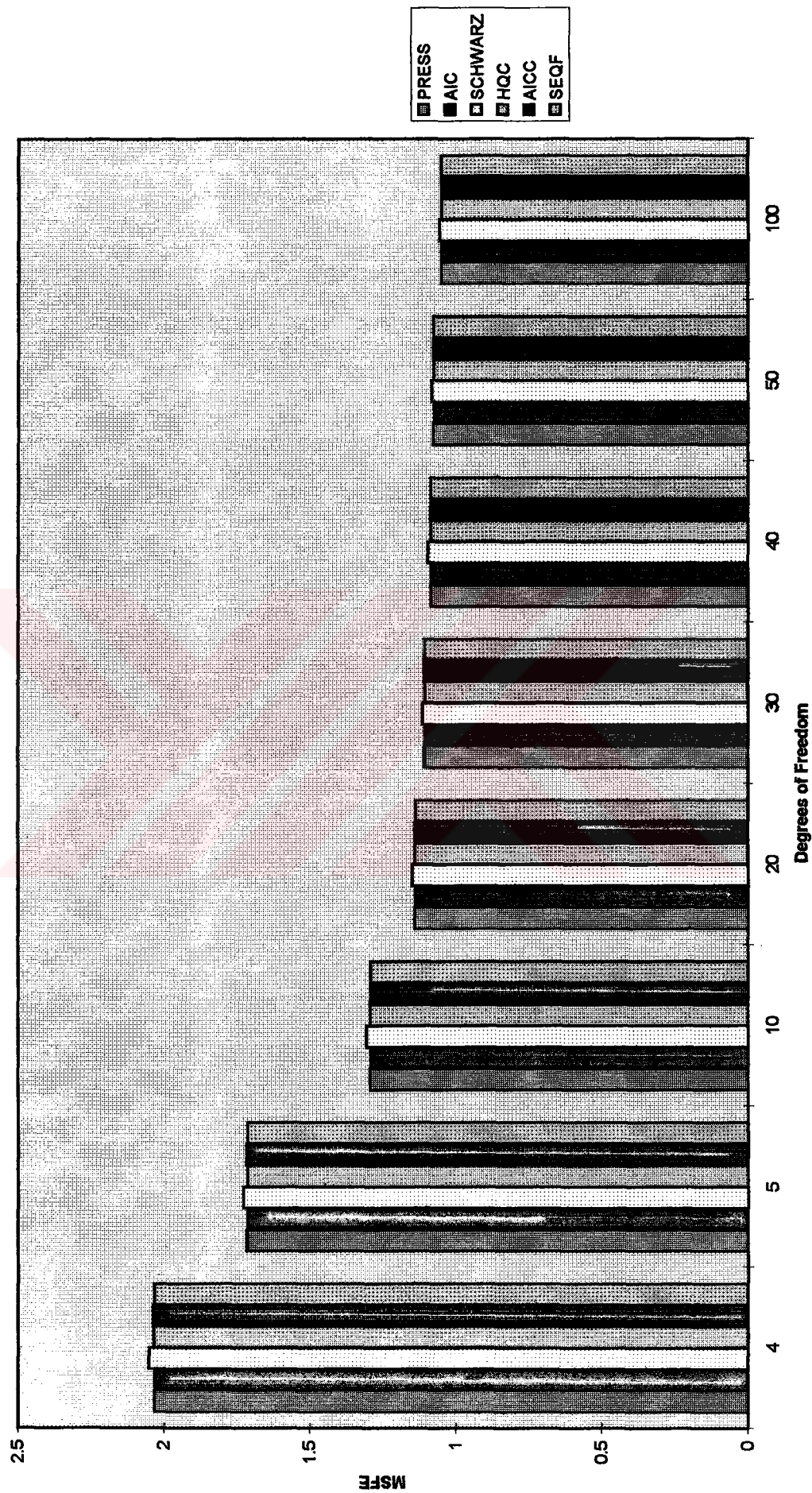
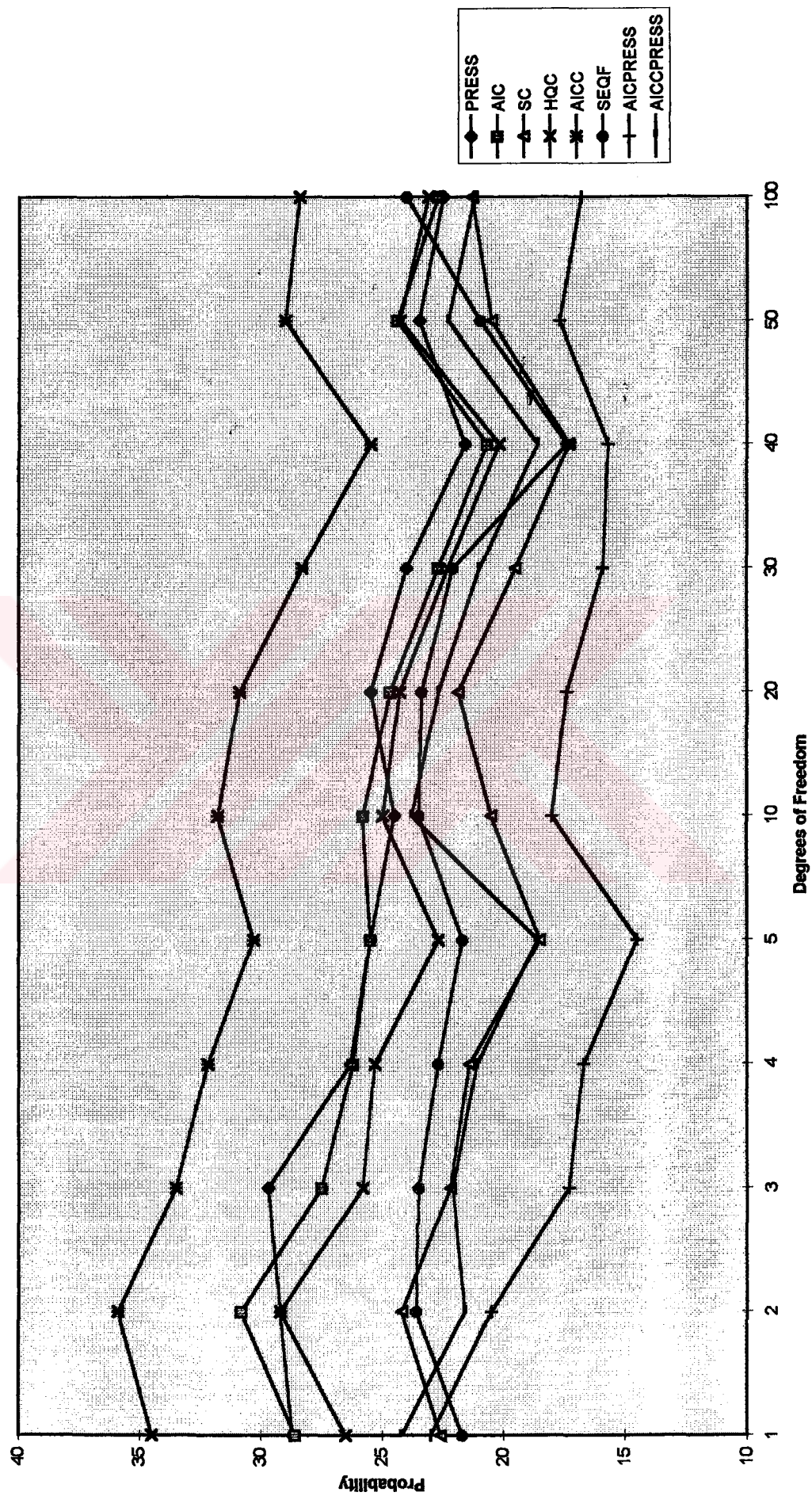


Figure 10a: Probability of Estimating the True Lag Order (True Lag Order = 2, Sample Size = 30, A skewed Distribution)



Journaling

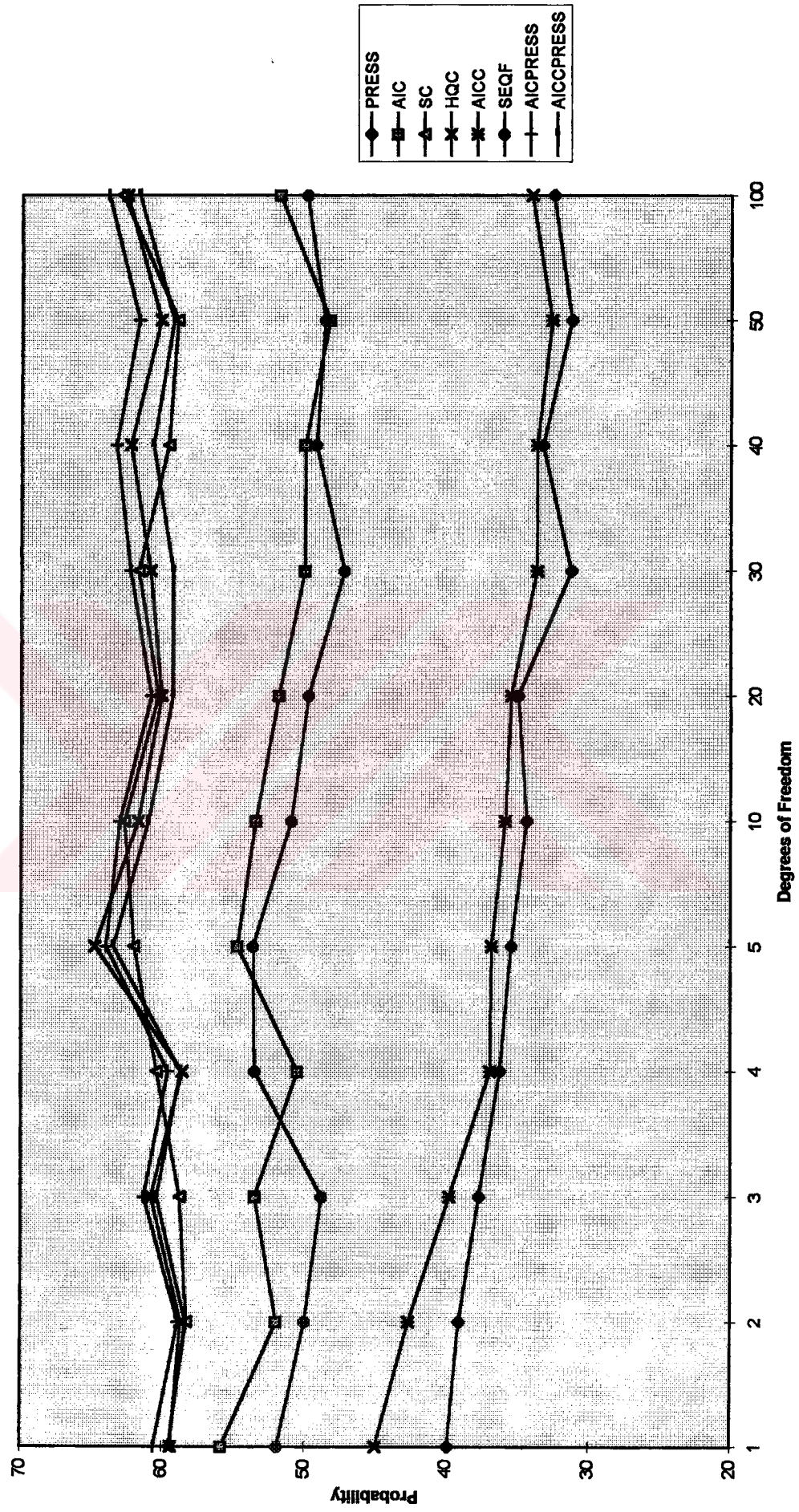


Figure 11a: Probability of Estimating the True Lag Order (True Lag Order = 2, Sample Size = 30, t Distribution)

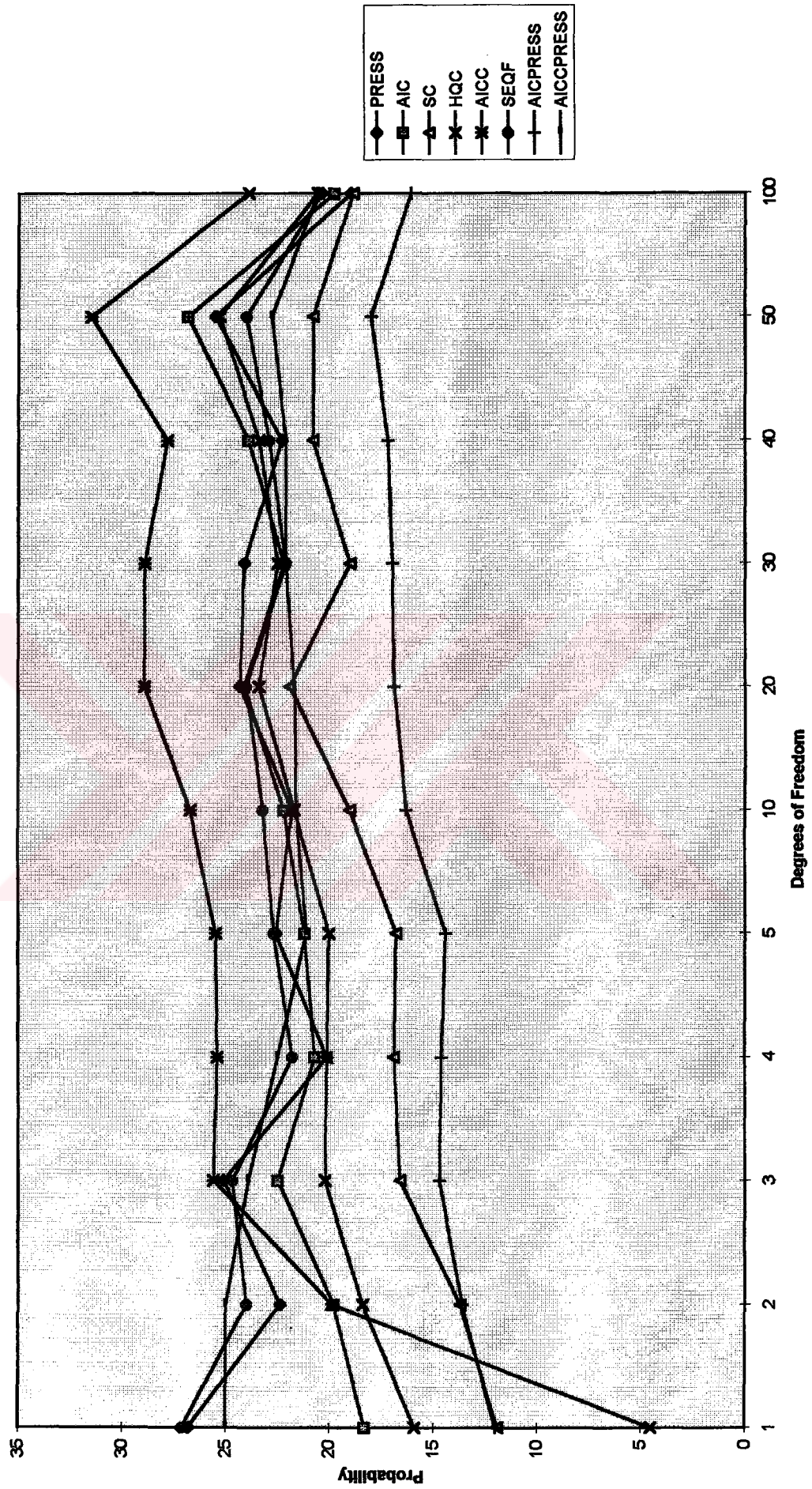


Figure 11b: Probability of Estimating the True Lag Order (True Lag Order = 2, Sample Size = 200, t Distribution)

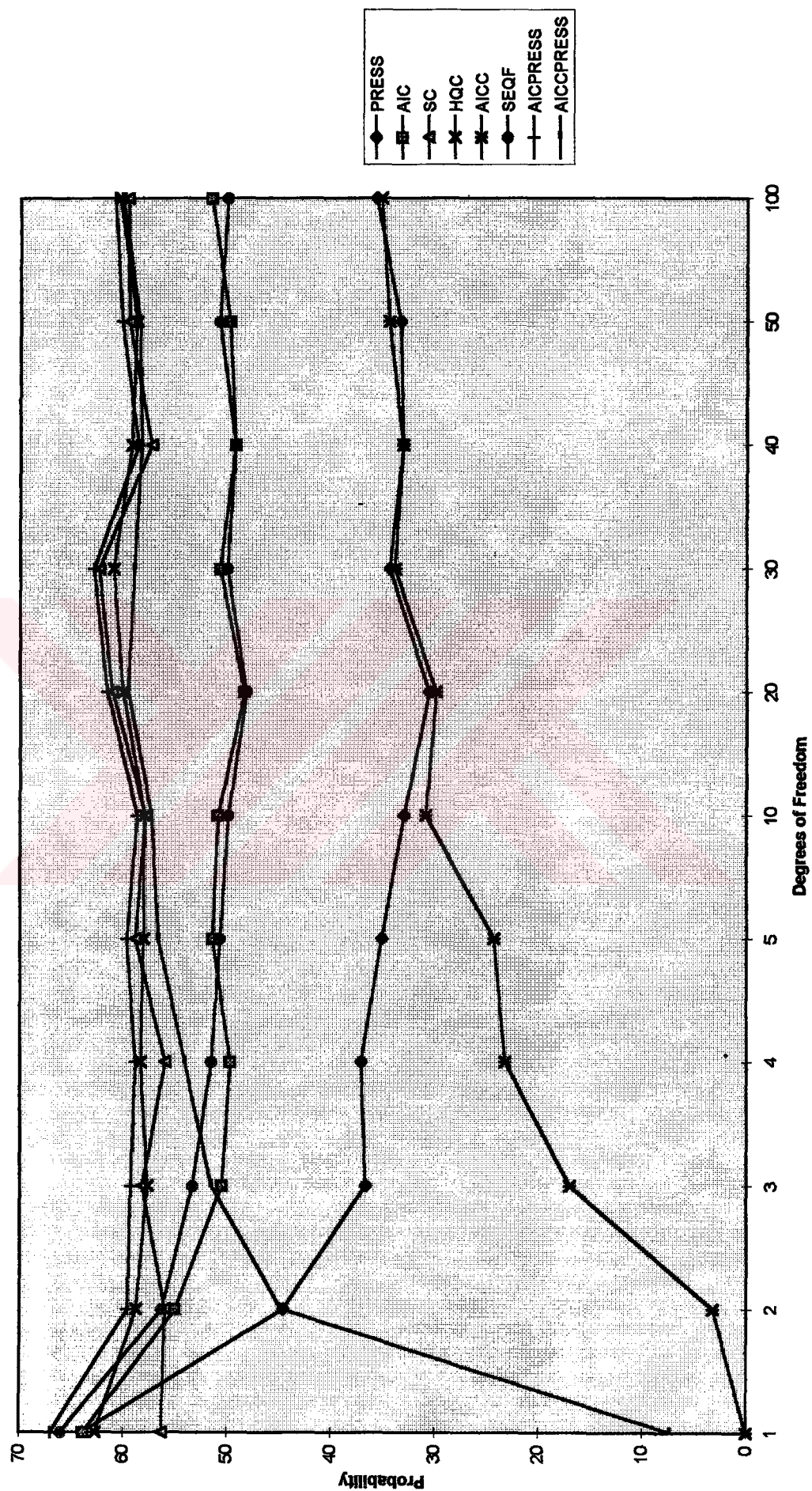


Figure 12a: Probability of Estimating the True Lag Order (True Lag Order = 3, Sample Size = 30, A Skewed Distribution)

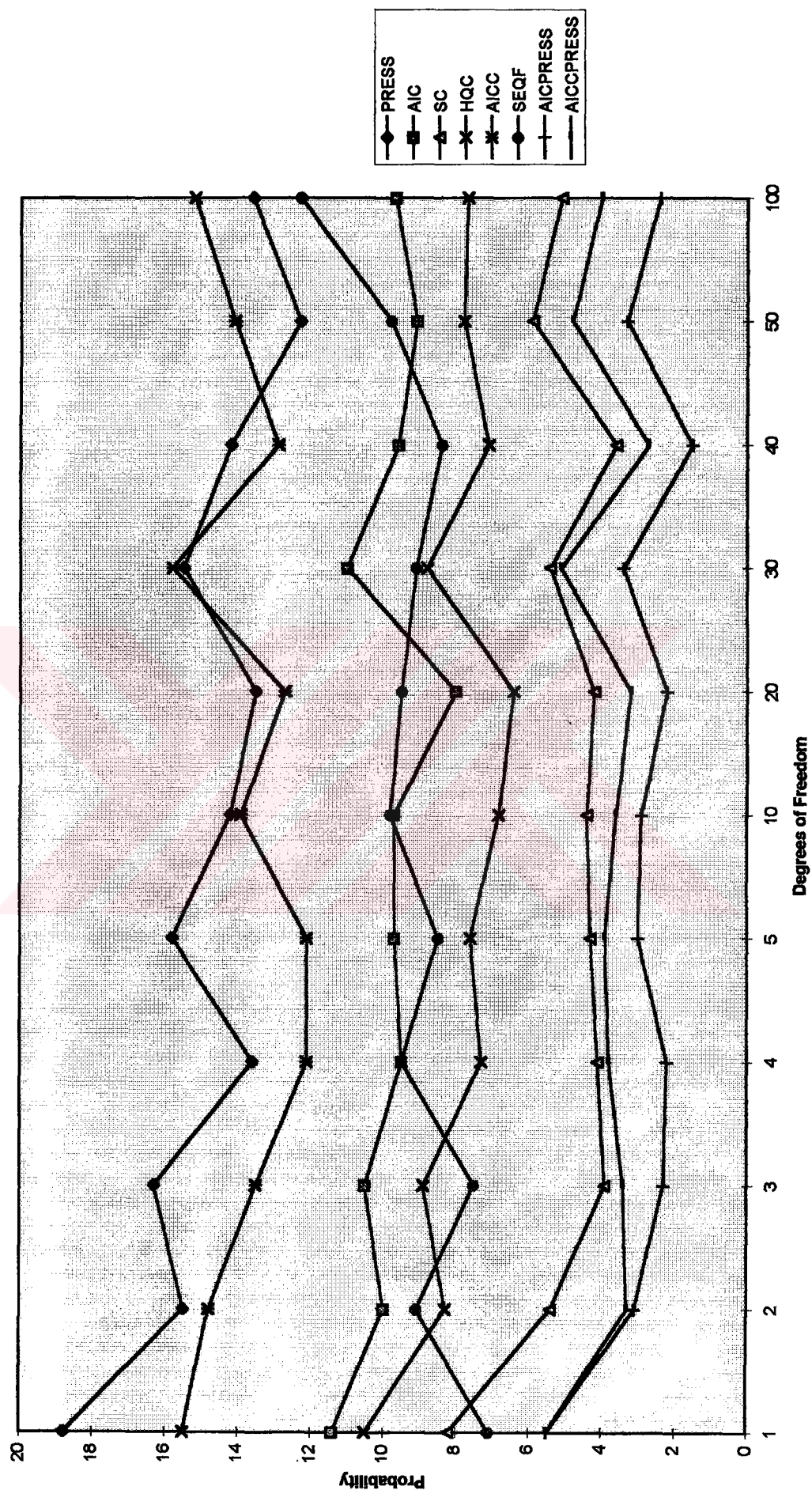


Figure 13a: Probability of Estimating the True Lag Order (True Lag Order = 3, Sample Size = 30, t Distribution)

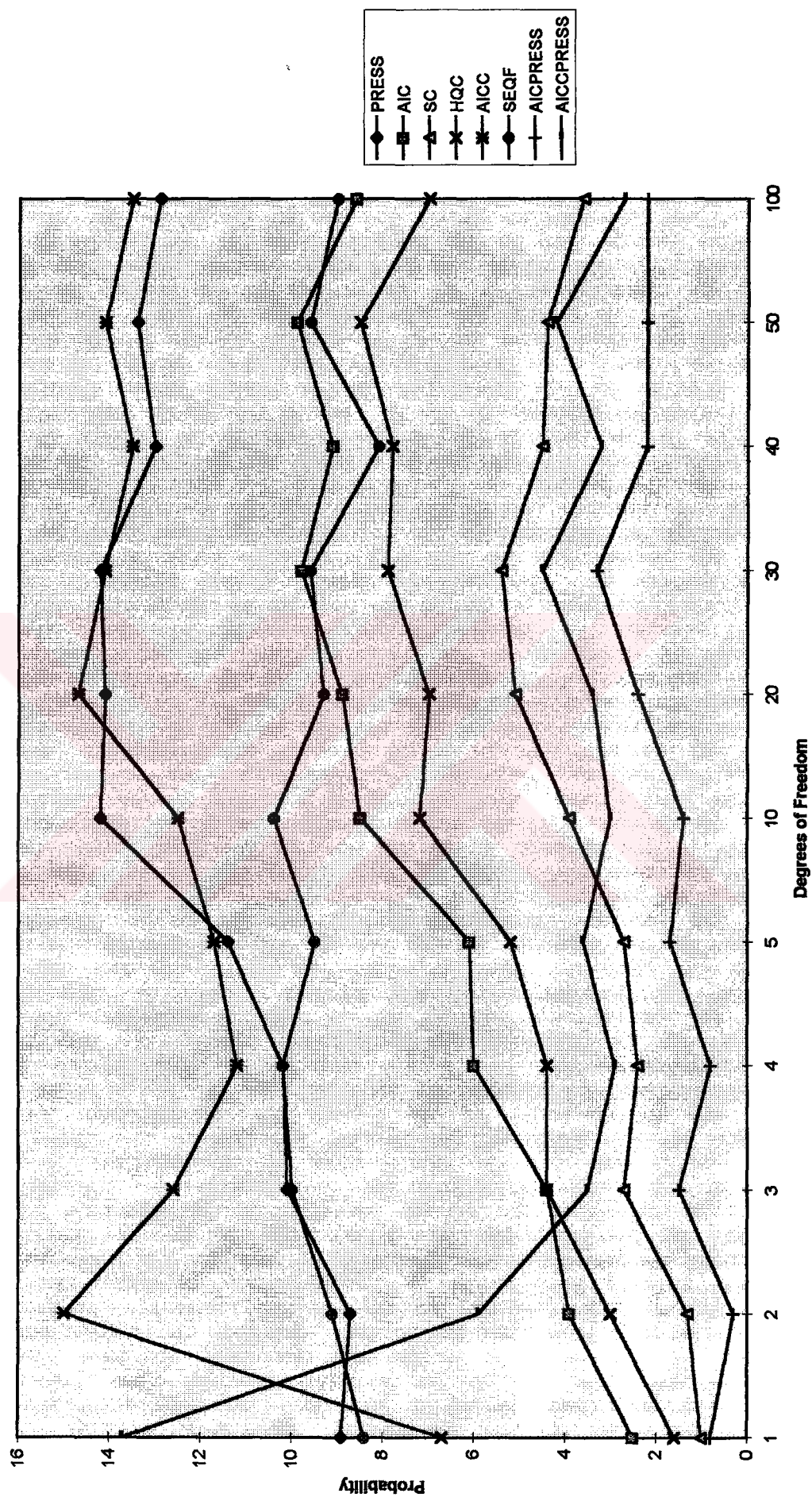


Figure 13b: Probability of Estimating the True Lag Order (True Lag Order = 3, Sample Size = 200, t Distribution)

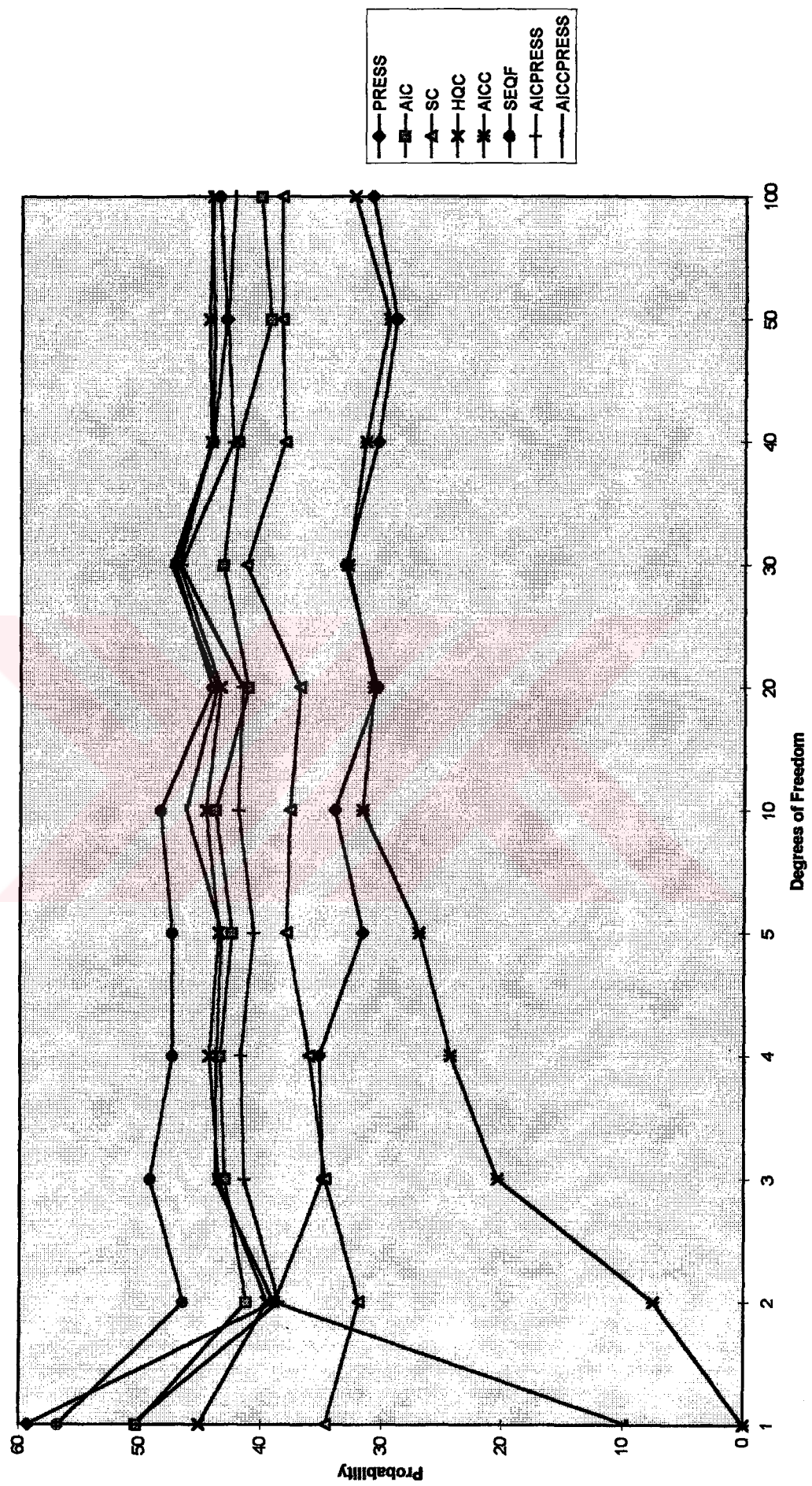


Figure 14a: Mean Squared Forecast Error (True Lag Order = 2, Sample Size = 30, A Skewed Distribution)

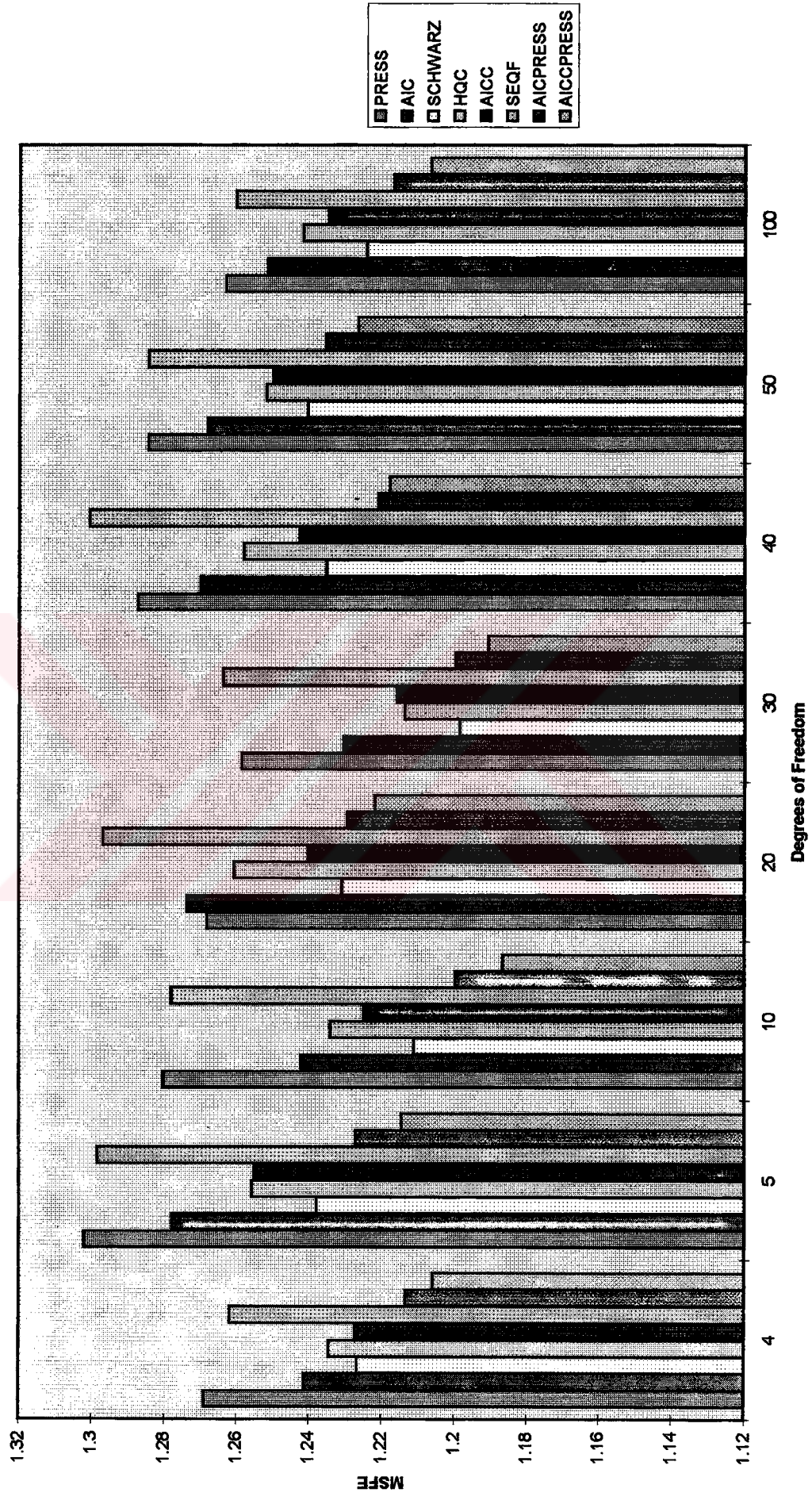


Figure 14b: Mean Squared Forecast Error (True Lag Order = 2, Sample Size = 200, A Skewed Distribution)

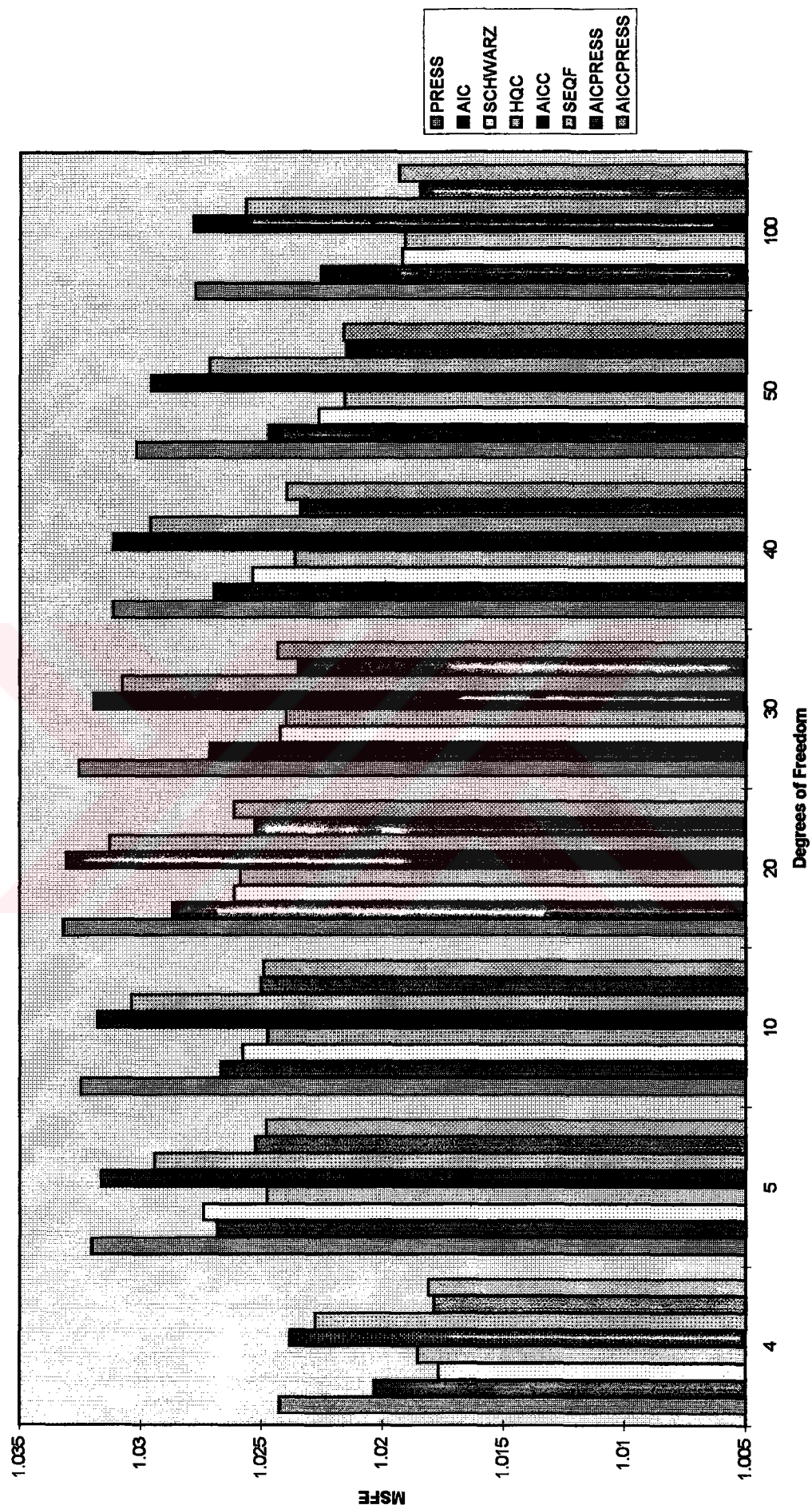


Figure 15a: Mean Squared Forecast Error (True Lag Order = 2, Sample Size = 30, t Distribution)

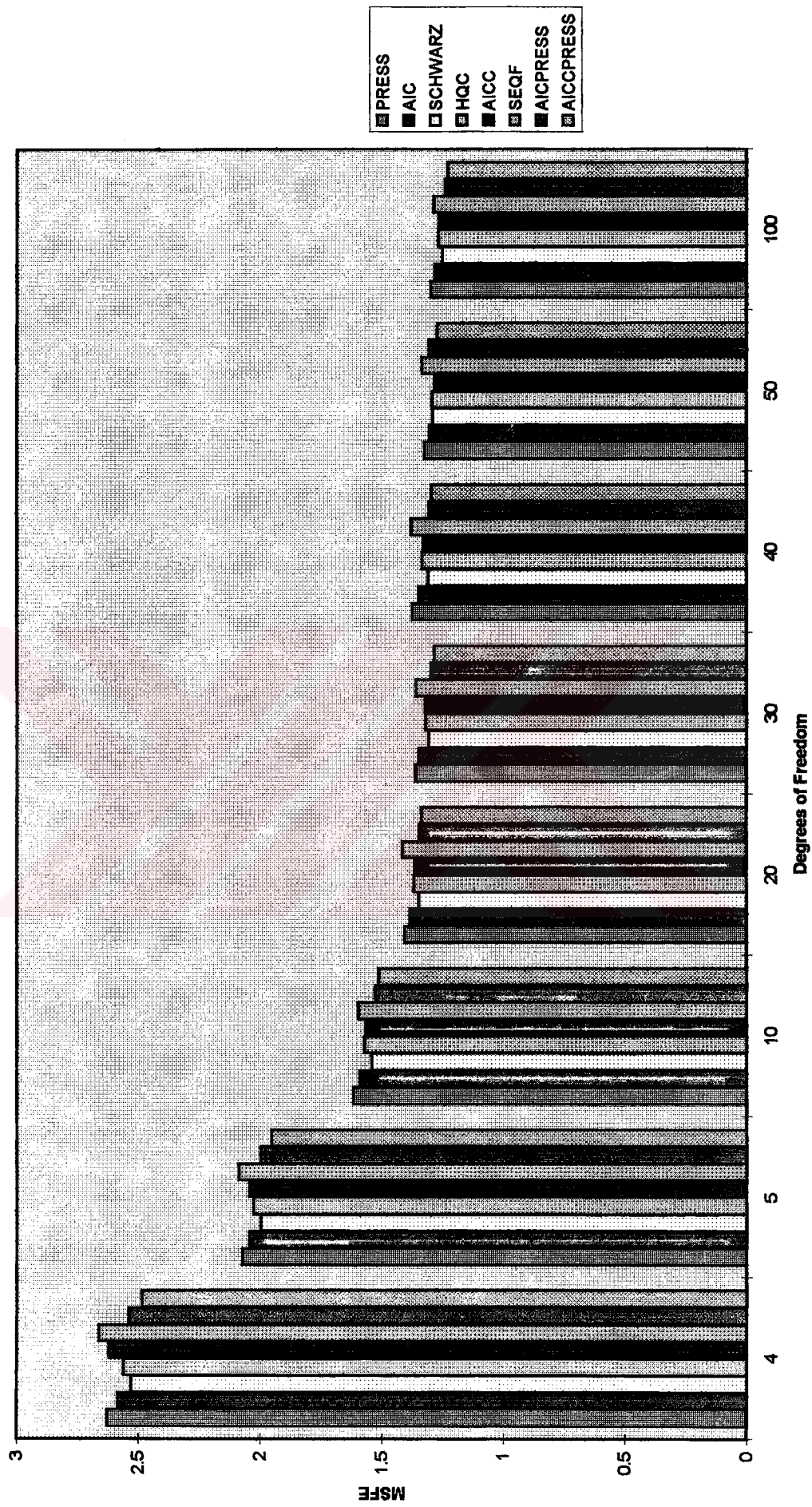


Figure 15b: (True Lag Order = 2, Sample Size = 200, t Distribution)



Figure 16a: Mean Squared Forecast Error (True Lag Order = 3, Sample Size = 30, A Skewed Distribution)

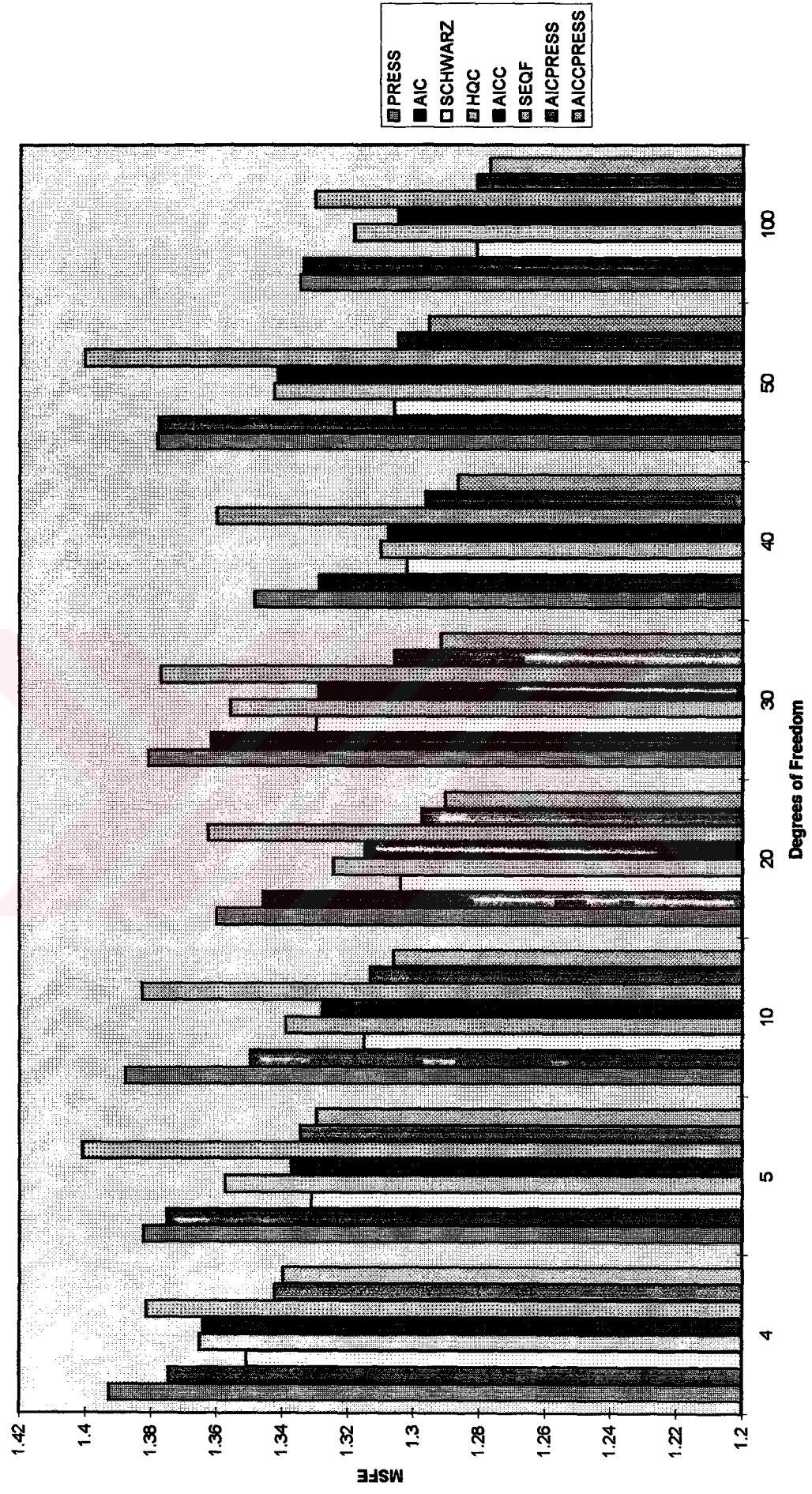


Figure 16b: Mean Squared Forecast Error (True Lag Order = 3, Sample Size = 200, A Skewed Distribution)

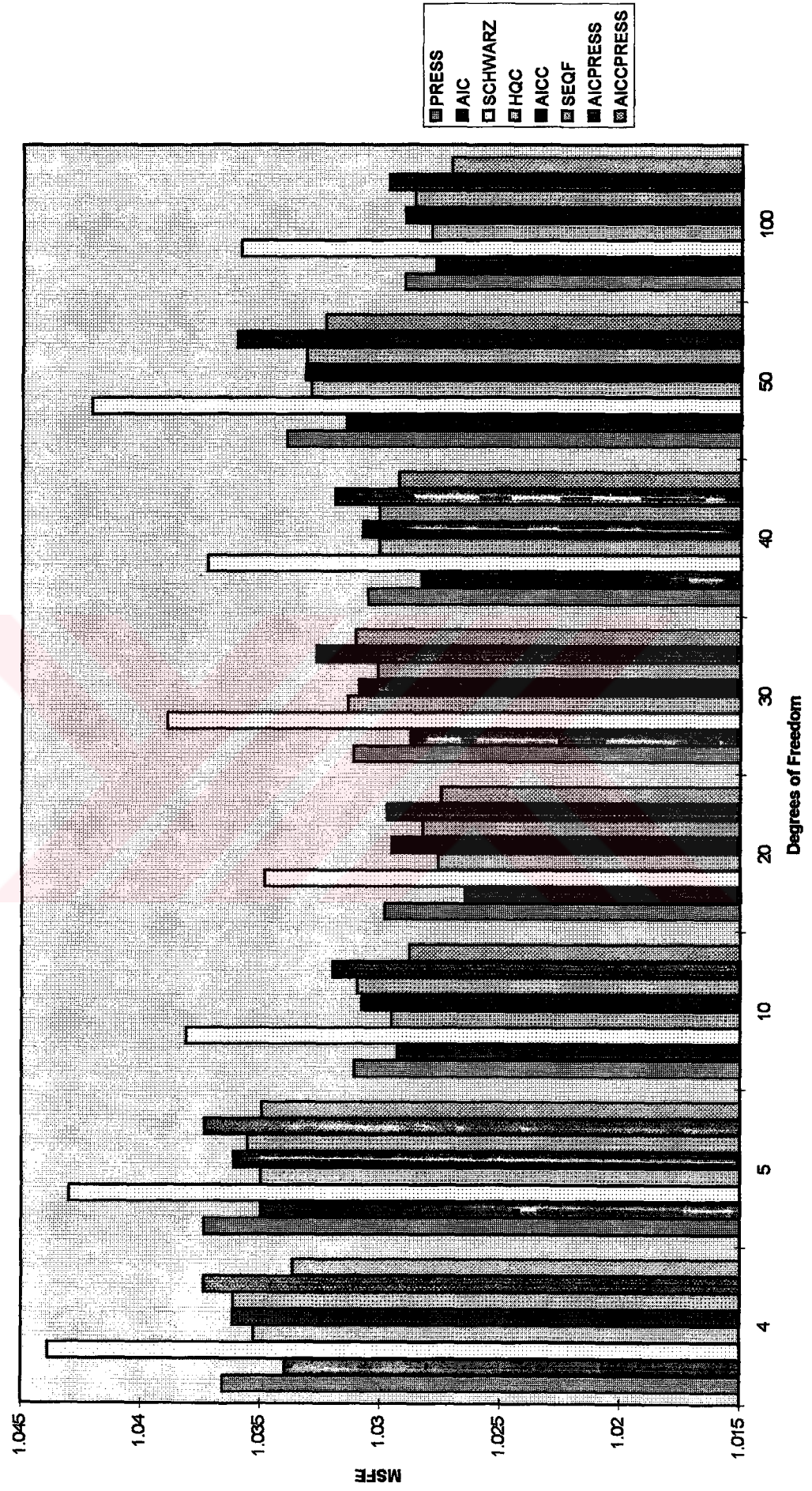


Figure 17b: Mean Squared Forecast Error (True Lag Order = 3, Sample Size = 200, t Distribution)

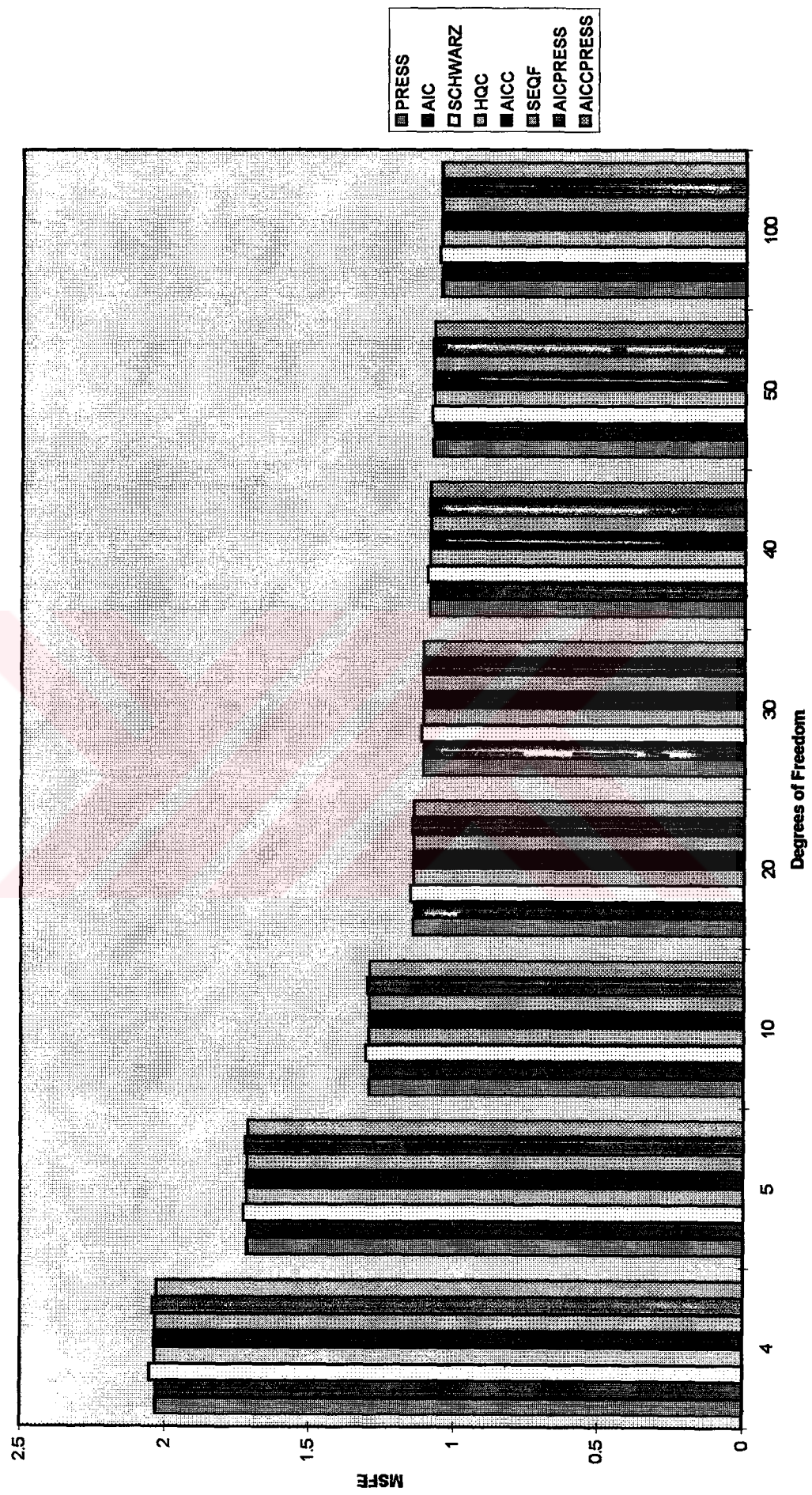


Figure 18: Expected Value of $\Delta_1 = I(AIC) - I(AICPRESS)$

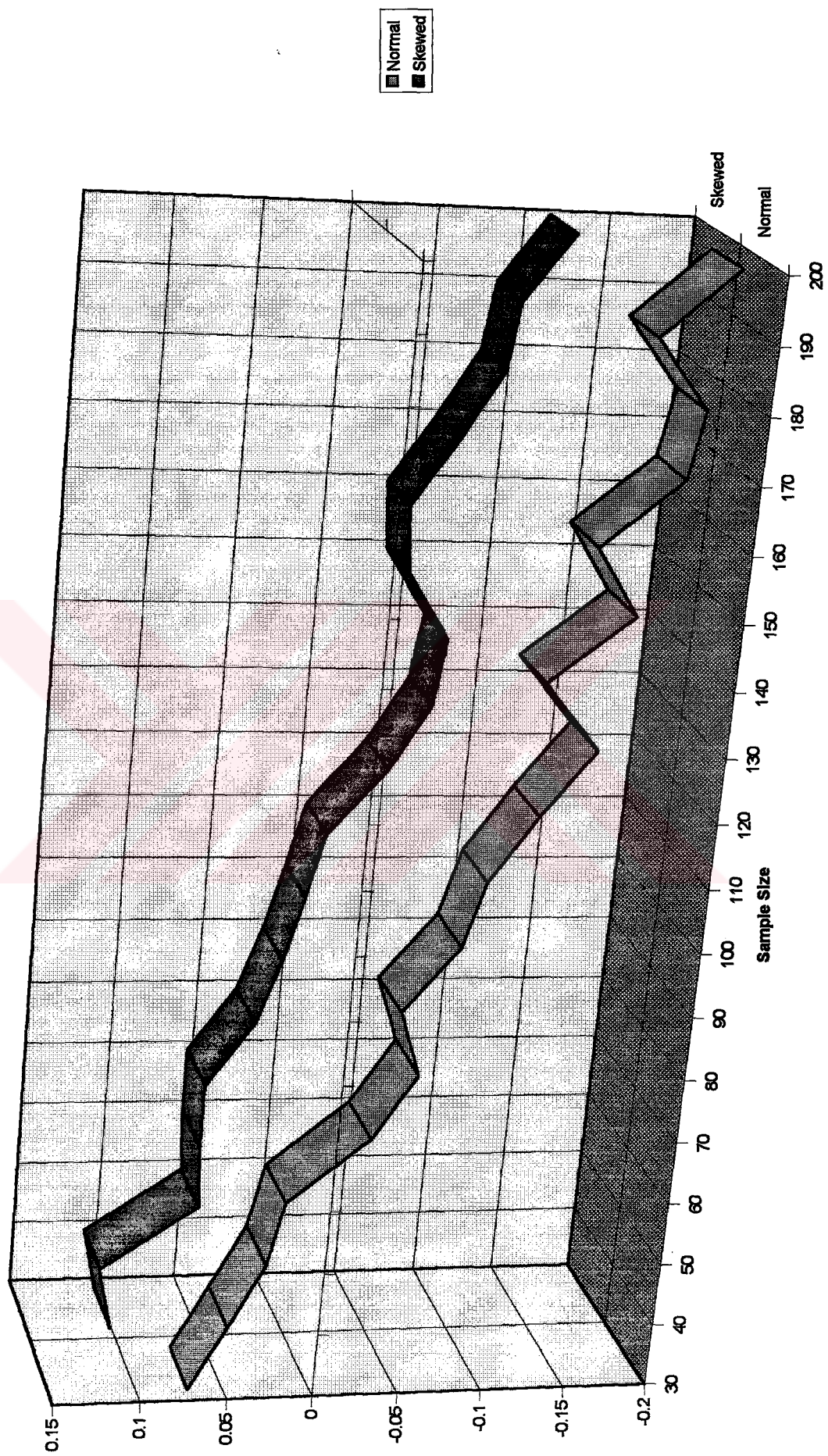


Figure 19: Expected Value of $\Delta_2 = I(AIC) - I(AICCV10)$

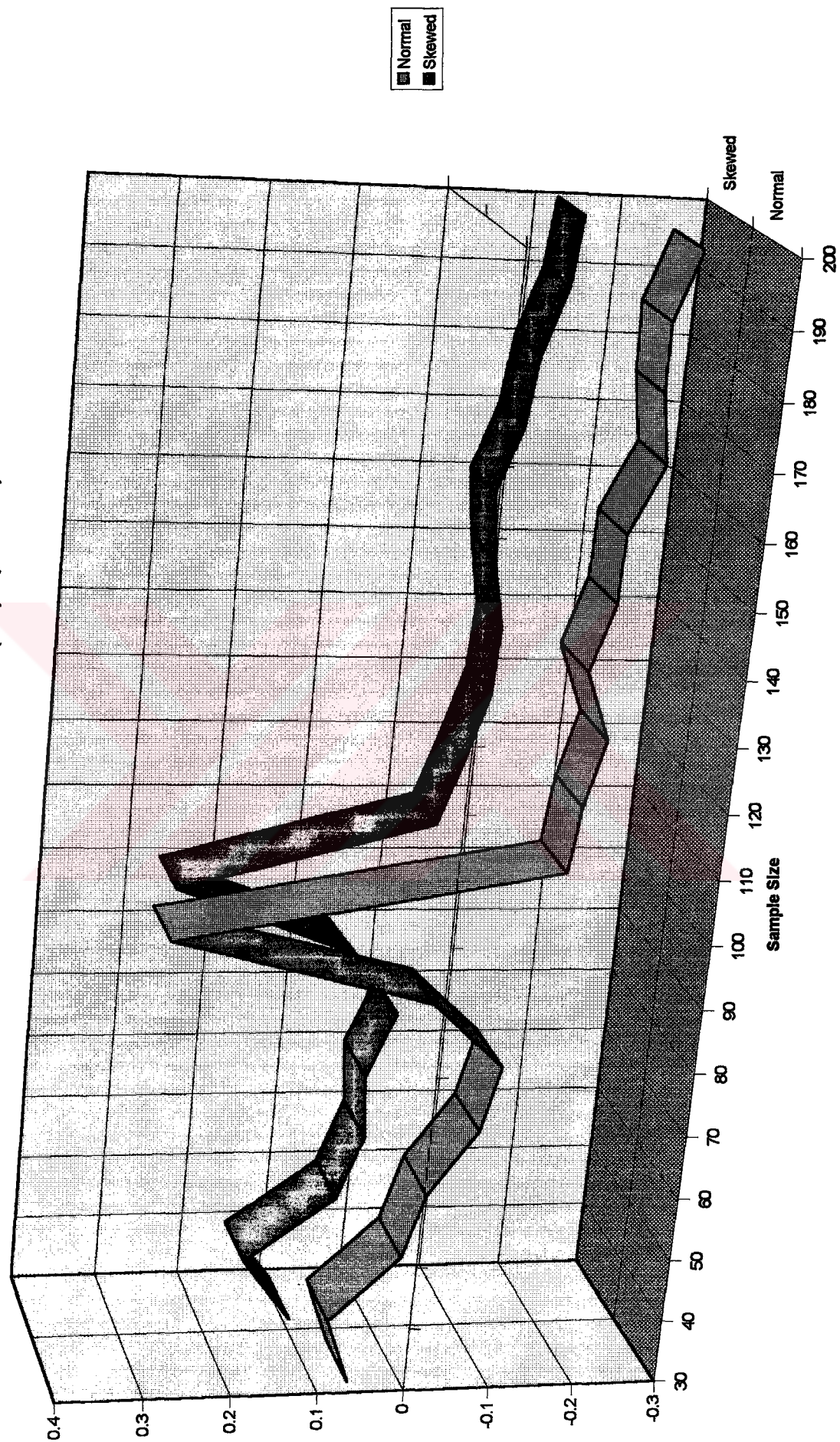


Figure 20: Expected Value of $\Delta_3 = I(\text{PRESS}) - I(\text{CV10})$

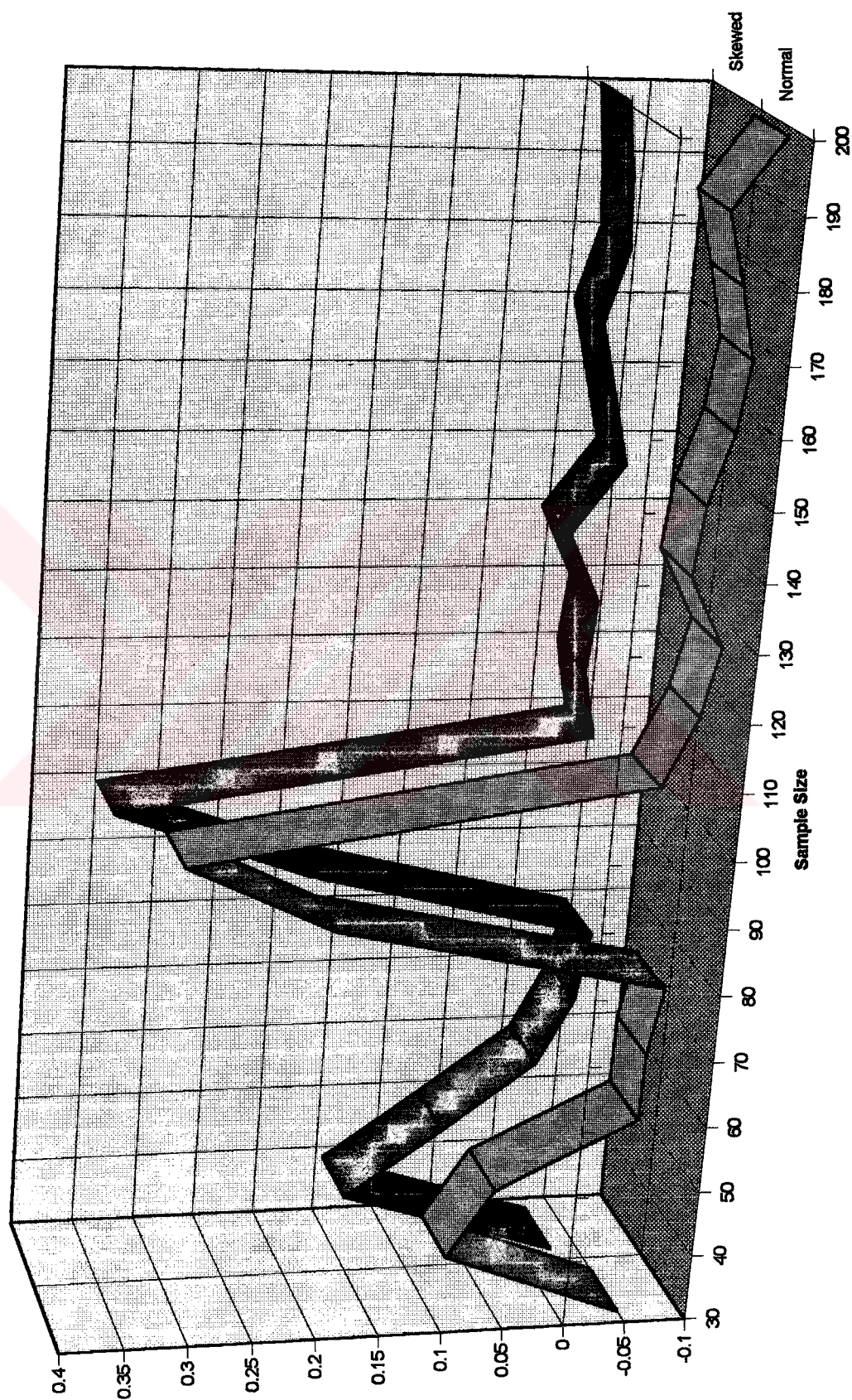


Figure 21a: Expected Value of $\delta_1 = \text{MSFE}(\text{AIC}) - \text{MSFE}(\text{AICPRESS})$, Normal Distribution

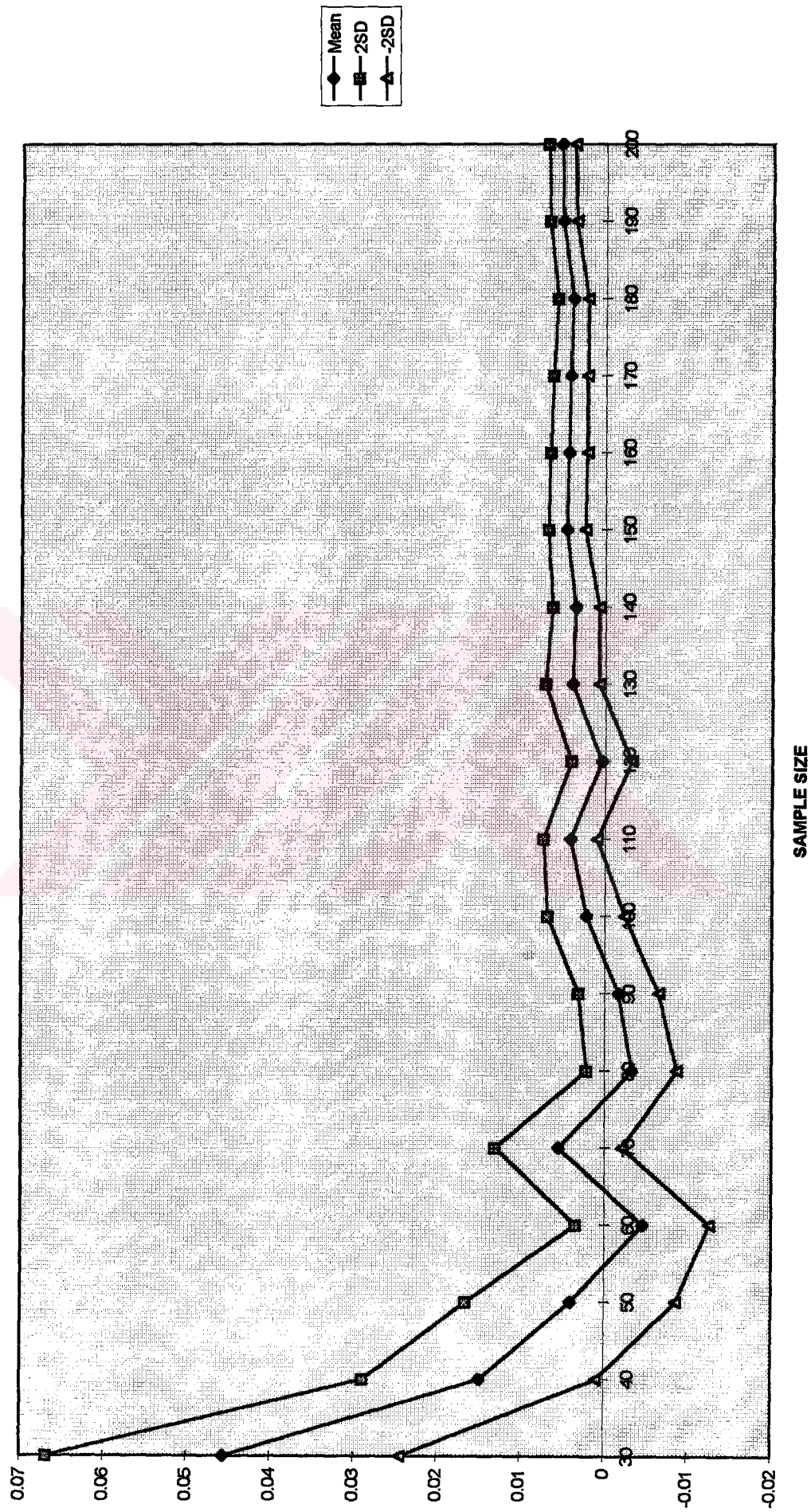


Figure 21b: Expected Value of $\delta_2 = \text{MSFE}(\text{AIC}) - \text{MSFE}(\text{AICCV10})$, Normal distribution

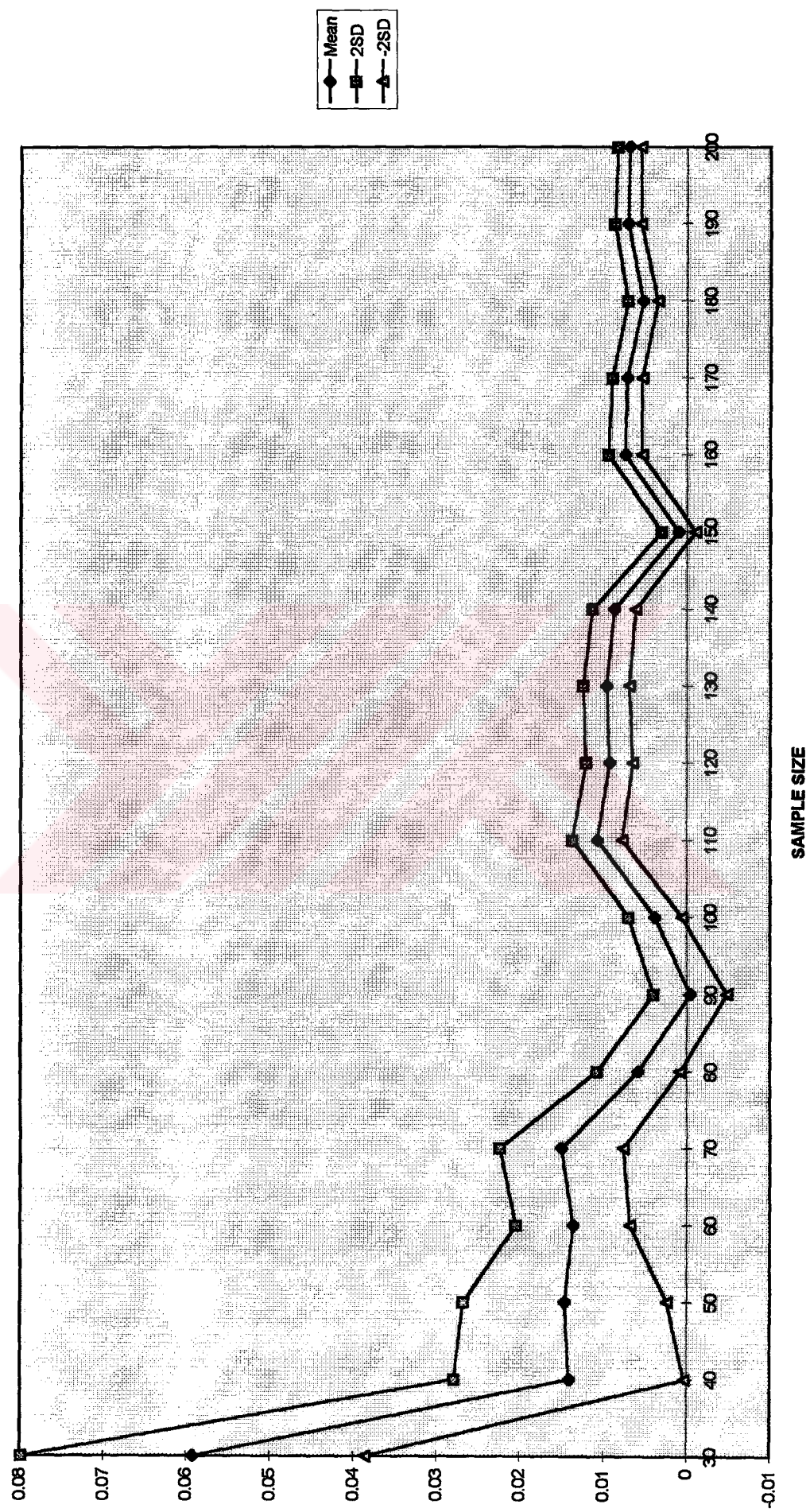


Figure 21c: Expected Value of $\delta_3 = \text{MSFE(PRESS)} - \text{MSFE(CV10)}$, Normal distribution

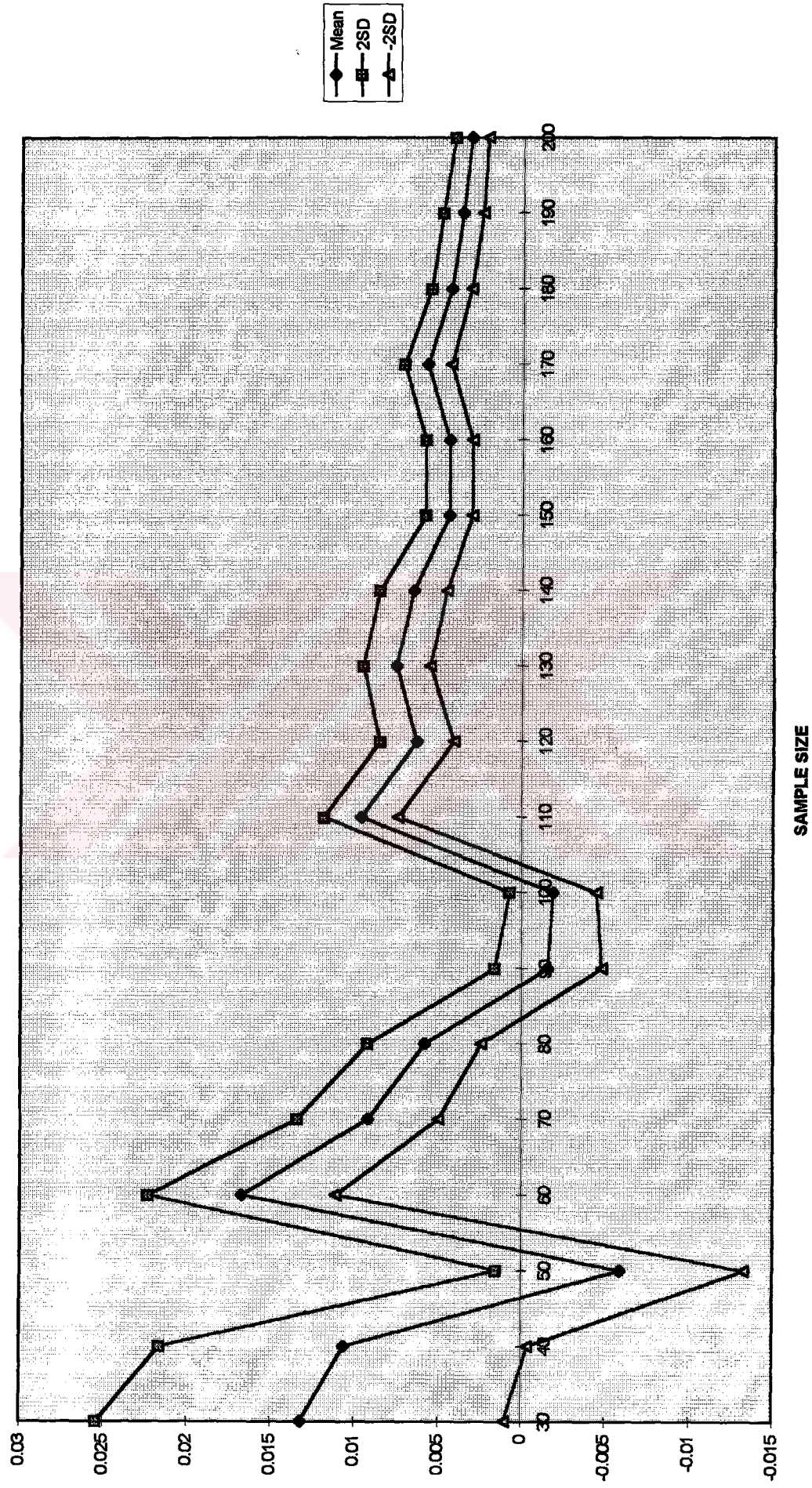


Figure 22a: Expected Value of $\delta_1 = \text{MSFE}(\text{AIC}) - \text{MSFE}(\text{AICPRESS})$, Skewed distribution

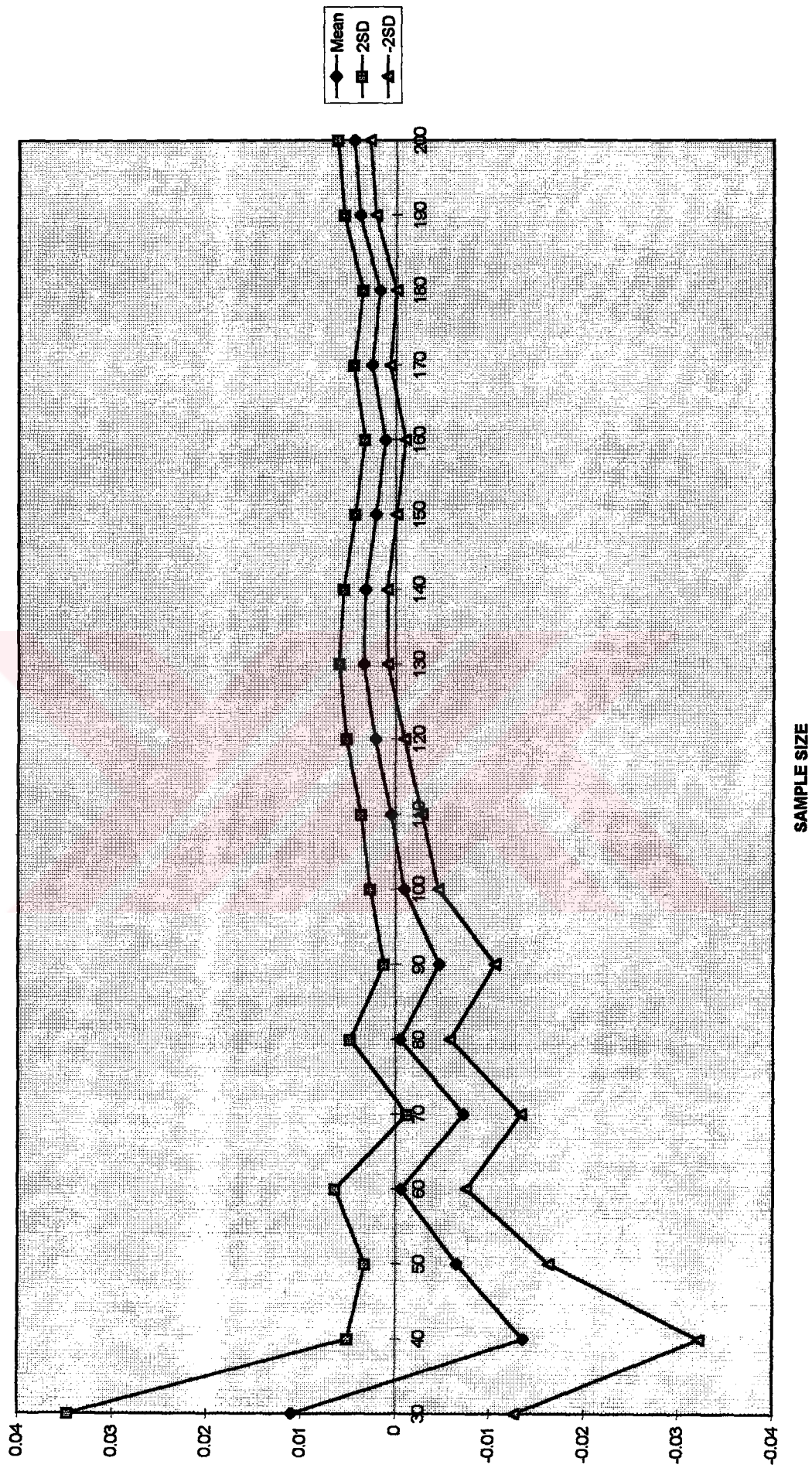


Figure 22b: Expected Value of $\delta_2 = \text{MSFE}(\text{AIC}) - \text{MSFE}(\text{AICCV10})$, Skewed distribution

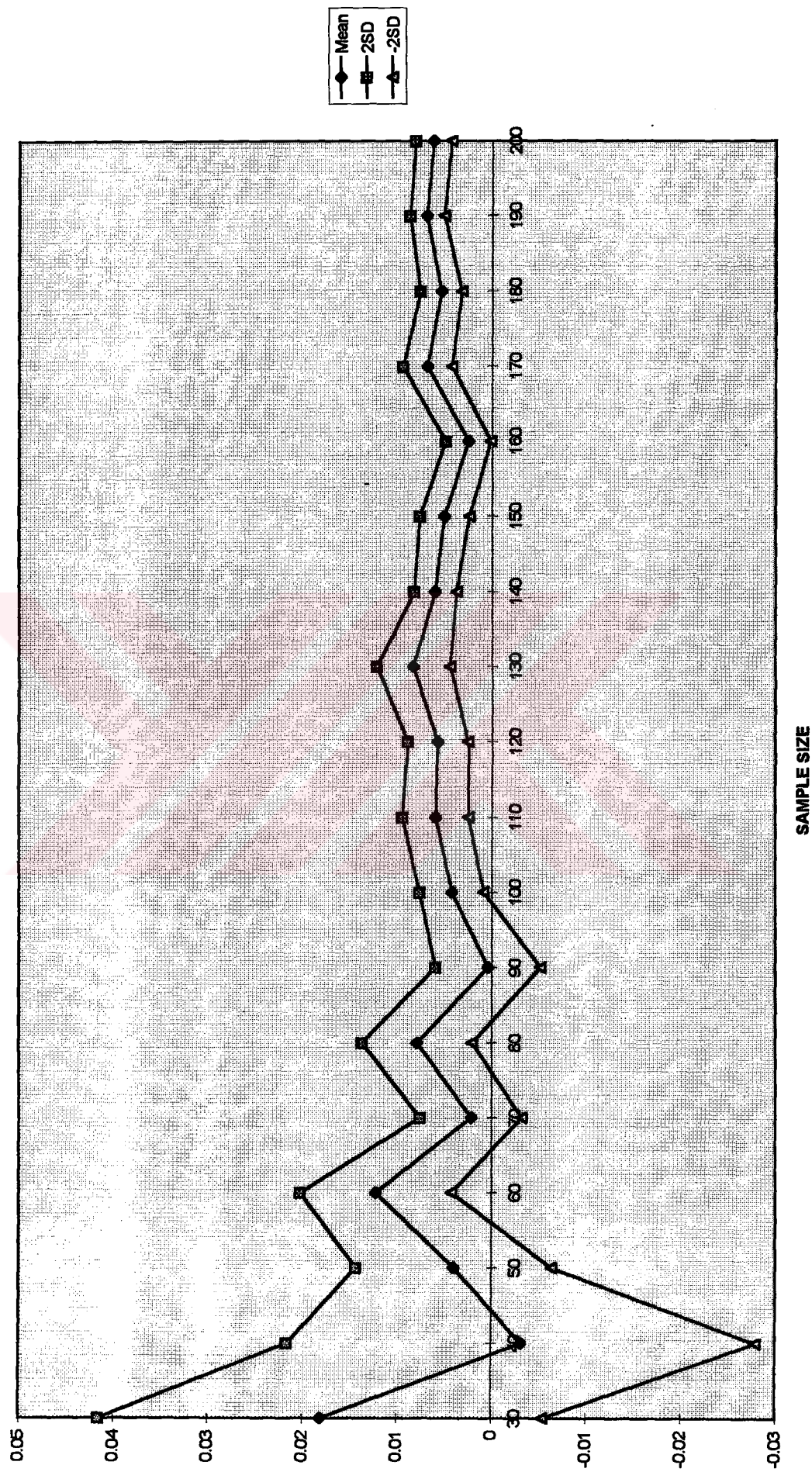


Figure 22c: Expected Value of $\delta_3 = \text{MSFE(PRESS)} - \text{MSFE(CV10)}$, Skewed distribution

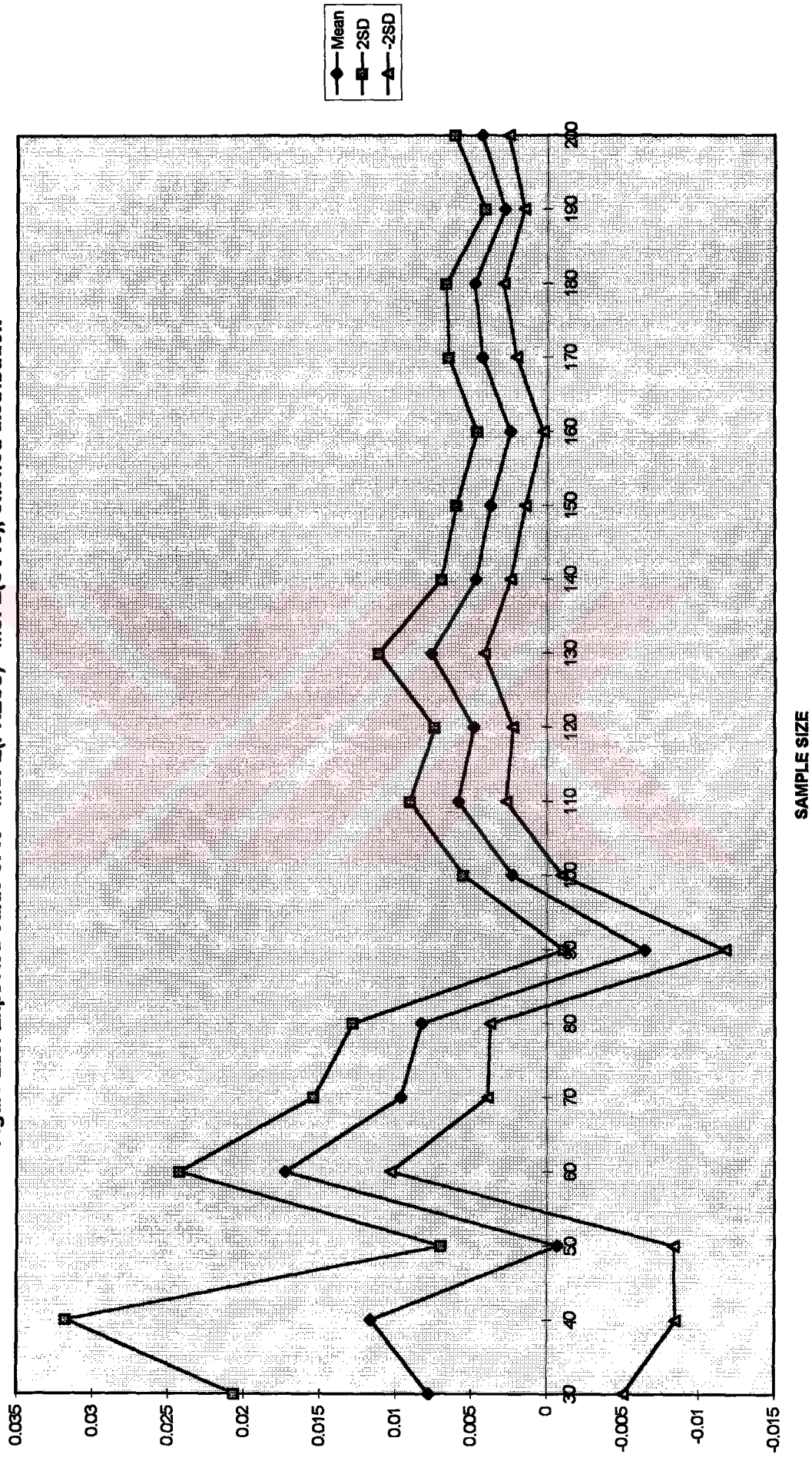


Figure 23a: Expected Value of $d_1 = \text{ABS}(\text{AIC}) - \text{ABS}(\text{AICPRESS})$, Normal distribution

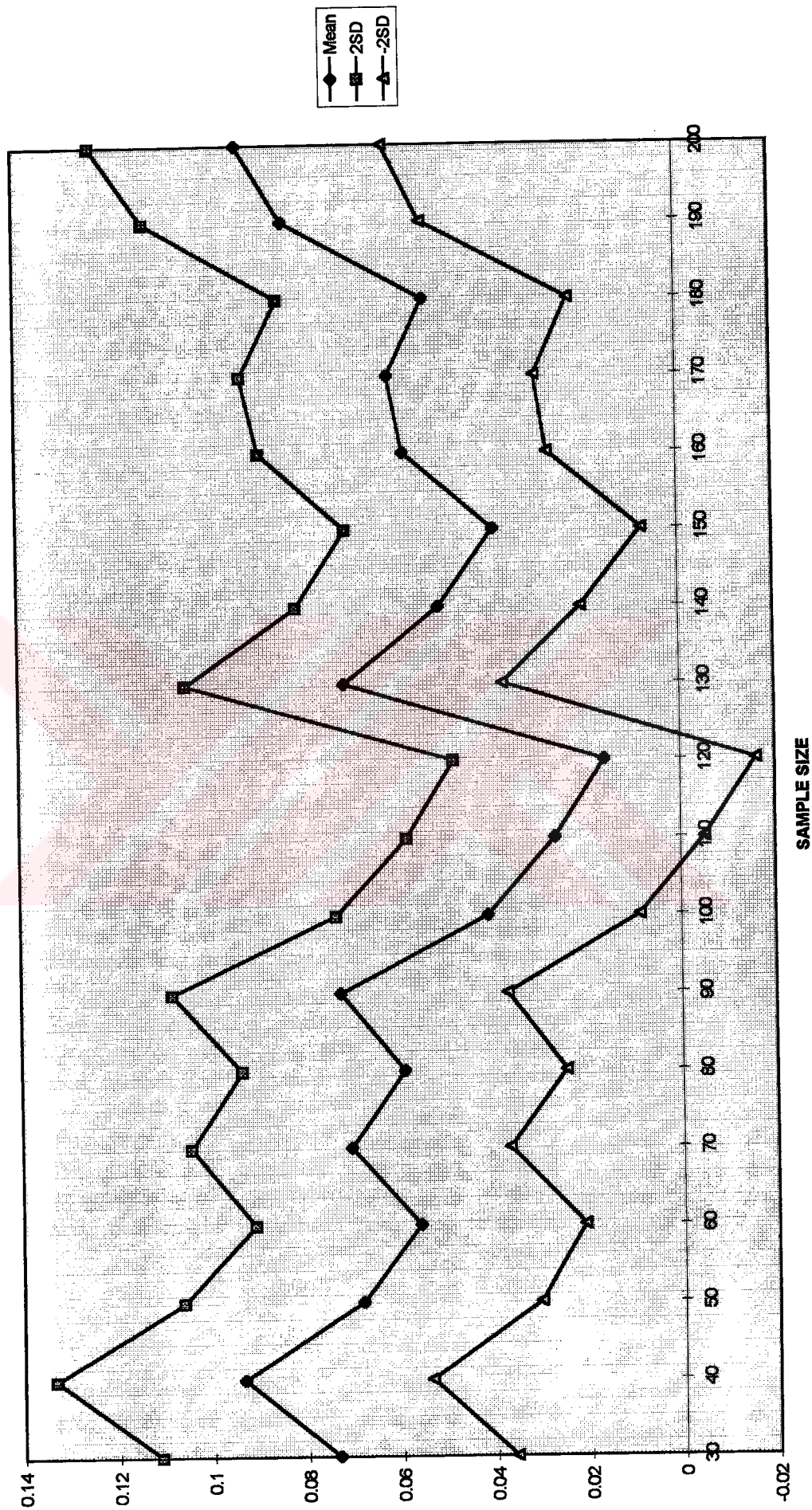


Figure 23b: Expected Value of $d_2 = \text{ABS}(\text{AIC}) - \text{ABS}(\text{AICCV10})$, Normal distribution

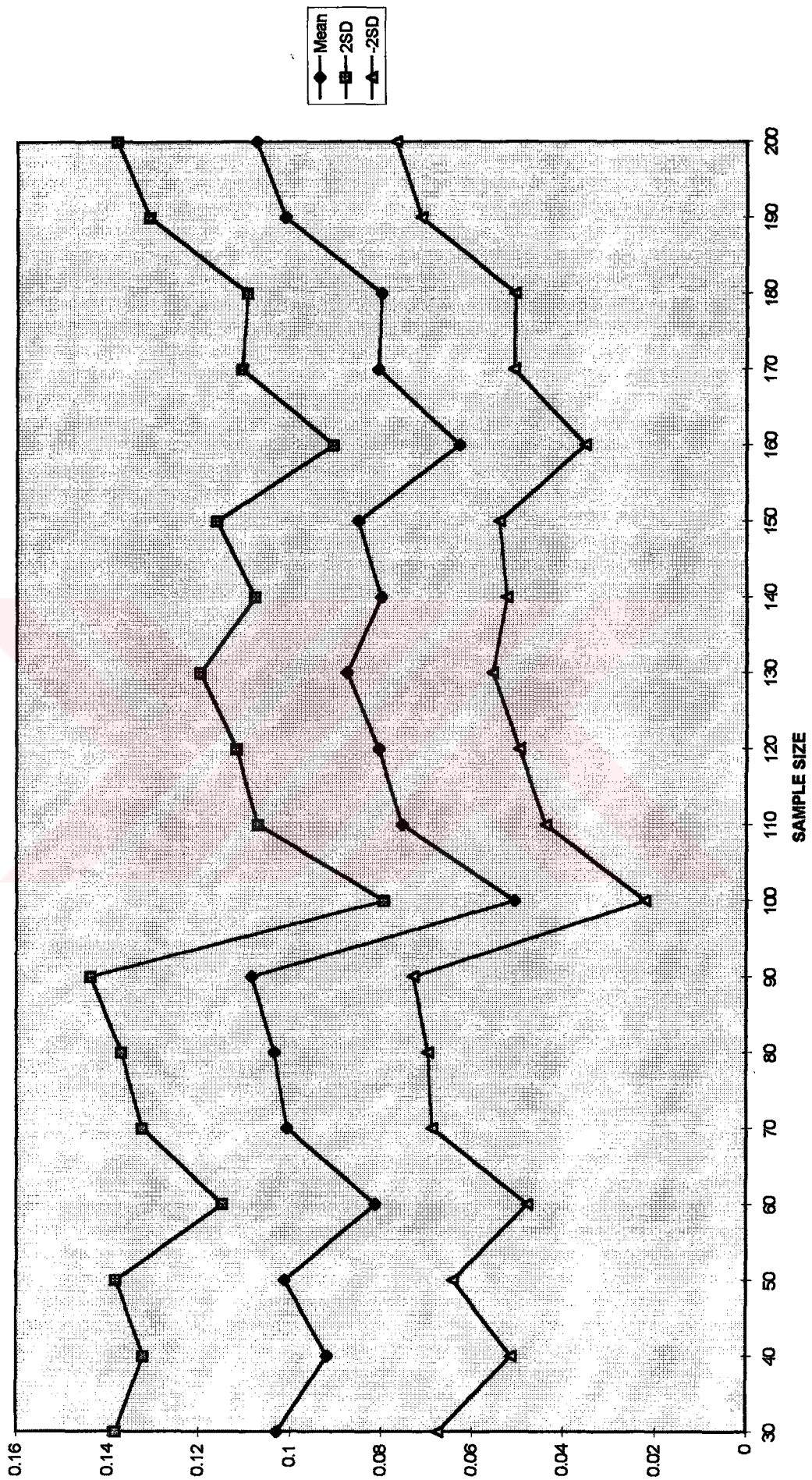


Figure 23c: Expected Value of $d_3 = \text{ABS}(\text{PRESS}) - \text{ABS}(\text{CV10})$, Normal distribution

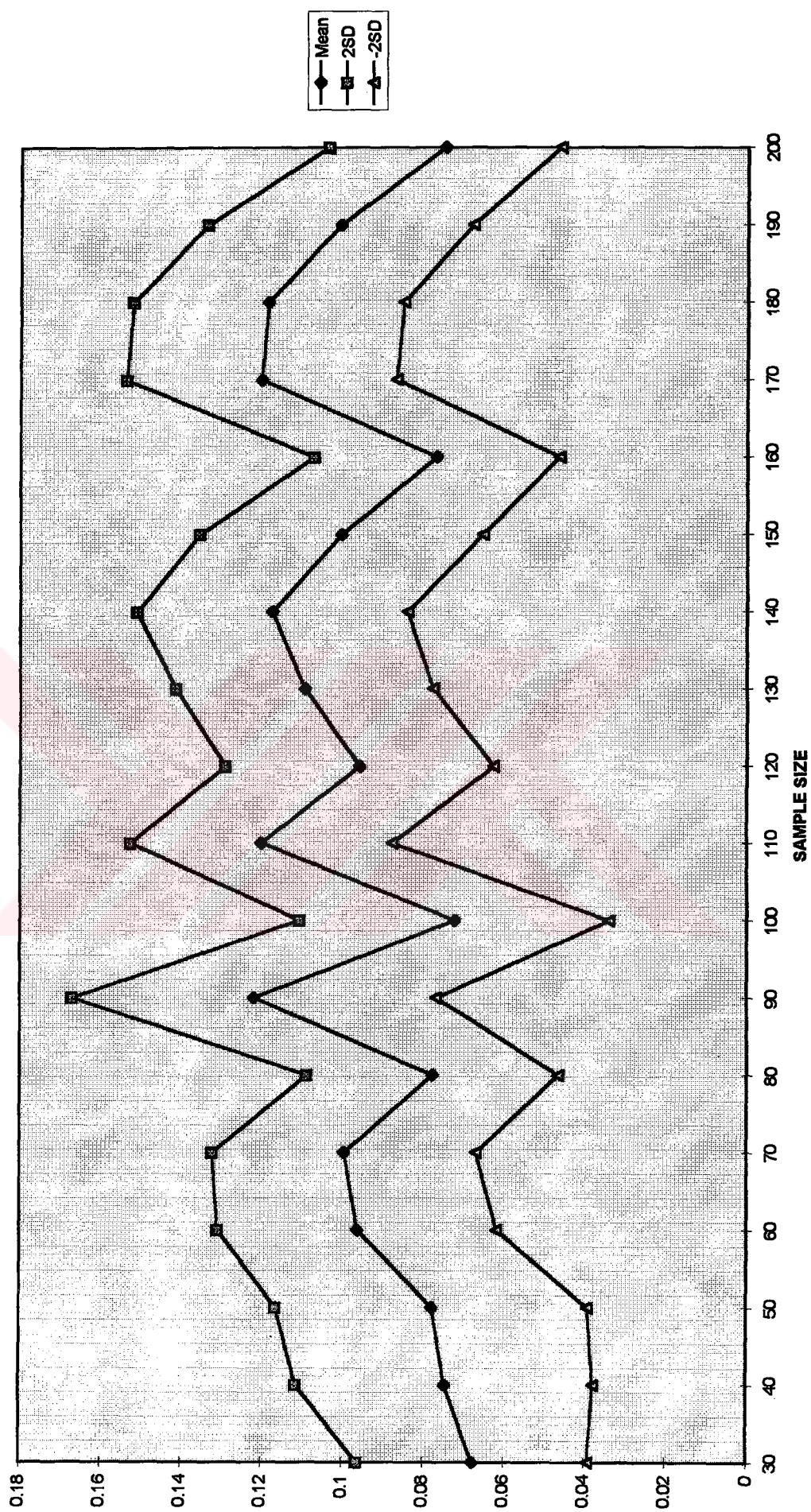


Figure 24a: Expected Value of $d_1 = \text{ABS}(\text{AIC}) - \text{ABS}(\text{AICPRESS})$, Skewed Distribution

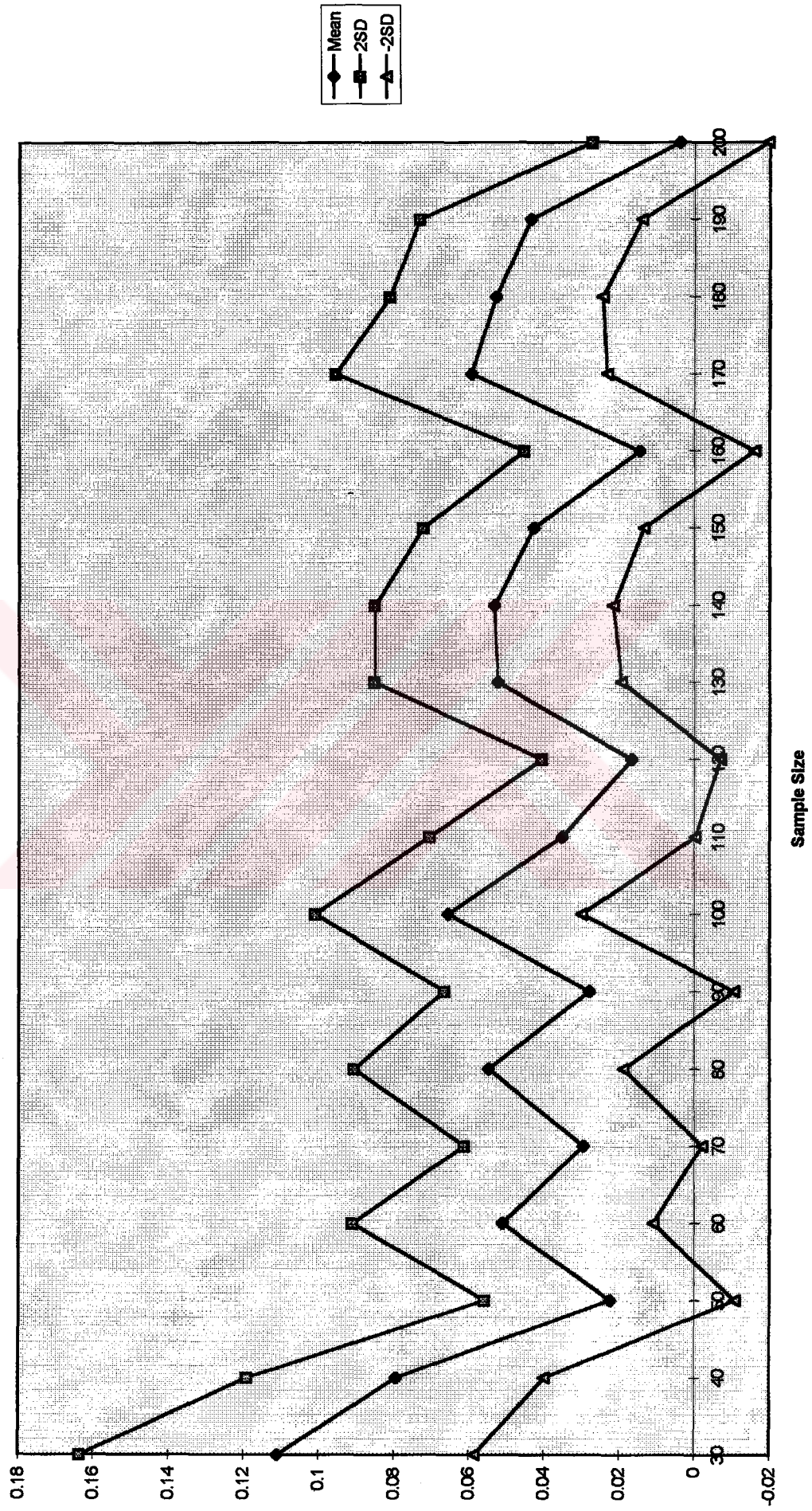


Figure 24b: Expected Value of $d_2 = \text{ABS}(\text{AIC}) - \text{ABS}(\text{ACCCV10})$, Skewed Distribution

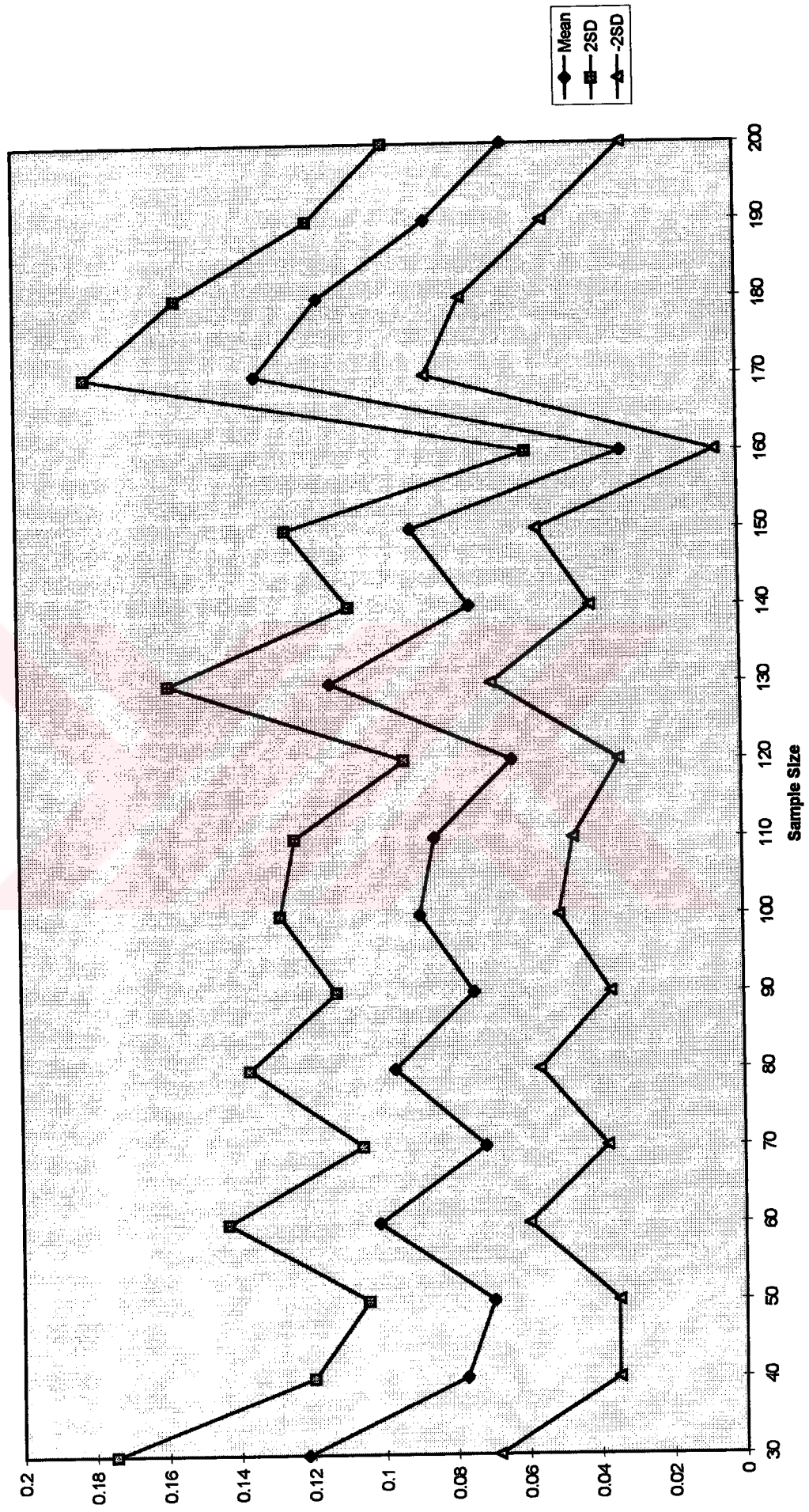


Figure 24c: Expected Value of $d_3 = \text{ABS(PRESS)} - \text{ABS(CV10)}$, Skewed Distribution

