ANALYSIS OF PREDICTIVE QUALITY

OF SOFTWARE RELIABILITY MODELS

by

Abdalla Ahmed ABDEL-GHALY

A Thesis Submitted for
The Degree of Doctor of Philosophy
in Statistics and Computer Science

DEPARTMENT OF MATHEMATICS
(STATISTICS)

The City University,
London,

APRIL 1986

CONTENTS

CHAPTER V. DATA ANALYSIS AND RESULTS

ACKNOWLEGEMENTS

The research described herein was undertaken in the Mathematics Department of The City University. During the course of the study I received valuable advice and helpful suggestions from several persons.

In particular, I wish to express my deepest gratitude to Professor B. Littlewood, the Director of the Centre for Software Reliability (C.S.R.) of The City University, who supervised this work, for his constant encouragement, his valuable advice and his guidance throughout the duration of the project.

My thanks also to the staff of C.S.R., The City University Computer Unit, and Mathematics Department.

Also I record my thanks to [redacted] for her skill typing of the manuscript.

Also I am grateful to the Department of Statistics of The Cairo University, the Egyptian Education Bureau in London and the Egyptian Government for the financial support of this study.

Finally, but not less significantly, I wish to express my thanks and love to my parents, my wife and my daughter for their patience and understanding during this work.

ABSTRACT

This thesis examines the problem of software reliability growth: how to measure it and how to know that the measures (predictions) are accurate.

Models in continuous time (i.e. complete inter-event time data) and discrete time (i.e. counts of numbers of events in successive observation periods) are considered. Several of these models are new, or are new versions of existing models.

Several statistical tools are presented which will allow a user of the models to analyse their respective merits (predictive accuracy) on a particular data set. Examples of the use of these tools on the predictions from several models on several real data sets are presented. The models perform with variable accuracy, which suggests that no model can be trusted to be of universal usefulness. The techniques presented here, then, form the beginnings of a tool-set which will enable a user to obtain reliability predictions for a particular context, and know that they are accurate.

KEY TO ABBREVIATIONS

BJM	Bayesian Jelinski-Moranda model
BL	Semi-Bayesian Littlewood model
cdf	Cumulative distribution function
D	Duane model
DD	Discrete Duane model
DJM	Discrete Jelinski-Moranda model
DJMNHPP	Discrete Non-homogeneous Poisson Process for Jelinski-Moranda model
DKL	Discrete Keiller-Littlewood model
DL	Discrete Littlewood model
DLNHPP	Discrete Non-homogeneous Poisson Process for Littlewood model
DLV	Discrete Littlewood-Verrall model
DW	Discrete Weibull model
i.i.d.	Identically independently distributed
IMTBF	Instantaneous Mean Time Between Failure
JM	Jelinski-Moranda model
JMNHPP	Non-homogeneous Poisson Process for Jelinski-Moranda model
KL	Keiller-Littlewood model
L	Littlewood model
LNHPP	Non-homogeneous Poisson Process for Littlewood model
LV	Littlewood-Verrall model
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimate
MTBF	Mean Time Between Failure
MTTF	Mean Time To Failure
NHPP	Non-homogeneous Poisson Process
pdf	Probability density function
pf	Probability function
PL	Prequential likelihood
PLR	Prequential likelihood ratio
ROCOF	Rate of Occurrence of Failure
W	Weibull model

CHAPTER I

INTRODUCTION

1.1. Background

Software reliability is defined as the probability of failure-free operation (of a software component or system) in a specified environment for a specified time. A failure is defined as an unacceptable departure of program operation from requirements. A fault is the software defect that causes a failure. The foregoing definition of software reliability is an operational one, it has been adopted because it offers the greatest utility to software engineers and managers, since it directly measures the impact on the user of failure of a system [Musa, 1980a].

Some of the models described here involve estimation of the number of residual faults in the software. Although the reliability is dependent upon this number, we shall show that the dependence is a complicated one and that no simple conclusions can be drawn about reliability from estimates of the number of remaining faults. We shall emphasise that it is the reliability itself, rather than such intermediate measures as fault-count, which are of importance.

The concern in this thesis will be exclusively with the measurement of software reliability. Although problems associated with the achievement of reliability are of immense importance, they will not be addressed here.

There is now a wide, and growing, acceptance of the importance of software reliability measurement. There are three main areas in which these techniques will be useful [Musa, 1980a, and 1980b].

Firstly, software reliability measures can be used to evaluate software engineering methodology. The difficulty of distinguishing between good and bad new methodology has led to a general resistance to change on the part of software managers and software engineers. Software reliability measures offer the promise of establishing at least one criterion for evaluating this new technology.

Secondly, a software reliability metric which can be established from actual test data on a single software product offers the possibility of evaluating status during the test phases of a project, and when the final product is shipped to a customer.

Finally, software reliability can be used as a means of monitoring the operational performance of software and controlling changes to software. Since a change usually involves a degradation of reliability, a reliability performance objective can be used as a means for determining when software changes will be allowed and perhaps even how large they can be.

Software reliability will often be achieved at the expense of some other characteristic of the product (program size, run time or response time, maintainability,....etc.) or the process of producing the product (cost, resource requirements, schedule,....etc.). A measurement technology is a vital prerequisite for making intelligent trade-offs in these areas.

The particular software reliability measurement problem to be considered in this thesis is the following. A system contains a certain number of faults, each of which eventually manifests itself at some time by causing a system failure. The system has to be redesigned in order to remove the cause of failure, i.e. to fix the fault which caused the failure. The system reliability will probably improve and the system should show reliability growth, at least in the long term.

Littlewood (1979a) treats the software failure process as a renewal process in continuous time, the successive renewals representing successive failures of the software. He assumes, for simplicity, that repairs are instantaneous. Such a process can be characterised either in terms of its successive inter-event times, t_1, t_2,....t_{i-1}, or via the number of events in fixed time intervals $n_1(\tau_1)$, $n_2(\tau_2)$,....$n_{i-1}(\tau_{i-1})$*. Reliability growth would show itself in such data set by, for example, a tendency for the $\{t_i\}$ sequence to be stochastically increasing. A user would be interested in using such a data set to predict future failure behaviour of the system being studied. These predictions will be statements about future (unobserved) random variables.

---

* t will be used instead of τ later in Chapter III.

Notice that the problem is difficult from a statistical point of view because of its intrinsic non-stationarity. It is precisely the structure of the reliability change which is of interest.

A user may express his predictions in various forms, depending upon his particular application. Examples are the reliability function $R_i(t) \equiv p(T_i > t)$, mean time to failure (MTTF) and rate of occurrence of failures (ROCOF), all of which may be calculated for the current or a future time. However, in all cases the problem is one of describing future $(T_i, T_{i+1}, \ldots)$ from the past $(t_1, t_2, \ldots t_{i-1})$ or future $(N_i(\tau_i), N_{i+1}(\tau_{i+1}), \ldots)$ from the past $(n_1(\tau_1), n_2(\tau_2), \ldots n_{i-1}(\tau_{i-1}))$. A <u>prediction system</u> to carry out this task comprises [Abdel-Ghaly, Chan and Littlewood, 1985a and 1985b]:

1. the probabilistic model which specifies the distribution of the T_j's ($N_j(\tau_j)$'s) conditional on a (unknown) parameter $\underline{\theta}$;
2. a statistical inference procedure for $\underline{\theta}$ involving the use of available data (realisation of T_j's ($N_j(\tau_j)$'s));
3. a prediction procedure combining (1) and (2) to allow the user to make probability statements about future T_j's ($N_j(\tau_j)$'s).

Of course, the <u>model</u> is an important part of this triad and it seems unlikely that good predictions can be obtained if the model is not in some sense "close to reality". However, a good model is not sufficient: stages (2) and (3) are vital components of the prediction system. In fact, disaster can strike at any of the three stages [Abdel-Ghaly, Chan and Littlewood, 1985b].

Analysing each of these three components separately is not possible for several reasons. In the first place, even the simplest models are too complicated for a conventional "goodness of fit" analysis of the model (stage 1 of the triad). The properties of the estimators of unknown parameters (stage 2) cannot be obtained exactly. Even the usual asymptotic theory of maximum likelihood breaks down in many models because the number of faults in the software (assumed finite) gives an upper bound to the sample size. For these reasons, a major part of the thesis concerns methods by which the accuracy of the predictions can be analysed directly.

1.2. Outline of the thesis

As stated earlier, this thesis will be concerned only with the software reliability growth problem. Other issues, such as structural problems, explanatory variables, costs, etc., will not be considered here.

The first aim of this study is to investigate the predictive ability of software reliability models. Thus in chapter II we describe some reliability growth models which can be used to analyse the inter-failure time data. These models are: Jelinski-Moranda model [Jelinski and Moranda, 1972], Bayesian versions of this model [Langberg and Singpurwalla, 1981; and Littlewood and Sofer, 1981], Littlewood model [Littlewood, 1981], its Bayesian version [new work], Littlewood-Verrall model [Littlewood and Verrall, 1973], Keiller-Littlewood model [Keiller, Littlewood, Miller and Sofer, 1983a], Duane model [Duane, 1964 and

Crow, 1977] the non-homogeneous Poisson process model for both Jelinski-Moranda and Littlewood models [Goel, 1980; Goel and Okumoto, 1979 and Littlewood, 1984] and finally the Weibull order statistic model [new work].

The second objective is to generalize these models (the Bayesian versions of Jelinski-Moranda and Littlewood models apart) to cope with discrete and summarized software data. This work, most of which is new, forms chapter III of the thesis.

Thirdly, in chapter IV we describe some statistical tools for the analysis of the accuracy of predictive quality. Although a considerable number of models have been developed during the past two decades, comparisons have tended to concentrate merely on analyses of the underlying assumptions. The intention in this thesis is to compare the predictive accuracy of the complete prediction systems.

Finally, the performance of all software reliability growth models considered in this research, as well as the statistical tools, will be examined for several sets of real failure data.

The final chapter contains conclusions and a summary of the current usefulness of these techniques together with a discussion of future research directions.

CHAPTER II

INTER-FAILURE TIME MODELS

2.1. Introduction

In this Chapter we shall discuss reliability growth models in which operating time between successive failures is a continuous random variable. When a fault manifests itself, by causing system failure, it is assumed that a repair attempt is made. The system is assumed to be returned to its operating environment immediately, i.e. we shall ignore repair times. Some of the models described here are well known and have been fairly widely used [Jelinski and Moranda, 1972; Littlewood and Sofer, 1981; Littlewood, 1981; Littlewood and Verrall, 1973; Keiller et at, 1983a; Duane, 1964; and Crow, 1977]; others are new.

In sections 2.2. and 2.3 we shall consider the early model proposed by Jelinski and Moranda (1972) and the Bayesian versions of this model due to Littlewood and Sofer (1981) and Langberg and Singpurwalla (1981).

In sections 2.4, 2.5, 2.6 and 2.7, Bayesian models will be considered for which the program is assumed to have faults of different size (i.e. rates). A distribution is assumed for the failure rate, λ_i, corresponding to the i^{th} inter-failure time random variable T_i. Details of this work can be found in Littlewood and Verrall (1973), Littlewood (1981), and Keiller, Littlewood, Miller and Sofer (1983a).

Non-homogeneous Poisson process models are considered in sections 2.8, 2.9 and 2.10. The earliest of these comes from an empirical observation of Duane (1964), which has been developed by Crow (1977). The others derive their rate functions by analogy with the models considered in previous sections.

Finally, in section 2.11 a new model is proposed based upon order statistics of Weibull random variables.

2.2. Jelinski-Moranda Model (JM)

This model seems to be the earliest one developed specifically for software reliability growth, although Shooman (1972) introduced a similar model almost contemporaneously. Jelinski and Moranda (1972) assumed that a program begins life with N faults, each having the same rate Φ. If we label the faults arbitrarily, it is assumed that fault i will show itself (by causing a failure), and be removed, after a time X_i, independently of other faults. The observed stochastic point process of failures is thus characterised by the order statistics of the random variables, $X_1, X_2, \ldots X_n$. That is, the first failure of the system will occur at time $T_1 \equiv X_{(1)}$, and in general the inter-failure time random variables $\{T_i\}$ are given by:

$$T_i = X_{(i)} - X_{(i-1)} \qquad (i=1,2,\ldots n) \qquad (2.2.1)$$

Alternatively, it is easy to see that when (i-1) failures have occurred, and hence this number of faults have been removed, the rate of occurrence of failures (ROCOF) of the system is $(N-i+1)\Phi$. It follows that the probability density function of T_i is:

$$f_i(t_i) = \lambda_i e^{-\lambda_i t_i} \qquad (2.2.2)$$

where $\lambda_i = (N-i+1)\Phi$

The likelihood function is:

$$L(t_1, t_2, \ldots . t_n / N, \Phi) = \prod_{i=1}^{n} f_i(t_i)$$

$$= \prod_{i=1}^{n} \left[(N-i+1)\Phi \, e^{-(N-i+1)\Phi t_i} \right] \qquad (2.2.3)$$

if the realisation terminates with a failure.

In order to get the maximum likelihood estimate (MLE) of (N,Φ), the natural logarithm will be taken for equation (2.2.3) which gives:

$$\ell(t_1, t_2, \ldots, t_n / N, \Phi) = \log[L(t_1, t_2, \ldots t_n / N, \Phi)]$$

$$= \sum_{i=1}^{n} \log(N-i+1) + n \log\Phi - \Phi \sum_{i=1}^{n} (N-i+1) t_i \qquad (2.2.4)$$

Littlewood and Verrall (1981) showed that this likelihood function will have a unique maximum at <u>finite</u> N and <u>non-zero</u> Φ if and only if:

$$\frac{\sum_{i=1}^{n} (i-1) t_i}{\sum_{i=1}^{n} (i-1)} > \frac{\sum_{i=1}^{n} t_i}{n} \quad , \qquad (2.2.5)$$

Otherwise the likelihood has its maximum at finite $\lambda = N\Phi$ for infinite N. Intuitively this means that the model can only be sensibly used when reliability growth is present in the data (Condition (2.2.5) states that the least squares regression line of t_i on i has positive slope, i.e. that the inter-failure times are tending to increase). If there is evidence of reliability decay, the best the model can do is suggest fault-fixes have no effect, i.e. that the inter-event times are i.i.d random variables. Littlewood and Verrall (1981) recommend that this test be carried out first, and only if the data set passes should the MLE for (Φ,N) be found in the usual way by differentiating (2.2.4) with respect to Φ and N giving:

$$\hat{\Phi} = \frac{n}{\sum_{i=1}^{n} (\hat{N}-i+1)t_i} \qquad (2.2.6)$$

and

$$\sum_{i=1}^{n} \frac{1}{(\hat{N}-i+1)} - \frac{n\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} (\hat{N}-i-1)t_i} = 0 \qquad (2.2.7)$$

The solution of the last equation can be found numerically, and $\hat{\Phi}$ calculated from (2.2.6). Alternatively (2.2.4) could be maximized numerically using an optimization routine to find the value of $\hat{N}$, $\hat{\Phi}$.

The MLE of (N,Φ) will be used to predict future failure behaviour of the system. For example, if we were interested in the time to next failure, T_{n+1}, after having seen (and removed) n faults, we could find the cdf, pdf, rate function, mean, median, etc.

The predictive pdf of T_{n+1} is:

$$\hat{f}(t_{n+1}) = (\hat{N} - n)\hat{\Phi} \exp\{-(\hat{N}-n)\hat{\Phi}t_{n+1}\} \tag{2.2.8}$$

and the predictive cdf is given by:

$$\hat{F}(t_{n+1}) = 1 - \exp\{-(\hat{N}-n)\hat{\Phi}t_{n+1}\} \tag{2.2.9}$$

From the last two equations we can get the rate function using its definition [Barlow and Proschan, 1975]:

$$\lambda_{n+1}(t) = \frac{f(t)}{1 - F(t)} \equiv \lim_{\delta \to 0} \frac{\Pr\{t < T_{n+1} < t+\delta / T_{n+1} > t\}}{\delta} = (\hat{N}-n)\hat{\Phi} \tag{2.2.10}$$

The MTTF is:

$$\frac{1}{(\hat{N}-n)\,\hat{\Phi}} \tag{2.2.11}$$

The median of the predictive distribution is:

$$\frac{\log 2}{(\hat{N}-n)\,\hat{\Phi}} \tag{2.2.12}$$

This model has been criticised for two related reasons: that the sequence of successive rates for the system is deterministic, and that fault rates are assumed to be equal. We shall discuss this issue in more detail in Chapter V.

2.3. Bayesian Jelinski-Moranda Model (BJM)

Two ways have been suggested of doing Bayesian analysis of the JM model. The direct method assumed a prior for N or Φ singly, or for (N,Φ) jointly, and was proposed by Langberg and Singpurwalla (1981). The other method involves a reparameterisation to (λ,Φ), where λ = NΦ. This approach by Littlewood and Sofer (1981) involves a slight modelling change. Both approaches will be described here, but our data analysis in later sections will only use the second.

2.3.1. Langberg-Singpurwalla Method

Langberg and Singpurwalla assume N,Φ are a *priori* independent and suggest three different cases.

Case 1: It is assumed that N has a Poisson distribution with parameter μ:

$$P(N) = \frac{\mu^N}{N!} e^{-\mu} \tag{2.3.1}$$

and Φ is degenerate at known Φ.

For the data vector $t_1, t_2, \ldots, t_n$, the likelihood function, $L(t_1, t_2, \ldots t_n / N, \Phi)$ is given by (2.2.3), so using Bayes theorem the posterior probability function of N is given by:

$$p^*(N/t_1, \ldots t_n, \Phi) = cL(t_1, t_2, \ldots t_n/N, \Phi)p(N)$$

where

$$c^{-1} = \sum_{N=n}^{\infty} L(t_1, t_2, \ldots t_n/N, \Phi)p(N)$$

$$= (\Phi\mu)^n \exp\{-\mu - \Phi \sum_{i=1}^{n} (n-i+1)t_i + \mu e^{-\Phi \sum_{i=1}^{n} t_i}\}$$

It follows that the posterior distribution of the number of remaining faults is Poisson:

$$p^*(N-n/t_1, \ldots t_n, \Phi) =$$

$$\frac{\left[\mu e^{-\Phi \sum_{i=1}^{n} t_i}\right]^{N-n}}{(N-n)!} \exp\{-\mu e^{-\Phi \sum_{i=1}^{n} t_i}\} \tag{2.3.2}$$

The conditional reliability function is:

$$R(t/N,\Phi) = 1 - F(t/N,\Phi) = \exp\{-(N-n)\Phi t\}$$

The posterior reliability function is:

$$R^*(t/t_1,\ldots,t_n\Phi) = P(T_{n+1} > t/t_1,\ldots,t_n,\Phi)$$

$$= \sum_{N=n}^{\infty} R(t/N,\Phi)p^*(N-n/t_1,\ldots,t_n,\Phi)$$

$$= \exp\{-\mu\, e^{-\Phi\tau_n}(1 - e^{-\Phi t})\} \qquad (2.3.3)$$

where $\tau_n = \sum_{i=1}^{n} t_i$

We shall return to this case in section (2.8) where we consider the non-homogeneous Poisson process analogue of the JM model.

Case 2: It is assumed that N is a known value and Φ has a gamma distribution with scale parameter β and shape parameter α, i.e.:

$$f(\Phi) = \frac{\beta^\alpha}{\Gamma\alpha}\Phi^{\alpha-1}e^{-\beta\Phi} \qquad (2.3.4)$$

It follows that Φ has a gamma posterior density:

$$f^*(\Phi/t_1,\ldots t_n,N) = \frac{\left[\beta + \sum_{i=1}^{n}(N-i+1)t_i\right]^{(n+\alpha)}}{\Gamma(n+\alpha)}\Phi^{n+\alpha-1}$$

$$.\exp\{-\Phi(\beta + \sum_{i=1}^{n}(N-i+1)t_i)\} \qquad (2.3.5)$$

The posterior reliability function is:

$$R^*(t/t_1,\ldots.t_n,N) = \left[\frac{\beta + \sum_{i=1}^{n}(N-i+1)t_i}{\beta + \sum_{i=1}^{n}(N-i+1)t_i + (N-n)t}\right]^{(n+\alpha)} \qquad (2.3.6)$$

Case 3: It is assumed that N has any specified prior distribution, p(N), and Φ has a gamma distribution as above, with N,Φ, independent:

$$p(N,\Phi) = \frac{\beta^{\alpha}}{\Gamma\alpha}\,\Phi^{\alpha-1}e^{-\beta\Phi}\,p(N) \qquad (2.3.7)$$

It follows that the posterior joint distributionof (N,Φ) is given by:

$$f^*(N,\Phi/t_1,\ldots t_n) = \frac{\left[\beta + \sum_{i=1}^{n}(N-i+1)t_i\right]^{n+\alpha}}{\Gamma(n+\alpha)}\,\Phi^{n+\alpha-1}$$

$$.\exp\{-\Phi(\beta + \sum_{i=1}^{n}(N-i+1)t_i)\}$$

$$\text{and}\ .\ \frac{N!}{(N-n)!}\,\frac{p(N)\left[\beta + \sum_{i=1}^{n}(N-i+1)t_i\right]^{-(n+\alpha)}}{\sum_{q=n}^{\infty}\frac{q!}{(q-n)!}\,p(N=q)\left[\beta + \sum_{i=1}^{n}(q-i+1)t_i\right]^{-(n+\alpha)}} \qquad (2.3.8)$$

The problems with this method are two-fold. In the first place, Cases 1 and 2 unrealistically assume that the user knows the true value of one of the parameters. Secondly, the complete version, Case 3, results in an intractable joint posterior distribution (2.3.8). A quasi-Bayesian approach to the first problem might be to use the "maximum likelihood" estimate of the unknown parameter based on the marginal posterior distribution. It may also be possible to approximate to the infinite sum in the denominator of (2.3.8). We shall, instead, use the following approach.

2.3.2. Littlewood Sofer Method

Littlewood and Sofer (1981) suggested a reparameterisation of the JM model, to (λ,Φ). Here λ is taken to be the initial rate of occurrence of failures. At the first repair the rate becomes $\lambda-\Phi$, and subsequently drops to $\lambda-2\Phi$, $\lambda-3\Phi$,..... This involves a slight modelling change, since it is not assumed that λ is an integer multiple of Φ: the last repair removes a rate which may be smaller than Φ. The likelihood is:

$$L(t_1,t_2,\ldots.t_n/\lambda,\Phi) = \prod_{i=1}^{n}(\lambda-(i-1)\Phi)\exp\{-\sum_{i=1}^{n}(\lambda-(i-1)\Phi)t_i\} \tag{2.3.9}$$

It is assumed that λ and Φ are independent and each has a gamma prior distribution:

$$f(\lambda) = \frac{c^b\lambda^{b-1}}{\Gamma b}\, e^{-c\lambda} \tag{2.3.10}$$

and

$$f(\Phi) = \frac{g^f\Phi^{f-1}}{\Gamma f}\, e^{-g\Phi} \tag{2.3.11}$$

The parameters c, b, g and f, which are positive, are chosen by the user. For reasons of mathematical tractability, b is restricted to integer values. Littlewood and Sofer (1981) use the improper, non-informative prior for λ and Φ, by letting $c,g\to 0$ and $f=b=1$:

$$f(\lambda,\Phi) = 1 \tag{2.3.12}$$

If we define:

$$\prod_{i=1}^{n} (X-i) = \sum_{i=0}^{n} a_{i,n} \, x^{n+i} \qquad (2.3.13)$$

where:

$$a_{j,n} = n \, a_{j-1,n-1} + a_{j,n-1} \qquad j \geqslant 1$$

$$a_{0,1} = 1, \quad a_{1,1} = 1, \quad a_{0,n} = 1 \quad \forall \, n \qquad (2.3.14)$$

then from (2.3.9) and (2.3.12) we can get the posterior joint density function of λ, Φ:

$$f^*(\lambda, \Phi / t_1, \ldots . t_n) = \frac{\sum_{j=0}^{n-1} a_{j,n-1} (-1)^j \Phi^j \lambda^{n-j} \exp\{-\lambda \sum_{i=1}^{n} t_i + \Phi \sum_{i=1}^{n} (i-1) t_i\}}{\sum_{j=0}^{n-1} a_{j,n-1} \dfrac{j! \, (n-j)!}{\left[\sum_{i=1}^{n} (n-i) t_i\right]^{j+1} \left[\sum_{i=1}^{n} t_i\right]^{n-j+1}}} \qquad (2.3.15)$$

for $\lambda > (n-1)\Phi$.

and zero otherwise.

If we are interested in the current reliability immediately after the occurrence of the n^{th} failure (and removal of n^{th} fault) we proceed as follows:

$$\begin{aligned} R(t_{n+1}/\lambda, \Phi) &= P(T_{n+1} > t_{n+1}/\lambda, \Phi) \\ &= \exp\{-(\lambda - n\Phi) t_{n+1}\} \qquad \text{if } \lambda > n\Phi \\ &= 1 \qquad \text{if } (n-1) < \lambda < n\Phi \end{aligned} \qquad (2.3.16)$$

The posterior reliability function is then, using (2.3.15) and (2.3.16):

$$R(t_{n+1}/t_1,\ldots t_n) = \iint R(t_{n+1}/\lambda,\Phi)f^*(\lambda,\Phi/t_1,\ldots t_n)d\lambda d\Phi$$

i.e. $$R(t_{n+1}/t_1,\ldots t_n) = c\left\{\sum_{j=o}^{n} a_{j,n} \frac{(n-j)!\ \ j!}{\left[t_{n+1}+\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i+1)t_i\right]^{j+1}}\right.$$

$$+\sum_{j=o}^{n-1} a_{j,n-1} \frac{(n-j)!\ \ j!}{\left[\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i)t_i\right]^{j+1}}$$

$$\left.-\sum_{j=o}^{n} a_{j,n} \frac{(n-j)!\ \ j!}{\left[\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i+1)t_i\right]^{j+1}}\right\} \qquad (2.3.17)$$

where

$$c^{-1} = \sum_{j=o}^{n-1} a_{j,n-1} \frac{(n-j)!\ \ j!}{\left[\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i)t_i\right]^{j+1}} \qquad (2.3.18)$$

At each stage there is a finite probability that the program is "perfect", i.e. that the last fix removed the only remaining fault. After the n^{th} fix this probability is:

$$P_o = 1 - c\sum_{j=o}^{n} a_{j,n} \frac{(n-j)!\ j!}{\left[\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i+1)t_i\right]^{j+1}} \qquad (2.3.19)$$

where c^{-1} is given by (2.3.18).

The posterior cdf of T_{n+1} can be obtained from (2.3.17), and so by differentiation we obtain the posterior pdf:

$$f^*(t_{n+1}/t_1,\ldots.t_n) = c\sum_{j=o}^{n} a_{j,n} \frac{(n-j+1)!\, j!}{\left[t_{n+1}+\sum_{i=1}^{n} t_i\right]^{n-j+2}\left[\sum_{i=1}^{n}(n-i+1)t_i\right]^{j+1}} \tag{2.3.20}$$

Then from (2.3.20) and (2.3.17), by division, we get the posterior rate function:

$$\lambda(t_{n+1}/t_1,\ldots t_n) = \sum_{j=o}^{n} a_{j,n} \frac{(n-j+1)!\; j!}{\left[t_{n+1}+\sum_{i=1}^{n} t_i\right]^{n-j+2}\left[\sum_{i=1}^{n}(n-i+1)t_i\right]^{j+1}} \Bigg/$$

$$\left\{\sum_{j=o}^{n} a_{j,n} \frac{(n-j)!\; j!}{\left[t_{n+1}+\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i+1)t_i\right]^{j+1}} - \sum_{j=1}^{n-1} a_{j,n-1} \frac{(n-j)!\; j!}{\left[\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i)t_i\right]^{j+1}} - \sum_{j=1}^{n} a_{j,n} \frac{(n-j)!\; j!}{\left[\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i+1)t_i\right]^{j+1}}\right\} \tag{2.3.21}$$

and if we let $t_{n+1} \rightarrow 0$ in (2.3.21), we get the posterior ROCOF evaluated at the epoch immediately following the n^{th} fix:

$$\text{ROCOF} = \frac{\sum_{j=o}^{n} a_{j,n} \frac{(n-j+1)!\quad j!}{\left[\sum_{i=1}^{n} t_i\right]^{n-j+2}\left[\sum_{i=1}^{n}(n-i+1)t_i\right]^{j+1}}}{\sum_{j=o}^{n-1} a_{j,n-1} \frac{(n-j)!\; j!}{\left[\sum_{i=1}^{n} t_i\right]^{n-j+1}\left[\sum_{i=1}^{n}(n-i)t_i\right]^{j+1}}} \tag{2.3.22}$$

It is easy to see that the MTTF does not exist because the posterior pdf of T_{n+1} is a linear combination of Pareto densities, one of which has a parameter equal to one. However, we can define the "instantaneous mean time between failure" (IMTBF) as the reciprocal of current ROCOF.

Although some of the expressions for reliability metrics are complicated, all are available in finite closed forms. This results in a more efficient algorithm than the search techniques required for ML estimators in some other models.

2.4. Littlewood Model (L)

A major criticism of the JM model (and its Bayesian variants) is that it assumes all faults to be of equal "size" (rate). This seems extremely implausible. Littlewood (1981) has suggested that if the fault rates are unequal there will be a tendency for earlier fixes to have greater effect than later ones. He proposed a model in which it is assumed that the program starts life with N faults, each of them with a different failure rate Φ_i. Thus at stage i, when (i-1) faults have been removed, the pdf of T_i is:

$$f(t_i/\lambda_i) = \lambda_i \exp\{-\lambda_i t_i\} \qquad (2.4.1)$$

where

$\lambda_i = \sum_{j=1}^{N-i+1} \Phi_j$, and $\Phi_1, \ldots\ldots \Phi_{N-i+1}$ are the (non-identical) rates of the remaining faults.

The novelty of this approach is that the Φ's are regarded as realisations of random variables. The occurrence rates of the N faults with which the program begins life can be regarded as i.i.d. random variables, Φ_i, from a gamma distribution:

$$f(\Phi) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \Phi^{\alpha-1} \exp\{-\beta\Phi\} \tag{2.4.2}$$

Faults which show themselves earlier in the debugging will tend to be the larger ones, thus at later stages of the debugging the remaining faults will tend to be smaller. This can be formalised by considering the distribution of the magnitude of a fault remaining when time τ has elapsed. Using Bayes theorem, this can be shown to be:

$$f(\Phi/\text{fault has survived for time } \tau)$$

$$= \frac{(\beta+\tau)^{\alpha}}{\Gamma(\alpha)} \Phi^{\alpha-1} \exp\{-\Phi(\beta+\tau)\} \tag{2.4.3}$$

i.e. a gamma $(\alpha,\beta+\tau)$ distribution. The average size (occurrence rate) is now $\alpha/(\beta+\tau)$, smaller, as expected, than the α/β at time zero.

Considering now the rate of occurrence of failures for the program, it is clear that λ_i is the sum of (N-i+1) independent random variables, each with gamma density (2.4.3). Thus the posterior density of λ_i is gamma $((N-i+1)\alpha, \beta+\tau)$:

i.e.

$$f(\lambda_i/\tau_{i-1},\beta,\alpha) = \frac{(\beta+\tau_{i-1})^{\alpha(N-i+1)}}{\Gamma\alpha(N-i+1)} \lambda_i^{\alpha(N-i+1)-1}$$

$$.\exp\{-\lambda_i(\beta+\tau_{i-1})\} \qquad (2.4.4)$$

Then mixing (2.4.4) and (2.4.1), and integrating for λ_i, we get the pdf of T_i as:

$$f(t_i/\tau_{i-1},\beta,\alpha) = \frac{(N-i+1)\alpha[\beta+\tau_{i-1}]^{(N-i+1)\alpha}}{[\beta+\tau_{i-1}+t_i]^{(N-i+1)\alpha+1}} \qquad (2.4.5)$$

which is a Pareto distribution.

This result can be obtained from the order statistic formulation of the model. If the times to detection (and removal) of the faults are the conditionally independent exponential random variables, X_i, where X_i has a rate Φ_i, with gamma (α,β) and independent of the other rates, it is easy to see that the unconditional distribution of X_i is Pareto

$$f(x_i) = \frac{\alpha\beta^{\alpha}}{(\beta+x_i)^{\alpha+1}}, \text{ independently for different } X_i.$$

The observed inter-event times are then, again, the spacings between the order statistics $X_{(i)}$. It is interesting how this parallels the exponential case: the spacings between o.s of i.i.d. Pareto are themselves Pareto, (2.4.5), but not independent (because the τ_{i-1} term in (2.4.5) is the sum of previous t_j's).

The likelihood function, assuming the realisation terminates in a failure is:

$$L(t_1,t_2,\ldots.t_n/N,\alpha,\beta) = \prod_{i=1}^{n} f(t_i/\tau_{i-1},N,\alpha,\beta)$$

$$= \prod_{i=1}^{n} \frac{(N-i+1)\alpha(\beta+\tau_{i-1})^{(N-i+1)\alpha}}{(\beta+\tau_{i-1}+t_i)^{(N-i+1)\alpha+1}} \qquad (2.4.6)$$

with log likelihood:

$$\ell(t_1,t_2,\ldots..t_n/N,\alpha,\beta) = \log[L(t_1,t_2\ldots.t_n/N,\alpha,\beta)]$$

$$= \sum_{i=1}^{n} \log(N-i+1) + n\log\alpha + N\alpha\log\beta$$

$$-(\alpha+1)\sum_{i=1}^{n} \log(\beta+\tau_{i-1}+t_i)-(N-n)\alpha\log(\beta+\tau_{n-1}+t_n) \qquad (2.4.7)$$

The MLE of α can be found analytically as a function $\hat{N}$, $\hat{\beta}$:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n}\log(\hat{\beta}+\tau_{i-1}+t_i)+(\hat{N}-n)\log(\hat{\beta}+\tau_{n-1}+t_n)-\hat{N}\log\hat{\beta}} \qquad (2.4.8)$$

This can be substituted into (2.4.7), so that the numerical optimisation of the likelihood occurs in 2-space. Then these MLE $\hat{N}$, $\hat{\beta}$, $\hat{\alpha}$ of N, β, α can be used to predict the pdf and cdf for the next failure time by substitution in (2.4.5):

$$\hat{f}(t_{n+1}/\tau_n) = \frac{(\hat{N}-n)\hat{\alpha}(\hat{\beta}+\tau_n)^{(\hat{N}-n)\hat{\alpha}}}{(\hat{\beta}+\tau_n+t_{n+1})^{(\hat{N}-n)\hat{\alpha}+1}} \tag{2.4.9}$$

and

$$\hat{F}(t_{n+1}/\tau_n) = 1 - \left[\frac{\hat{\beta} + \tau_n}{\hat{\beta} + \tau_n + t_{n+1}}\right]^{(\hat{N}-n)\hat{\alpha}} \tag{2.4.10}$$

Then, using the last two forms, we can get the rate function at (t_{n+1}):

$$\hat{\lambda}(t_{n+1}) = \frac{(\hat{N}-n)\hat{\alpha}}{\hat{\beta} + \tau_n + t_{n+1}} \tag{2.4.11}$$

and if $t_{n+1} = 0$ we get the current ROCOF:

$$\text{ROCOF} = \frac{(\hat{N}-n)\hat{\alpha}}{\hat{\beta} + \tau_n} \tag{2.4.12}$$

The MTTF, $E(T_{n+1})$ is:

$$\text{MTTF} = \frac{\hat{\beta} + \tau_n}{(\hat{N}-n)\hat{\alpha} - 1}$$

and this measure will exist as long a $(\hat{N}-n)\hat{\alpha} > 1$.

Also the median is given by:

$$\text{Median} = (\beta + \tau_n)\left[(.5)^{-\frac{1}{(\hat{N}-n)\hat{\alpha}}} - 1\right] \tag{2.4.13}$$

Longer term prediction can be conducted by substitution into suitable conditional expressions.

2.5. Semi-Bayesian Littlewood Model (BL)

It is very hard to do a full Bayesian analysis for L model because of the role played by β in the density (and hence likelihood). For simplicity we assume β to be known in the following analysis and a MLE method will be used later to estimate it. The analysis will follow the similar one for the BJM model.

2.5.1. Langberg-Singpurwalla Method

Case 1: Let N have a Poisson distribution with parameter μ (2.3.1) and α a known value. Then to find the posterior distribution of N, we have from (2.3.1) and (2.4.6):

$$p^*(N-n/t_1,\ldots.t_n,\alpha,\beta) = \frac{\left[\mu \exp\{-\alpha \log\left\{\frac{\beta+\tau_{n-1}+t_n}{\beta}\right\}\right]^{N-n}}{(N-n)!} .\exp\left\{-\mu \exp\left[-\alpha \log \left\{\frac{\beta+\tau_{n-1}+t_n}{\beta}\right\}\right]\right\} \qquad (2.5.1)$$

From (2.4.10) the reliability function $R(t_{n+1})$ is:

$$R(t_{n+1}/N,\alpha,\beta,\tau_n) = \left[\frac{\beta+\tau_n}{\beta+\tau_n+t_{n+1}}\right]^{(N-n)\alpha} \qquad (2.5.2)$$

So the unconditional reliability function is, from (2.5.1) and (2.5.2):

$$R(t_{n+1}/\alpha,\beta,\tau_n) = \exp\left\{-\mu\left\{\frac{\beta}{\beta+\tau_n}\right\}^{\alpha}\left[1-\left\{\frac{\beta+\tau_n}{\beta+\tau_n+t_{n+1}}\right\}^{\alpha}\right]\right\} \qquad (2.5.3)$$

A full analysis related to this case will be given later in section (2.10) under the non-homogeneous Poisson process for L model.

Case 2: Let α have a gamma density with parameters γ, θ, and N a known value. Then the pdf of α is:

$$P(\alpha/\gamma,\theta) = \frac{\theta^{\gamma}}{\Gamma\gamma}\,\alpha^{\gamma-1}\,\exp\{-\theta\alpha\} \qquad (2.5.4)$$

Then α has a posterior gamma density:

$$f^{*}(\alpha/t_1,t_2,\ldots\ldots t_n,N,\beta) = \frac{\left[\theta+\sum_{i=1}^{n}(N-i+1)\log\left\{\frac{\beta+\tau_{i-1}+t_i}{\beta+\tau_{i-1}}\right\}\right]^{n+\gamma}}{\Gamma(n+\gamma)}$$

$$.\alpha^{n+\gamma-1}\exp\left\{-\alpha\left\{\theta+\sum_{i=1}^{n}(N-i+1)\log\left\{\frac{\beta+\tau_{i-1}+t_i}{\beta+\tau_{i-1}}\right\}\right\}\right\} \qquad (2.5.5)$$

The unconditional reliability function is given by:

$$R(t_{n+1}/t_1,\ldots\ldots t_n,N,\beta)$$

$$= \left[\frac{\theta+\sum_{i=1}^{n}(N-i+1)\log\left\{\frac{\beta+\tau_{i-1}+t_i}{\beta+\tau_{i-1}}\right\}}{\theta+\sum_{i=1}^{n}(N-i+1)\log\left\{\frac{\beta+\tau_{i-1}+t_i}{\beta+\tau_{i-1}}\right\}+(N-n)\log\left\{\frac{\beta+\tau_n+t_{n+1}}{\beta+\tau_n}\right\}}\right] \qquad (2.5.6)$$

<u>Case 3:</u> Let α be gamma distributed (2.5.4), and N have any specified prior probability function p(N), and assume N,α to be independent. Again we can find the joint posterior for (N,α) in the following form:

$$p^*(N,\alpha/t_1,\ldots t_n,\beta) = \frac{\left[\theta + \sum_{i=1}^{n}(N-i+1)\log\left\{\frac{\beta+\tau_{i-1}+t_i}{\beta+\tau_{i-1}}\right\}\right]^{n+\gamma}}{\Gamma(n+\gamma)}$$

$$.\alpha^{\,n+\gamma-1}\exp\left\{-\alpha\left[\theta + \sum_{i=1}^{n}(N-i+1)\log\left\{\frac{\beta+\tau_{i-1}+t_i}{\beta+\tau_{i-1}}\right\}\right]\right\}$$

$$.\frac{\frac{N!}{(N-n)!}\,p(N)\left[\theta + \sum_{i=1}^{n}(N-i+1)\log\left\{\frac{\beta+\tau_{i-1}+t_i}{\beta+\tau_{i-1}}\right\}\right]^{-(n+\gamma)}}{\sum_{q=n}^{\infty}\frac{q!}{(q-n)!}\,p(N=q)\left[\theta + \sum_{i=1}^{n}(q-i+1)\log\left\{\frac{\beta+\tau_{i-1}+t_i}{\beta+\tau_{i-1}}\right\}\right]^{-(n+\gamma)}} \tag{2.5.7}$$

Finally, as in section (2.3.1), this method is intractable and for that we shall, instead, use the following approach.

<u>2.5.2. Littlewood-Sofer Method</u>

The reparameterisation suggested by Littlewood and Sofer (1981) for the Bayesian JM model will be used here by letting $\lambda=N\alpha$ and $\phi=\alpha$. (Note β is assumed to be known). Then the probability density function of T_i in (2.4.5) can be rewritten in a convenient exponential form as:

$$f(t_i/\lambda,\phi,\beta) = \frac{\lambda-\phi(i-1)}{\beta+T_{i-1}+t_i}\exp\left\{-(\lambda-\phi(i-1))\log\left\{\frac{\beta+T_{i-1}+t_i}{\beta+T_{i-1}}\right\}\right\} \tag{2.5.8}$$

which produces the likelihood function:

$$L(t_1,\ldots t_n/\lambda,\phi,\beta) = \prod_{i=1}^{n}\frac{(\lambda-\phi(i-1))}{(\beta+T_{i-1}+t_i)}\exp\left\{-(\lambda-\phi(i-1))\log\left\{\frac{\beta+T_{i-1}+t_i}{\beta+T_{i-1}}\right\}\right\} \tag{2.5.9}$$

Let λ,ϕ be independent random variables with density functions given by (2.3.10) and (2.3.11) respectively. For simplicity the non-informative uniform distribution versions will be used here so the posterior probability density function of (λ,ϕ) is:

$$f^*(\lambda,\phi/t_1,\ldots t_n,\beta) = c.\prod_{i=1}^{n}(\lambda-(i-1)\phi)$$

$$.\exp\left\{-\sum_{i=1}^{n}(\lambda-(i-1)\phi)\log\left\{\frac{\beta+T_{i-1}+t_i}{\beta+T_{i-1}}\right\}\right\} \tag{2.5.10}$$

where

$$c^{-1} = \sum_{j=0}^{n-1}a_{j,n-1}\frac{(n-j)!\ j!}{\left[\log\left\{\frac{\beta-T_n}{\beta}\right\}\right]^{n-j+1}\left[\sum_{i=1}^{n-1}\log\left\{\frac{\beta-T_i}{\beta}\right\}\right]^{j+1}} \tag{2.5.11}$$

and the a's are given by (2.3.14).

All the above is conditional on β. A natural approach to estimate β would be to find the unconditional joint density function of $T_1,\ldots T_n$. However, the joint density function of (T_1,T_2) will be improper because of the improper prior used for (λ,Φ). Instead, then, we shall use the posterior joint density of $T_3,\ldots.T_n$ conditional on T_1,T_2 to find the MLE of β. We proceed as follows.

Let n faults have been removed in (o,τ_n). The probability density function for T_{n+1} will be:

$$f(t_{n+1}/\lambda,\Phi,\beta) = \frac{(\lambda - n\Phi)}{(\beta+\tau_n+t_{n+1})}\exp\left\{-(\lambda-n\Phi)\log\left\{\frac{\beta+\tau_n+t_{n+1}}{\beta+\tau_n}\right\}\right\} \tag{2.5.12}$$

Multiply (2.5.12) and (2.5.10), then integrate over λ,Φ with the constraint $\lambda > n\Phi$, which is the only condition for $f(t_{n+1}/\lambda,\Phi,\beta)$ to exist. The posterior density function of T_{n+1} is:

$$f^*(t_{n+1}/t_1,\ldots t_n,\beta)$$

$$= \frac{\sum_{j=o}^{n} a_{j,n} \dfrac{(n-j+1)!\; j!}{\left[\log\left\{\frac{\beta+\tau_n+t_{n+1}}{\beta}\right\}\right]^{n-j+2}\left[\sum_{i=1}^{n}\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}}}{(\beta+\tau_n+t_{n+1})\sum_{j=o}^{n-1} a_{j,n-1} \dfrac{(n-j)!\; j!}{\left[\log\left\{\frac{\beta+\tau_n}{\beta}\right\}\right]^{n-j+1}\left[\sum_{i=1}^{n-1}\log\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}}} \tag{2.5.13}$$

Then:

$$f^*(t_3,t_4,\ldots t_n/t_1,t_2,\beta) = \prod_{i=3}^{n} f(t_1/t_{i=1},\ldots t_1)$$

$$= \frac{\sum_{j=o}^{n-1} a_{j,n-1} \frac{(n-j+1)!\ j!}{\left[\log\left\{\frac{\beta+\tau_{n-1}+t_n}{\beta}\right\}\right]^{n-j+1}\left[\sum_{i=1}^{n-1}\log\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}}}{\prod_{i=3}^{n}(\beta+\tau_{i-1}+t_i)\sum_{j=o}^{1} a_{j,1}\frac{(2-j)!\ \ j!}{\left[\log\left\{\frac{\beta+\tau_2}{\beta}\right\}\right]^{3-j}\left[\log\left\{\frac{\beta+\tau_1}{\beta}\right\}\right]^{j+1}}} \tag{2.5.14}$$

By maximising (2.5.14) with respect to β we get the MLE, $\hat{\beta}$, of β which will be used in the following analysis. Now, consider the reliability function $R_{n+1}(t/\lambda,\phi,\beta)$. If the pdf of T_{n+1} exists (that is if $\lambda>n\phi$):

$$R_{n+1}(t/\lambda,\phi,\beta) = \left\{\frac{\beta+\tau_n}{\beta+\tau_n+t}\right\}^{\lambda-n\phi}$$

but

$$R_{n+1}(t/\lambda,\phi,\beta) = 1 \quad \text{if } n\phi>\lambda>(n-1)\phi,$$

remembering that T_{n+1} is not observed if $\lambda<n\phi$, i.e. that the program is now perfect and the last fault has been removed in the n^{th} fix.

We can write $R_{n+1}(t/\lambda,\phi,\beta)$ as follows:

$$R_{n+1}(t/\lambda,\phi,\beta) = \begin{cases}\exp\left\{-(\lambda-n\phi)\log\left\{\frac{\beta+\tau_n+t}{\beta+\tau_n}\right\}\right\} & \lambda>n\phi \\ 1 & (n-1)\phi<\lambda<n\phi\end{cases} \tag{2.5.15}$$

Then from (2.5.10) and (2.5.15), the current posterior reliability function will be:

$$R_{n+1}(t/t_1,\ldots.t_n,\beta) = c\left\{\sum_{j=o}^{n} a_{j,n}\frac{(n-j)!\ j!}{\left[\log\left\{\frac{\beta+\tau_n+t}{\beta}\right\}\right]^{n-j+1}\left[\sum_{i=1}^{n}\log\left\{\frac{\beta+\tau_n}{\beta}\right\}\right]^{j+1}}\right.$$

$$+\sum_{j=o}^{n-1} a_{j,n-1}\frac{(n-j)!\ j!}{\left[\log\left\{\frac{\beta+\tau_n}{\beta}\right\}\right]^{n-j+1}\left[\sum_{i=1}^{n-1}\log\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}}$$

$$\left.-\sum_{j=o}^{n} a_{j,n-1}\frac{(n-j)!\ j!}{\left[\log\left\{\frac{\beta+\tau_n}{\beta}\right\}\right]^{n-j+1}\left[\sum_{i=1}^{n}\log\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}}\right\} \qquad (2.5.16)$$

where c^{-1} is given by (2.5.11).

Thus we can obtain, for example, the probability that the program is currently perfect, P_o, which is:

$$\begin{aligned} P_o &= R_{n+1}(\infty/t_1,\ldots.t_n) \\ &= 1-c\sum_{j=o}^{n} a_{j,n}\frac{(n-j)!\ j!}{\left[\log\left\{\frac{\beta-\tau_n}{\beta}\right\}\right]^{n-j+1}\left[\sum_{i=1}^{n}\log\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}} \end{aligned} \qquad (2.5.17)$$

The predictive cdf of T_{n+1} is given by:

$$\tilde{F}(t_{n+1}/t_1,\ldots.t_n) = 1 - R_{n+1}(t/t_1,\ldots t_n)$$

Finally, to find the ROCOF, we can get first the posterior density function of $\Lambda = \frac{\lambda-n\Phi}{\beta+\tau_n}$, the rate function after n faults have been removed. The posterior density function will be obtained by transformation from (λ,Φ) to (Λ,Φ) and integration over Φ:

$$f(\Lambda/\Lambda>o,t,\ldots t_n) =$$

$$\frac{\sum_{j=o}^{n} a_{j,n} \dfrac{(\beta+\tau_n) \quad j!\ \Lambda^{n-j+1}}{\left[\sum_{i=1}^{n} \log\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}} \exp\left\{-\Lambda(\beta+\tau_n)\right\} \log\left\{\frac{\beta+\tau_n}{\beta}\right\}}{\sum_{j=o}^{n} a_{j,n} \dfrac{(n-j)!\ j!}{\left[\log\left\{\frac{\beta+\tau_n}{\beta}\right\}\right]^{n-j+1} \left[\sum_{i=1}^{n} \log\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}}} \qquad (2.5.18)$$

Then the ROCOF is given by the posterior mean of Λ. Alternatively, it could be obtained from $R_{n+1}(t/t_1,\ldots t_n)$:

$$\text{ROCOF} = -\frac{d}{dt} R_{n+1};(t/t_1,\ldots t_n)\Big|_{t=o}$$

Then the ROCOF is given by:

$$\text{ROCOF} = \frac{c}{\beta+\tau_n} \sum_{j=o}^{n} a_{j,n} \frac{(n-j+1)!\ j!}{\left[\log\left\{\frac{\beta+\tau_n}{\beta}\right\}\right]^{n-j+2} \left[\sum_{i=1}^{n} \log\left\{\frac{\beta+\tau_i}{\beta}\right\}\right]^{j+1}} \qquad (2.5.19)$$

where c^{-1} is given by (2.5.11).

The MTBF does not exist for this model (intuitively this is obvious: there is always a finite probability P_O that the program contains no more faults). We shall resort to use of IMTBF in later sections.

2.6. Littlewood-Verrall Model (LV)

The full details of this early model are given by Littlewood and Verrall (1973). They assume that $T_1, T_2 \ldots$ are the successive inter-failure times and these T's are independent and conditionally exponentially distributed:

$$f(t_i/\lambda_i) = \lambda_i \exp\{-\lambda_i t_i\} \qquad (2.6.1)$$

It is assumed that Λ_i are independent random variables with gamma $(\alpha, \psi(i))$ distribution:

$$f(\lambda_i/\alpha, \psi(i)) = \frac{[\psi(i)]^{\alpha}}{\Gamma\alpha} \lambda_i^{\alpha-1} \exp\left\{-\psi(i)\lambda_i\right\} \qquad (2.6.2)$$

From (2.6.1),(2.6.2), and integrating over λ_i, the pdf of T_i is:

$$f(t_i/\alpha_i \psi(i)) = \frac{\alpha[\psi(i)]^{\alpha}}{[\psi(i)+t_i]^{\alpha+1}} \qquad (2.6.3)$$

That is, $T_1, T_2, \ldots$ are independent Pareto distributed random variables with the same α, but different $\psi(i)$'s.

The function $\psi(i)$ determines the form of reliability growth in the model. A parametric family is chosen by the user, and here we shall use:

$$\psi(i, \beta) = \beta_1 + i\beta_2 \qquad (2.6.4)$$

However, different functions have been used [Keiller, Littlewood, Miller and Sofer, 1983a]. It is easy to see that if $\Psi(i)$ is monotonically increasing, $\{\Lambda_i\}$ are stochastically decreasing and $\{T_i\}$ are stochastically increasing (reliability growth). Conversely, if $\Psi(i)$ is decreasing, $\{\Lambda_i\}$ are stochastically increasing and $\{T_i\}$ are stochastically decreasing (reliability decay).

Now, if $t_1, \ldots t_n$ are the inter-failure times, terminating in a failure, each has a density given by (2.6.3) and so the likelihood function is:

$$L(t_1, \ldots t_n/\alpha, \Psi(i)) = \alpha^n \prod_{i=1}^{n} \frac{[\Psi(i)]^\alpha}{[\Psi(i)+t_i]^{\alpha+1}} \qquad (2.6.5)$$

with

$$\ell(t_1, \ldots t_n/\alpha, \Psi(i)) = \log[L(t_1, \ldots t_n/\alpha, \Psi(i))]$$

$$= n \log \alpha + \alpha \sum_{i=1}^{n} \log[\Psi(i)]$$

$$-(\alpha+1) \sum_{i=1}^{n} \log [\Psi(i)+t_i] \qquad (2.6.6)$$

The MLE of α is:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \left\{ \log[\hat{\Psi}(i)+t_i] - \log[\hat{\Psi}(i)] \right\}} \qquad (2.6.7)$$

The parameters of $\Psi(i,\beta)$ are found by maximising (2.6.6) numerically with respect to β_1, β_2 and these estimates of the parameters will be used to predict the pdf and cdf of T_{n+1} by substituting in (2.6.3):

$$\hat{f}(t_{n+1}/\hat{\alpha},\hat{\psi}(n+1)) = \frac{\hat{\alpha}[\hat{\psi}(n+1)]^{\hat{\alpha}}}{[\hat{\psi}(n+1)+t_{n+1}]^{\hat{\alpha}+1}} \tag{2.6.8}$$

$$\hat{F}(t_{n+1}/\hat{\alpha},\hat{\psi}(n+1)) = 1 - \left[\frac{\hat{\psi}(n+1)}{\hat{\psi}(n+1)+t_{n+1}}\right]^{\hat{\alpha}} \tag{2.6.9}$$

Then from (2.6.8) and (2.6.9), the rate function for T_{n+1} is:

$$\hat{\lambda}(t_{n+1}/\hat{\alpha},\hat{\psi}(n+1)) = \frac{\hat{\alpha}}{\hat{\psi}(n+1)+t_{n+1}} \tag{2.6.10}$$

By letting $t_{n+1} = 0$, we get the current ROCOF, i.e.

$$\text{ROCOF} = \frac{\hat{\alpha}}{\hat{\psi}(n+1)} \tag{2.6.11}$$

The mean time to failure (MTTF) (if it exists) and the median are:

$$\text{MTTF} = \frac{\hat{\psi}(n+1)}{\hat{\alpha} - 1} \qquad \text{iff } \alpha > 1 \tag{2.6.12}$$

and

$$\text{Median} = \hat{\psi}(n+1)[(.5)^{1/\hat{\alpha}} - 1] \tag{2.6.13}$$

2.7. Keiller-Littlewood Model (KL)

This model is described in Keiller, Littlewood, Miller and Sofer (1983a). It is similar to the above model, except that the reliability growth is induced via the shape parameter of the gamma distribution for the rate instead of the scale parameter as in LV. Thus the pdf of Λ_i is rewritten as:

$$f(\lambda_i/\psi(i),\beta) = \frac{\beta^{\psi(i+1)}}{\Gamma\psi(i)} \lambda_i^{\psi(i)-1} \exp(-\beta\lambda_i) \tag{2.7.1}$$

which implies that the T_i's are independently Pareto distributed.

$$f(t_i/\Psi(i),\beta) = \frac{\Psi(i)\beta^{\Psi(i)}}{[\beta+t_i]^{\Psi(i)+1}} \tag{2.7.2}$$

$\Psi(i)$ is a parametric family chosen by the user.

Here we shall use:

$$\Psi(i,\alpha) = \frac{1}{\alpha_1 + i\alpha_2} \tag{2.7.3}$$

For this model it is easy to see that monotonically decreasing $\Psi(i)$ implies stochastically decreasing $\langle\Lambda_i\rangle$, hence stochastically increasing $\langle T_i\rangle$, i.e. reliability growth. Conversely increasing $\Psi(i)$ implies reliability decay.

In order to find the MLE of the parameters, let $t_1,\dots.t_n$ be the recorded inter-failure times for the first n failures, then the likelihood function will be:

$$L(t_1,\dots.t_n/\beta,\Psi(i)) = \prod_{i=1}^{n} \frac{\Psi(i)\ \beta^{\Psi(i)}}{[\beta + t_i]^{\Psi(i)+1}} \tag{2.7.4}$$

so

$$\begin{aligned} \ell(t_1,\dots.t_n/\beta,\Psi(i)) &= \log[L(t_1,\dots.t_n/\beta,\Psi(i)] \\ &= \sum_{i=1}^{n} \log(\Psi(i)) + [\log(\beta)]\sum_{i=1}^{n}\Psi(i) \\ &\quad - \sum_{i=1}^{n} (\Psi(i) + 1)\log(\beta+t_i) \end{aligned} \tag{2.7.5}$$

We can obtain the MLE $\hat{\beta}$, $\hat{\Psi}(i)$ of β, $\Psi(i)$ by numerically maximization. Substituting $\hat{\beta}$, $\hat{\Psi}(i)$ into (2.7.3) we obtain the predictive pdf of T_{n+1}:

$$f(t_{n+1}/\beta,\Psi(n+1)) = \frac{\hat{\Psi}(n+1)\hat{\beta}^{\hat{\Psi}(n+1)}}{[\hat{\beta}+t_{n+1}]^{\hat{\Psi}(n+1)+1}} \tag{2.7.6}$$

and the predictive cdf:

$$\begin{aligned}\hat{F}(t/\hat{\beta},\hat{\Psi}(n+1)) &= \hat{P}_r(t_{n+1} \leqslant t)\\ &= 1 - \left[\frac{\hat{\beta}}{\beta + t}\right]^{\hat{\Psi}(n+1)}\end{aligned} \tag{2.7.7}$$

From (2.7.6) and (2.7.7) the rate function for the next failure time distribution will be:

$$\lambda(t) = \frac{\hat{\Psi}(n+1)}{[\hat{\beta} + t]} \tag{2.7.8}$$

Letting t=0 in (2.7.8), we obtain the ROCOF.

The MTTF (if it exists) and Median are given by:

$$MTTF = \frac{\hat{\beta}}{\hat{\Psi}(n+1)-1} \qquad \text{iff } \hat{\Psi}(n+1) > 1$$

$$Median = \hat{\beta}\,[(.5)^{1/\hat{\Psi}(n+1)} - 1]$$

2.8. Duane Model (D)

This model has been discussed widely in the literature: see, for example, Duane (1964), Finkelstein (1976), Braun and Paine (1977), Crow (1977), Crow (1979), Bruan and Schenker (1980), Littlewood (1984), and Ascher and Feingold (1984).

The model originally comes from an observation of Duane, that for hardware systems, the cumulative failure number approximates a power law in time. Crow (1977) added the non-homogeneous Poisson process (NHPP) assumption, taking a rate function of form $\gamma\beta t^{\beta-1}$.

Several authors have misleadingly referred to this NHPP as the Weibull process because of the similarity of the rate function to the hazard rate of the Weibull distribution. This confusion between the rate of a stochastic process and the rate of a life-time distribution is discussed extensively in Ascher and Feingold (1984).

The intensity function $\gamma\beta t^{\beta-1}$ has three possibilities, it is decreasing when $\beta < 1$, increasing when $\beta > 1$ and constant when $\beta = 1$, corresponding to reliability growth, reliability decay, and constant reliability (simple Poisson process), respectively.

It is easy to show that the pdf of T_i will be given by:

$$f(t_i/\gamma,\beta,\tau_{i-1}) = \gamma\,\beta(\tau_{i-1}+t_i)^{\beta-1}\exp(-\gamma[(\tau_{i-1}+t_i)^{\beta}-\tau_{i-1}^{\beta}]) \tag{2.8.1}$$

Now, if $t_1,\ldots t_n$ are the observed inter-event times for the n faults detected, the likelihood function is:

$$L(t_1,\ldots.t_n/\gamma,\beta) = \gamma^n\beta^n[\prod_{i=1}^{n}(\tau_{i-1}+t_i)^{\beta-1}]\exp(-\gamma(\tau_{n-1}+t_n)^{\beta}) \tag{2.8.2}$$

where τ_{i-1} is the total elapsed time until the fault (i-1) has been fixed, i.e. $\tau_{i-1} = \sum_{j=1}^{i-1} t_j$. Taking logs:

$$\ell(t_1,\ldots t_n/\gamma,\beta) = n\log\gamma + n\log\beta + (\beta-1)\sum_{i=1}^{n}\log(\tau_{i-1}+t_i) - \gamma(\tau_{n-1}+t_n)^{\beta} \tag{2.8.3}$$

and the likelihood equations are:

$$\frac{\partial}{\partial\gamma}\ell(t_1,\ldots t_n/\gamma,\beta) \equiv \frac{n}{\gamma}-(\tau_{n-1}+t_n)^{\beta} = 0 \tag{2.8.4}$$

$$\frac{\partial}{\partial\beta}\ell(t_1,\ldots.t_n/\gamma,\beta) \equiv \frac{n}{\beta} + \sum^{n}\log(\tau_{i-1}+t_i) - \gamma(\tau_{n-1}-t_n)^{\beta}\log(\tau_{n-1}+t_n) = 0 \tag{2.8.5}$$

It follows that the MLE's of γ,β are:

$$\hat{\gamma} = \frac{n}{(\tau_{n-1}+t_n)^{\hat{\beta}}} \tag{2.8.6}$$

and

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n-1} \log\left(\frac{\tau_{n-1}+t_n}{\tau_{i-1}+t_i}\right)} \tag{2.8.7}$$

Various predictions about future behaviour can now be made by substitution of $\hat{\gamma},\hat{\beta}$ into appropriate expressions. For example, the predictive density for T_{n+1} is:

$$\hat{f}(t_{n+1}/\hat{\gamma},\hat{\beta}) = \hat{\gamma}\,\hat{\beta}(\tau_n+t_{n+1})^{\hat{\beta}-1}\exp\{-\hat{\gamma}[(\tau_n+t_{n+1})^{\hat{\beta}} - \tau_n^{\hat{\beta}}]\} \tag{2.8.8}$$

with cdf

$$\hat{F}(t_{n+1}/\hat{\gamma},\hat{\beta}) = 1 - \exp\{-\hat{\gamma}[(\tau_n+t_{n+1})^{\hat{\beta}} - \tau_n^{\hat{\beta}}]\} \tag{2.8.9}$$

From (2.8.8) and (2.8.9), the rate function $\lambda(t_{n+1})$ for T_{n+1} is:

$$\lambda(t_{n+1}) = \hat{\gamma},\hat{\beta}\,(\tau_n+t_{n+1})^{\hat{\beta}-1} \tag{2.8.10}$$

and when $t_{n+1} \to 0$, we obtain the estimated curent ROCOF. Because the MTTF cannot be easily obtained in a simple form, we use instead the IMTBF (see section (2.11)), defined as the reciprocal of the ROCOF, Braun and Paine (1977). The median for T_{n+1} is:

$$\text{Median} = \tau_n \left[\left(1 - \frac{\log(.5)}{\hat{\gamma}\, \tau_n^{\hat{\beta}}}\right)^{1/\hat{\beta}} - 1 \right] \tag{2.8.11}$$

It should be pointed out that the rate function of this NHPP has two minor disadvantages as a model of reliability growth: it is infinite at $t = 0$ and zero at $t = \infty$. Clearly real systems will have neither of these properties: Littlewood (1984). However, it may be a good model for real behaviour for finite times.

2.9. Non-Homogeneous Poisson Process for JM Model (JMNHPP)

Goel (1980) proposed a NHPP variant of the JM model, arguing informally that the rate of detection (and removal) of faults should be proportional to the number present. This reasoning suggests a rate function of the form $\mu(1 - e^{-\Phi t})$, but does not justify the NHPP assumption. A more satisfactory approach is to assume that N, the initial number of faults in the JM model, is a Poisson (μ) random variable. It then follows that the process is a NHPP with rate:

$$\mu(1 - e^{-\Phi t}) \tag{2.9.1}$$

The distribution of the i^{th} inter-event time, T_i, for this process, given that the $(i-1)^{th}$ event occurs at epoch, τ_{i-1}, is:

$$f(t_i/\Phi,\mu) = \mu\,\Phi\, e^{-\Phi(\tau_{i-1}+t_i)}\exp\{-\mu e^{-\Phi\tau_{i-1}}(1 - e^{-\Phi t_i})\} \quad (2.9.2)$$

Thus, if $t_1,\ldots t_n$ are the observed inter-failure times for the 1st, 2nd,...and n^{th} fault, the likelihood function is:

$$L(t_1,\ldots t_n/\mu,\Phi) = \mu^n\Phi^n \exp\{-\Phi \sum_{i=1}^{n}(\tau_{i-1}+t_i)$$

$$-\mu(1 - \exp[-\Phi(\tau_{n-1}+t_n)])\} \quad (2.9.3)$$

so

$$\ell(t_1,\ldots .t_n/\mu,\Phi) = n\log\mu + n\log\Phi = \Phi\sum_{i=1}^{n}(\tau_{i-1}+t_i)$$

$$-\mu[1 - \exp(-\Phi(\tau_{n-1}+t_n])] \quad (2.9.4)$$

Then by maximizing (2.9.4) numerically with respect to μ,Φ we get the MLE's $\hat{\mu}$, $\hat{\Phi}$, which can be substituted appropriately to obtain various predictions of future behaviour. The predictive pdf of the next inter-failure time, T_{n+1}, is:

$$\hat{f}(t_{n-1}/\hat{\mu},\hat{\Phi}) = \hat{\mu}\,\hat{\Phi}\,\exp\{-\hat{\Phi}(\tau_n+t_{n+1})$$

$$- \hat{\mu} e^{-\hat{\Phi}\tau_n} (1 - e^{-\hat{\Phi}t_{n+1}})\} \quad (2.9.5)$$

with cdf:

$$\hat{F}(t_{n+1}/\hat{\mu},\hat{\Phi}) = 1 - \exp\{-\hat{\mu}\exp(-\hat{\Phi}\tau_n)[1 - e^{-\hat{\Phi}t_{n+1}}]\} \tag{2.9.6}$$

and the rate function (hazard function):

$$\lambda(t) = \hat{\mu}\,\hat{\Phi}\,\exp\{-\hat{\Phi}(\tau_n+t)\}$$

The current ROCOF is:

$$\hat{\mu}\,\hat{\Phi}\,\exp\{-\hat{\Phi}\tau_n)$$

and

$$\text{IMTBF} = 1/\text{ROCOF}$$

The median of T_{n+1} is:

$$\frac{1}{\hat{\Phi}}[\log\hat{\mu} - \log[\hat{\mu}e^{-\hat{\Phi}\tau_n} + \log(.5)]]-\tau_n$$

Notice that this model is Case 1 in the Langberg and Singpurwalla (1981) approach to find the Bayesian JM model, section (2.3.1). Here the Poisson distribution for N is interpreted as a Bayesian prior distribution. It is also of interest to observe that:

$$F(\infty) = 1 - e^{-\mu e^{-\Phi\tau_n}}$$

which means that there is a finite probability that the program is currently perfect, which, in turn, imples that the MTTF does not exist for this model.

2.10. Non-Homogeneous Poisson Process for L Model (LNHPP)

The L model discussed in section (23.4) is an order statistic model where N is the number of faults in the program. Miller (1984) suggested that any order statistic mixture model mixed over the Poisson N will be a non-homogeneous Poisson process (NHPP). For the L model the result will be a NHPP with rate function

$$\mu[1 - (\frac{\beta}{\beta+\tau})^{\alpha}]$$

which is the rate function of the stochastic point process defined by Littlewood (1984).

The pdf of T_i given that the $(i-1)^{th}$ failure occurs at epoch $\tau_{i-1}(\equiv \sum_{j=1}^{i-1} t_j)$ is:

$$f(t_i/\tau_{i-1},\mu,\alpha,\beta) = \frac{\alpha\mu\beta^{\alpha}}{(\beta+\tau_{i-1}+t_i)^{\alpha+1}} \exp\left\{-\mu(\frac{\beta}{\beta+\tau_{i-1}})^{\alpha}\right.$$

$$\left.\left[1 - (\frac{\beta - \tau_{i-1}}{\beta + \tau_{i-1} + t_i})^{\alpha}\right]\right\} \qquad (2.10.1)$$

The log likelihood at the epoch immediately after the n^{th} failure is:

$$\ell(t_i,\ldots t_n/\mu,\alpha,\beta) = n\log\mu + n\log\alpha + n\alpha\log\beta$$

$$-(\alpha+1)\sum_{i=1}^{n}\log(\beta+\tau_{i-1}+t_i) - \mu\left[1-\left(\frac{\beta}{\beta+\tau_{n-1}+t_n}\right)^{\alpha}\right] \qquad (2.10.2)$$

Maximizing (2.10.2) numerically will give the MLE's of α,μ and β. Hence, by substituting in (2.10.1) we get the usual predictors of future behaviour. Thus the pdf of T_{n+1} is:

$$\hat{f}(t_{n+1}/\hat{\mu},\hat{\alpha},\hat{\beta},\tau_n) - \frac{\hat{\alpha}\,\hat{\mu}\,\hat{\beta}^{\alpha}}{(\hat{\beta}+\tau_n+t_{n-1})^{\hat{\alpha}+1}}\exp\left\{-\hat{\mu}\left(\frac{\hat{\beta}}{\hat{\beta}+\tau_n}\right)^{\hat{\alpha}}\left[1-\left(\frac{\beta+\tau_n}{\beta+\tau_n+t_{n+1}}\right)^{\hat{\alpha}}\right]\right\} \qquad (2.10.3)$$

with cdf

$$\hat{F}(t_{n+1}/\hat{\mu},\hat{\alpha},\hat{\beta},\tau_n) = 1-\exp\left\{-\hat{\mu}\left(\frac{\hat{\beta}}{\hat{\beta}+\tau_n}\right)^{\hat{\alpha}}\left[1-\left(\frac{\hat{\beta}+\tau_n}{\hat{\beta}+\tau_n+t_{n+1}}\right)^{\hat{\alpha}}\right]\right\} \qquad (2.10.4)$$

and rate function:

$$\lambda(t) = \frac{\hat{\mu}\,\hat{\alpha}\,\hat{\beta}^{\hat{\alpha}}}{(\hat{\beta}+\tau_n+t)^{\hat{\alpha}+1}} \qquad (2.10.5)$$

The current ROCOF is obtained by putting $t = 0$ in this:

$$\text{ROCOF} = \frac{\hat{\mu}\,\hat{\alpha}\,\hat{\beta}^{\hat{\alpha}}}{(\hat{\beta}+\tau_n)^{\hat{\alpha}+1}}$$

and

$$\text{IMTBF} = 1/\text{ROCOF}$$

Finally, it is easy to see that the model considered here is Case 1 in section (2.5.1), where the mixture is interpreted as a Bayesian prior distribution for N.

Again:

$$F(\infty) = 1 - \exp\left\{-\hat{\mu}\left(\frac{\hat{\beta}}{\beta + \tau_n}\right)^{\hat{\alpha}}\right\}$$

which means that there is a finite probability that the program is currently perfect, implying that the MTTF does not exist for this model.

2.11. The Weibull Model (W)

Assume, as before, that the program starts life containing N faults. When a fault shows itself by causing a system failure, this fault will be removed and the system will immediately return to its operating environment. Let Xi be the total time on test until the fault i is removed. The X_i's are assumed to be i.i.d. random variables with common Weibull density function, i.e.:

$$f(x) = \alpha \beta x^{\beta-1} \exp(-\alpha x^{\beta}) \qquad (2.11.1)$$

Then $X_{(1)}$, $X_{(2)}$... are the order statistics of the time of occurrence of the 1st, 2nd,....failures and T_1,T_2,...are the inter-failure times so:

$$T_1 = X_{(1)}$$

$$T_1 = X_{(1)} - X_{(i-1)} \tag{2.11.2}$$

It is easy to show that the random variables $T_i(i=1,2,...n)$ are conditionally Weibull distributed with the pdf of T_i given by:

$$f(t_i/\tau_{i-1},\alpha,\beta,N) = (N-i+1)\alpha\beta(\tau_{i-1}+t_i)^{\beta-1}$$

$$.\exp\{-(N-i+1)\alpha[(\tau_{i-1}-t_i)^{\beta} - \tau_{i-1}^{\beta}] \tag{2.11.3}$$

where τ_{i-1} is the total elapsed time at the point (i-1), i.e.:

$$t_{i-1} \equiv \sum_{j=1}^{i-1} t_j$$

The rate function (hazard rate) associated with r.v. T_i is:

$$\lambda(t_i) = (N-i+1)\alpha\beta(\tau_{i-1}+t_i)^{\beta-1} \tag{2.11.4}$$

and the current ROCOF:

$$(N-i+1)\alpha\beta\ \tau_{i-1}^{\beta-1} \tag{2.11.5}$$

It is clear that for $\beta = 1$ the model becomes the JM model. When $\beta < 1$ the individual times to detection, X_i, have a distribution with decreasing rate function. When $\beta > 1$ these rates are increasing, so the reliability growth of the programs happends solely through the fault removal term (N-i+1). The model reduces to the Duane model as $N\to\infty$ and $\alpha\to 0$ with $N\alpha$ fixed at, say, γ.

To obtain predictions for the Weibull model we substitute MLE estimates of the parameters into suitable expressions. The likelihood function is:

$$L(t_1,\dots t_n/\alpha,\beta,N) = \prod_{i=1}^{n} f(t_i/\tau_{i-1},\alpha,\beta,N)$$

$$= \alpha^n \beta^n \prod_{i=1}^{n} (N-i+1)(\tau_{i-1}+t_i)^{\beta-1}$$

$$\exp\{-\alpha \sum_{i=1}^{n} (\tau_{i-1}+t_i)^{\beta} - (N-n)\alpha(\tau_{n-1}+t_n)^{\beta}\} \qquad (2.11.6)$$

So taking logs:

$$\ell(t_1,\dots t_n/\alpha,\beta,N) = n\log\alpha + n\log\beta + \sum_{i=1}^{n} \log(N-i+1)$$

$$+ (\beta-1)\sum_{i=1}^{n} \log(\tau_{i-1}+t_i)\alpha \sum_{i=1}^{n} (\tau_{i-1}+t_i)^{\beta}$$

$$-(N-n)\alpha(\tau_{n-1}+t_n)^{\beta} \qquad (2.11.7)$$

Hence the likelihood equations are:

$$\frac{\partial}{\partial\alpha}\ell(t_1,\dots t_n/\alpha,N,\beta) = \frac{n}{\alpha} - \sum_{i=1}^{n} \tau_i^{\beta} - (N-n)\tau_n^{\beta} = 0$$

$$\frac{\partial}{\partial\beta}\ell(t_1,\dots t_n/\alpha,N,\beta) = \frac{n}{\beta} + \sum_{i=1}^{n} \log\tau_i - \alpha \sum_{i=1}^{n} \tau_i^{\beta}\log\tau_i$$

$$-(N-n)\alpha\,\tau_n^{\beta}\log\tau_n = 0$$

and

$$\frac{\partial}{\partial N}\ell(t_1,\dots t_n/\alpha,N,\beta) = \sum_{i=1}^{n} \frac{1}{(N-i+1)} - \alpha\,\tau_n^{\beta} = 0$$

The first equation can be solved in closed form:

$$\hat{\alpha} = \frac{n}{(\hat{N}-n)\tau_n^{\beta} + \sum_{i=1}^{n} \tau_i^{\beta}}$$

This can be substituted into the previous equations which can be solved by a version of Newton's method. Alternatively, it can be substituted into (2.11.7) which can be maximised in (β,N) space.

The predictive pdf of T_{n+1} is:

$$\hat{f}(t_{n+1}/\tau_n, \hat{\alpha}, \hat{\beta}, \hat{N}) = (\hat{N}-n)\hat{\alpha}\ \hat{\beta}(\tau_n+t_{n+1})^{\hat{\beta}-1}$$

$$.\exp\{-(\hat{N}-n)\hat{\alpha}[(\tau_n-t_{n+1})^{\hat{\beta}} - \tau_n^{\hat{\beta}}]\} \qquad (2.11.8)$$

the predictive cdf is:

$$\hat{F}(t_{n+1}/\tau_n, \hat{\alpha}, \hat{\beta}, \hat{N}) = 1 - \exp\{-(\hat{N}-n)\hat{\alpha}[(\tau_n+t_{n+1})^{\hat{\beta}} - \tau_n^{\hat{\beta}}]\} \qquad (2.11.9)$$

The rate function (hazard rate) for T_{n+1} is, from (2.11.8) and (2.11.9):

$$\lambda(t_{n+1}) = \frac{f(t_{n+1})}{1-F(t_{n+1})} = (\hat{N}-n)\hat{\alpha}\ \hat{\beta}(\tau_n+t_{n+1})^{\hat{\beta}-1}$$

and by letting $t_{n+1} \to 0$, we obtain the current ROCOF:

$$\text{ROCOF} = (\hat{N}-n))\hat{\alpha}\ \hat{\beta}\ \tau_n^{\hat{\beta}-1}$$

The current MTTF is:

$$E(T_{n+1}) = \int_0^\infty t_{n+1}\ \hat{f}(t_{n+1}/\tau_n, \hat{\alpha}, \hat{\beta}, \hat{N})\ d\,t_{n+1}$$

$$= \tau_n \left\{ \int_0^\infty \left[1 + \frac{Z}{(\hat{N}-n)\hat{\alpha}\tau_n^{\hat{\beta}}}\right]^{1/\hat{\beta}} e^{-Z}\ dZ - 1 \right\}$$

$$= \frac{1}{(\hat{N}-n)\hat{\alpha}\tau_n^{\hat{\beta}-1}} + \tau_n \sum_{i=1}^{[1/\hat{\beta}]} \frac{(1/\hat{\beta})!}{(1/\hat{\beta}-i)!} \frac{1}{[(\hat{N}-n)\hat{\alpha}\tau_n^{\hat{\beta}}]^i} \qquad (2.11.10)$$

We note that the first term is the reciprocal of the ROCOF, which is the IMTBF. For simplicity this term will be used instead of the MTTF in much of the later data analysis.

The median is given by:

$$\text{Median} = \tau_n \left[\left\{ 1 - \frac{\log(.5)}{(\hat{N}-n)\hat{\alpha}\ \tau_n^{\beta}} \right\}^{1/\hat{\beta}} - 1 \right] \qquad (2.11.11)$$

CHAPTER III

DISCRETE SOFTWARE RELIABILITY MODELS

3.1. Introduction

The models considered in the previous chapter depend on the availability of data in the form of successive times between failure. Such extensive data may not be available in practice, so it is important to be able to analyse data in the form of a counting process. Such data will be in the form of pairs $(n_i(t_i),t_i)$ where $n_i(t_i)$ represents the number of failures during the time interval t_i.

The main objective of this chapter is to present some models, related to the others mentioned in the previous chapter, which treat this kind of discrete data.

There has been little work in this area over the past five years. Goel (1980) has proposed a model based on the NHPP; Brooks and Motley (1980) have considered variants of the Jelinski-Moranda using Binomial and Poisson distributions; Misra (1983) has used the Goel model to analyse software failure data for different projects.

In section (3.2) the discrete JM model will be considered. The discrete Littlewood model will be considered in section (3.3) and discrete LV and KL models will be discussed in sections (3.4) and (3.5) respectively.

Sections (3.6), (3.7) and (3.8) will treat the non-homogeneous Poisson Process (NHPP) models, and finally, the discrete Weibull model will be presented in section (3.9).

3.2. Discrete Jelinski-Moranda Model (DJM)

Let us assume that the program will start its life containing N faults, each of which has a time to occurrence which is exponentially distributed with rate Φ. Let n_1, n_2,... be the numbers of failures observed during the successive test periods of lengths t_1, t_2,... respectively. We assume that perfect fixing takes place at the end of each interval, which means that the program failure rate will improve by the amount $n_i\Phi$. So during period i the system failure rate will be $(N-C_{i-1})\Phi$, where $C_{i-1} = \sum_{j=1}^{i-1} n_i$ is the total number of faults fixed since debugging began.

Consider the random variable $N_i(t_i)$, representing the number of failures in the i^{th} time interval of length t_i. It is easy to show that $N_i(t_i)$ has a Poisson distribution with mean value function $M(\tau_{i-1},\tau_{i-1}+t_i) = (N-C_{i-1})\Phi t_i$. That is:

$$P(n_i/N,\Phi) = \frac{[(N-C_{i-1})\Phi t_i]^{n_i}}{n_i!} \exp(-(N-C_{i-1})\Phi t_i) \qquad (3.2.1)$$

The two parameters N and Φ will be estimated by the MLE method. Let $(n_1,t_1),(n_2,t_2)...(n_k,t_k)$ be the observations for the first k test periods. The likelihood function of the observations is:

$$L(n_1,...n_k/N,\Phi) = \prod_{i=1}^{k} \frac{[(N-C_{i-1})\Phi t_i]^{n_i}}{n_i!}$$

$$.\exp(-\sum_{i=1}^{k}(N-C_{i-1})\Phi t_i) \tag{3.2.2}$$

and by taking the natural logarithm we get:

$$\ell(n_1,...n_k/N,\Phi) = (\sum_{i=1}^{k} n_i)\log\Phi + \sum_{i=1}^{k} n_i \log t_i$$

$$+\sum_{i=1}^{k} n_i \log(N-C_{i-1}) - \sum_{i=1}^{k}(N-C_{i-1})\Phi t_i$$

$$-\sum_{i=1}^{k} \sum_{j=1}^{n_i} \log j \tag{3.2.3}$$

The condition given by Littlewood and Verrall (1981) for the likelihood function, in the continuous case, to have a unique maximum of finite N and non-zero Φ has an equivalent form here. The function (3.2.2) or (3.2.3) will have a unique maximum at finite N and non-zero Φ if and only if:

$$\frac{\sum_{i=1}^{k} C_{i-1} t_i}{\sum_{i=1}^{k} n_i C_{i-1}} > \frac{\sum_{i=1}^{k} t_i}{\sum_{i=1}^{k} n_i} \tag{3.2.4}$$

otherwise the likelihood function has its maximum at finite λ = NΦ for infinite N (see Appendix A).

The condition (3.2.4) should be tested for each data set, and if satisfied, the MLEs can be obtained by maximizing (3.2.3) numerically. Then the MLE $\hat{\Phi}$, $\hat{N}$ of Φ,N will be used to predict the expected number of failures (the mean function) in the next test period, which will be given by:

$$M(\tau_k, \tau_k + t_{k+1}) = (\hat{N} - C_k)\hat{\Phi} t_{k+1} \qquad (3.2.5)$$

Also, the predictive probability function for the number of failures during this period could be obtained in the following form:

$$\hat{P}(n_{k+1}/(\hat{N},\hat{\Phi}) = \frac{[(\hat{N}-C_k)\hat{\Phi} t_{k+1}]^{n_{k+1}}}{(n_{k+1})!} \cdot \exp\{-(\hat{N}-C_k)\hat{\Phi} t_{k+1}\} \qquad (3.2.6)$$

and the predictive distribution function (cdf) could be obtained by summing for n_{k+1} from zero to some specified number, say, n, i.e.:

$$\hat{F}(n) = \Pr\{n_{k+1} \leqslant n\}$$

$$= \sum_{n_{k+1}=o}^{n} \hat{P}(n_{k+1}/\hat{N},\hat{\Phi}) \qquad (3.2.7)$$

The ROCOF is:

$$\lambda(t_{k+1}) = (\hat{N} - C_k)\hat{\Phi} \qquad (3.2.8)$$

It can be shown that the reliability function, which is the probability that no failure will occur during the next t time units, is:

$$\hat{R}\ (t) = \exp\{-(\hat{N} - C_k)\hat{\Phi} t\} \qquad (3.2.9)$$

This form is similar to the reliability function obtained for JM model in section (2.2). Other reliability growth measures such as MTTF and median can be used here with little change, i.e.

$$MTTF = 1/(\hat{N} - C_k)\hat{\Phi} \tag{3.2.10}$$

$$Median = -\log(.5)/(\hat{N} - C_k)\hat{\Phi} \tag{3.2.11}$$

3.3. Discrete Littlewood Model (DL)

The model described in (3.2) assumed that each fault contributed the same amount to the overall failure rate. This assumption has been challenged and an alternative one was proposed by Littlewood (1981) (see Section (2.4)). In this model each fault has a different failure rate represented by random variables Φ_i, while the other assumptions remain unchanged. After (i-1) test intervals, the program failure rate for the i^{th} interval is:

$$\lambda_i = \sum_{j=1}^{N-C_{i-1}} \Phi_j \tag{3.3.1}$$

where C_{i-1} is the total fixes in the past intervals. The probability function for the number of failures during the test period i, again assuming that fixes are introduced only at the end of intervals, is:

$$P(n_i/\lambda_i) = \frac{[\lambda_i t_i]^{n_i}}{n_i!} \; e^{-\lambda_i t_i} \tag{3.3.2}$$

If the probability density function of the Φ's is known, the unconditional probability funtion of N_i is obtainable in principle. We assume, as before, that the Φ's are independent and identically gamma distributed, i.e.:

$$p(\Phi/\alpha,\beta) = \frac{\beta^\alpha}{\Gamma\alpha} \Phi^{\alpha-1} e^{-\beta\Phi} \tag{3.3.3}$$

Using Bayes theorem, the posterior probability density function of each Φ can be obtained in the following form:

P(Φ/this fault has not been fixed in the first (i-1) intervals)

$$\equiv \frac{\Pr(\text{no failure is caused by this fault in the first } (i-1)\text{interval}/\Phi)p(\Phi)}{\int_0^\infty \Pr(\text{no failure is caused by this fault in the first } (i-1) \text{ interval}/\Phi)p(\Phi)d\Phi}$$

It is easy to show that this is:

$$f(\Phi/\text{Data}) = \frac{(\beta+\tau_{i-1})^\alpha}{\Gamma\alpha} \Phi^{\alpha-1} \exp\{-\Phi(\beta+\tau_{i-1})\} \tag{3.3.4}$$

where τ_{i-1} is the total elapsed time of the (i-1) test intervals.

From (3.3.1), it is clear that λ_i is the sum of independent random variables each with pdf given by (3.3.4). It follows that λ_i has a gamma density, i.e.

$$f(\lambda_i/\text{Data},N,\alpha,\beta) = \frac{(\beta+\tau_{i-1})^{(N-C_{i-1})\alpha}}{\Gamma[(N-C_{i-1})\alpha]} \lambda_i^{(N-C_{i-1})\alpha-1} . \exp(-\lambda_i(\beta+\tau_{i-1})) \quad (3.3.5)$$

By mixing (3.3.2) and (3.3.5), the unconditional probability function of N_i can be obtained:

$$P(n_i/N,\alpha,\beta,\text{Data}) = \int_0^\infty P(n_i/\lambda_i)p(\lambda_i/N,\alpha,\beta,\text{Data})d\lambda_i$$

$$= \binom{M_i+n_i-1}{M_i-1} P_i^{M_i}(1-P_i)^{n_i} \quad (3.3.6)$$

where

$$P_i = \left(\frac{\beta+\tau_{i-1}}{\beta+\tau_{i-1}+t_i}\right) \text{ and } M_i = (N-C_{i-1})\alpha$$

It is clear that the last form is the negative binomal probability function, and the likelihood function for the observations will be:

$$L(n_1,\ldots n_k/N,\alpha,\beta) = \prod_{i=1}^{k} P(n_i/N,\alpha,\beta,\text{Data})$$

$$= \prod_{i=1}^{k} \binom{M_i+n_i-1}{M_i-1} P_i^{M_i}(1-P_i)^{n_i} \quad (3.3.7)$$

By taking the natural logarithm, we get:

$$\ell(n_1,\ldots n_k/N,\alpha,\beta) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\log\left(\frac{(N-C_{i-1})\alpha+j-1}{j}\right)$$

$$+\sum_{i=1}^{k}(N-C_{i-1})\alpha\log\left(\frac{\beta+\tau_{i-1}}{\beta+\tau_{i-1}+t_i}\right)$$

$$+\sum_{i=1}^{k}n_i\log\left(\frac{t_i}{\beta+\tau_{i-1}+t_i}\right) \qquad (3.3.8)$$

In order to get the MLE of the parameters N, α and β, the last form should be maximized numerically to get such estimates, $\hat{N}$, $\hat{\alpha}$ and $\hat{\beta}$. These estimates can be used to predict, for example, the reliability function which is the probability of no failure in the next t test time units. This is:

$$\hat{R}(t/\hat{\alpha},\hat{N},\hat{\beta}) = \left(\frac{\hat{\beta}+\tau_n}{\hat{\beta}+\tau_k+t}\right)^{(\hat{N}-C_k)\hat{\alpha}} \qquad (3.3.9)$$

The last form is identical to the reliability function obtained for the L model in section (2.4). It is easy to obtain the other reliability growth measures such as the rate function, ROCOF, MTTF and the median:

$$\lambda(t/\hat{\alpha},\hat{N},\hat{\beta}) = \frac{(\hat{N}-C_k)\hat{\alpha}}{\hat{\beta}+\tau_k+\tau} \qquad (3.3.10)$$

$$\text{ROCOF} = \frac{(\hat{N}-C_k)\hat{\alpha}}{\hat{\beta}+\tau_k} \qquad (3.3.11)$$

$$\text{MTTF} = \frac{\hat{\beta}+\tau_k}{(\hat{N}-C_k)\hat{\alpha}-1} \qquad (\hat{N}-C_k)\hat{\alpha} > 1 \qquad (3.3.12)$$

$$\text{Median} = (\hat{\beta}+\tau_k)[(.5)^{1/(\hat{N}-C_k)\hat{\alpha}}-1] \qquad (3.3.13)$$

The predictive probability function of the number of failures during the next test period is obtained by substituting $\hat{\alpha}$, $\hat{N}$, and $\hat{\beta}$ in (3.3.6) which gives:

$$\hat{P}(n_{k+1}/\hat{\alpha},\hat{N},\hat{\beta}) = \binom{M_{k+1}+n_{k+1}-1}{M_{k+1}-1} P_{k+1}^{M_{k+1}}(1 - P_{k+1})^{n_{k+1}} \tag{3.3.14}$$

where

$$P_{k+1} = \left(\frac{\hat{\beta} + \tau_k}{\beta+\tau_k+t_{k+1}}\right) \qquad \text{and } M_{k+1} = (\hat{N} - C_k)\hat{\alpha}$$

and the expected number of failures in this interval is given by:

$$M(\tau_k,\tau_n + t_{k+1}) = \frac{(\hat{N} - C_k)\hat{\alpha}\, t_{k+1}}{\hat{\beta} + \tau_k} \tag{3.3.15}$$

3.4. Discrete Littlewood-Verrall Model (DLV)

The probability function of N_i given by (3.3.2) depends only on λ_i, the rate of occurrence of failures. If the prior density function of λ_i is known, the unconditional probability function of N_i can be obtained. Littlewood and Verrall (1973) assumed that λ_i is a random variable with a gamma density function (see (2.6.2)), with shape parameter α and scale parameter $\Psi(i)$. We shall assume for this discrete time version of the model that $\Psi(i)$ is linear:

$$\Psi(i) = \beta_1 + (C_{i-1} + 1)\beta_2 \tag{3.4.1}$$

where C_{i-1} is the total observed number of failures up to the stage (i-1). Again it is assumed that fixes are introduced only at the ends of intervals.

It is easy to shown that N_i will have a negative binomal distribution with probability function given by:

$$P(n_i/\alpha,\psi(i)) = \int_0^\infty P(n_i/\lambda_i)P(\lambda_i/\alpha,\psi(i))d\lambda_i$$

$$= \int_0^\infty \frac{(\lambda_i t_i)^{n_i}}{n_i!} e^{-\lambda_i t_i} \frac{(\psi(i))^\alpha}{(\alpha-1)!} \lambda_i^{\alpha-1} e^{-\lambda_i \psi(i)} d\lambda_i$$

$$= \binom{\alpha+n_i-1}{\alpha-1} \left(\frac{\psi(i)}{\psi(i)+t_i}\right)^\alpha \left(\frac{t_i}{\psi(i)+t_i}\right)^{n_i} \qquad (3.4.2)$$

Let $n_1,....n_k$ be the numbers of failures in the first k test periods with lengths $t_1,... t_k$ respectively. The likelihood function of these observations is:

$$L(n_1,...n_k/\alpha,\psi(i)) = \prod_{i=1}^{k} \binom{\alpha+n_i-1}{\alpha-1} \left(\frac{\psi(i)}{\psi(i)+t_i}\right)^\alpha \left(\frac{t_i}{\psi(i)+t_i}\right)^{n_i} \qquad (3.4.3)$$

Taking natural logarithms gives:

$$\ell(n_1,....n_k/\alpha,\psi(i)) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log\left(\frac{\alpha+j-1}{j}\right)$$

$$+ \alpha \sum_{i=1}^{k} \log\left(\frac{\psi(i)}{\psi(i)+t_i}\right)$$

$$+ \sum_{i=1}^{k} n_i \log\left(\frac{t_i}{\psi(i)+t_i}\right) \qquad (3.4.4)$$

Then the MLE $\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ of the parameters α, β_1, and β_2 will be obtained by maximizing the form (3.3.4) numerically. These estimates will be used to predict, for example, the reliability function, which is the probability of no failure in the next interval of length t time units.

$$\hat{R}(t/\hat{\alpha},\hat{\psi}(k+1)) = \left[\frac{\hat{\psi}(k+1)}{\hat{\psi}(k+1)+t}\right]^{\hat{\alpha}} \qquad (3.4.5)$$

Clearly this is the same function as obtained for LV model in the continuous case in section (2.6). The rate function which is given by (2.6.11) can be written in the form:

$$\lambda(t_{i+1}) = \frac{\hat{\alpha}}{\psi(k+1)+t_{k+1}} \qquad (3.4.6)$$

with ROCOF

$$\text{ROCOF} = \hat{\alpha}/\hat{\psi}(k+1) \qquad (3.4.7)$$

$$\text{MTTF} = \frac{\hat{\psi}(k+1)}{\hat{\alpha}-1} \qquad \hat{\alpha} > 1 \qquad (3.4.8)$$

$$\text{Median} = \hat{\psi}(k+1)[(.5)^{1/\hat{\alpha}}-1] \qquad (3.4.9)$$

The cdf can be predicted by substituting $\hat{\alpha}$, $\hat{\psi}(k+1)$ in (3.4.2) and summing over N_{k+1}. More importantly, the expected number of failures in the present test interval is given by the following form:

$$M(\tau_k,\tau_k+t_{k+1}) = \frac{\hat{\alpha}\, t_{k+1}}{\hat{\psi}(k+1)} \qquad (3.4.10)$$

where, of course, t_{k+1} is assumed known.

3.5. Discrete Keiller-Littlewood Model (DKL)

This model is similar to the previous model. Here the distribution of λ_i is assumed to be gamma, with shape parameter $\psi(i)$ and scale β: see (2.7.1). Fixes are carried out only at the ends of intervals.

The probability function for N_i, the number of failures in t_i, is:

$$P(n_i/\beta,\psi(i)) = \int_0^\infty P(n_i/\lambda_i)P(\lambda_i/\beta,\psi(i))d\lambda_i$$

$$= \binom{\psi(i)+n_i-1}{\psi(i)-1} \left(\frac{\beta}{\beta+t_i}\right)^{\psi(i)} \left(\frac{t_i}{\beta+t_i}\right)^{n_i} \qquad (3.5.1)$$

Again, there is flexibility in the choice of $\psi(i)$, we shall assume a reciprocal linear form:

$$\psi(i) = \frac{1}{\alpha_1 + (C_{i-1}+1)\alpha_2} \qquad (3.5.2)$$

where $C_{i-1} = \sum_{j=1}^{i-1} n_j$

Let $n_1,....n_k$ be the observed numbers of failures during the test intervals $t_1,....t_k$ respectively. The likelihood function can be written as:

$$L(n_1,\ldots n_k/\beta,\psi(i)) = \prod_{i=1}^{k} \binom{\psi(i)+n_i-1}{\psi(i)-1} \left(\frac{\beta}{\beta+t_i}\right)^{\psi(i)} \left(\frac{t_i}{\beta+t_i}\right)^{n_i} \tag{3.5.3}$$

Taking natural logarithms we obtain:

$$\ell(n_1,\ldots n_k/\beta,\psi(i)) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log\left(\frac{\psi(i)+j+1}{j}\right)$$

$$+ \sum_{i=1}^{k} \psi(i) \log\left(\frac{\beta}{\beta+t_i}\right)$$

$$+ \sum_{i=1}^{k} n_i \log\left(\frac{\beta}{\beta+t_i}\right) \tag{3.5.4}$$

It is clearly not easy to find the MLE of α_1, α_2, and β analytically from the above form, and numerical maximization must be used. After obtaining the MLE $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\beta}$ for the parameters α_1, α_2 and β respectively, it can be shown that the reliability function, which is the probability that no failure will occur during the next period of length t is given by:

$$\hat{R}(t) = \left[\frac{\hat{\beta}}{\hat{\beta}+t}\right]^{\hat{\psi}(k+1)} \tag{3.5.5}$$

Again this form is identical to the reliability function obtained for the KL model in section (2.7). Other software reliability growth measures, such as the rate function, ROCOF, MTTF and the Median time to failure, which were given in section (2.7) can be used here without any change. It is important to note that the ψ functions in both continuous and discrete cases depend on the number of failures.

The predictive cdf can be obtained by summation of the probability mass function, and the expected number of failures during the next test phase is:

$$M(\tau_k,\tau_k+t_{k+1}) = \frac{\hat{\psi}(k+1)t_{k+1}}{\hat{\beta}} \tag{3.5.6}$$

where

$$\hat{\psi}(k+1) = 1/(\hat{\alpha}_1 + (C_{k+1})\hat{\alpha}_2)$$

3.6. Discrete Duane Model (DD)

In section (2.8), the Duane model is considered as a non-homogeneous Poisson process (NHPP) with intensity function at τ given by:

$$\lambda(\tau) = \gamma\beta\ \tau^{\beta-1} \tag{3.6.1}$$

It is well known that the mean function for a process is the integral of this, i.e:

$$M(\tau) = \gamma\ \tau^{\beta} \tag{3.6.2}$$

is the expected number of failures during the interval $(0,\tau)$.

Suppose that (i-1) test phases have been observed. The system will run for the next test phase (the i^{th} period) of length t. The expected number of failures during this period can be easily seen as the difference between $M(\tau_{i-1}+t_i)$ and $M(\tau_{i-1})$, where

$\tau_{i-1} = \sum_{j=1}^{i-1} t_j$ is the total elapsed time at the start of the interval.

That is:

$$M(\tau_{i-1},\tau_{i-1}+t_i) = \gamma\,[(\tau_{i-1} + t_i)^{\beta} - \tau_{i-1}^{\beta}] \qquad (3.6.3)$$

The probability function of the number of failures during this current test interval is Poisson distributed:

$$P(n_i) = \frac{[M(\tau_{i-1},\tau_{i-1}+t_i)]^{n_i}}{n_i!}\ \exp.(-M(\tau_{i-1},\tau_{i-1}+t_i))$$

$$= \frac{[\gamma((\tau_{i-1}+t_i)^{\beta}-\tau_{i-1}^{\beta})\ ]^{n_i}}{n_i!}\ \exp.(-\gamma[(\tau_{i-1}+t_i)^{\beta}-\tau_{i-1}^{\beta}]) \qquad (3.6.4)$$

The likelihood function is therefore:

$$L(n_1,\ldots n_k/\gamma,\beta) = \prod_{i=1}^{k} \left(\frac{[\gamma((\tau_{i-1}+t_i)^{\beta}-\tau_{i-1}^{\beta})]^{n_i}}{n_i!}\right)$$

$$.\exp.(-\gamma(\tau_{k-1}+t_k)^{\beta}) \qquad (3.6.5)$$

Taking natural logarithms:

$$\ell(n_1,\ldots n_k/\gamma,\beta) = (\sum_{i=1}^{k} n_i)\log\gamma + \sum_{i=1}^{k} n_i \log[(\tau_{i-1}+t_i)^{\beta}-\tau_{i-1}^{\beta}\ ]$$

$$-\gamma(\tau_{k-1}+t_k)^{\beta} - \sum_{i=1}^{k}\sum_{j=1}^{k} \log j \qquad (3.6.6)$$

Again by differentiating with respect to γ and equating to zero:

$$\hat{\gamma} = \frac{\sum_{i=1}^{k} n_i}{(\tau_{k-1}+t_k)^{\beta}} \qquad (3.6.7)$$

Then $\hat{\gamma}$ can be substituted into (3.6.6) which can then be solved numerically for $\hat{\beta}$.

The MLE $\hat{\gamma}$ and $\hat{\beta}$ of γ and β are substituted in (3.6.4) to obtain the predictive probability function of N_{k+1} (the number of failures during the $(k+1)^{th}$ test phase). Also the probability that no failure will occur during an interval t will be given by:

$$\hat{R}(t) = \exp.(-\hat{\gamma}\ [(\tau_k+t)^{\hat{\beta}} - \tau_k^{\hat{\beta}}]) \qquad (3.6.8)$$

This is the reliability function for Duane model in the continuous case, hence the other software reliability growth measures, such as the rate function, ROCOF, IMTBF and the Median time to failure can be used from section (2.8).

Finally, the expected number of failures is given by the mean value function $M(\tau_k, \tau_k+t_{k+1})$, and the predictive distribution function (cdf) is obtainable by summing the probability function.

3.7. Discrete Non-Homogeneous Poisson Process for JM Model (DJMNHPP)

Goel (1980) proposed this model to deal with the situation where inter-failure time data is not available. He assumed that the number of undetected faults at any time is finite, and the initial number of faults to be detected is finite and equal μ.

Goel argued informally and deterministically that the number of faults detected in $(t,t+\Delta t)$ should be proportional to the number of undetected faults, essentially the JM assumption that all faults contribute equally to the ROCOF of the system. Specifically, if $M(t)$ is the mean function, $E(N(t))$, Goel assumes that:

$$M'(t) = (\mu - M(t)).\Phi \qquad (3.7.1)$$

so that:

$$M(t) = \mu(1 - e^{-\Phi t}) \qquad (3.7.2)$$

Goel then goes on to assume that the observed stochastic process is a NHPP with mean function (3.7.2). This extra assumption is not justified in Goel's work, but does not seem unreasonable as an approximation to the exact JM model. A more formal justification comes from observing that if, in the JM model, we assume that N is a random variable with Poisson (μ) distribution, the unconditional process is exactly the above NHPP.

The expected number of failures in $(\tau_{i-1}, \tau_{i-1}+t_i)$ will be:

$$M(\tau_{i-1}, \tau_{i-1}+t_i) = M(\tau_{i-1}+t_i) - M(\tau_{i-1})$$

$$= \mu\, e^{-\Phi\tau_{i-1}} \quad (1 - e^{-\Phi t_i}) \qquad (3.7.3)$$

The probability mass function of N_i, the number of failures in the phase i with length t is:

$$P(n_i) = \frac{[\mu\, e^{-\Phi\tau_{i-1}}\,(1 - e^{-\Phi t_i})]^{n_i}}{n_i\,!}$$

$$\exp.(-\mu\, e^{-\Phi\tau_{i-1}}(1 - e^{-\Phi t_i})) \qquad (3.7.4)$$

Now, let (n_1,t_1), (n_2,t_2),....(n_k,t_k) be the data points for the first k test intervals, where n_i is the number of failures in the interval i with length t_i. The likelihood function is:

$$L(n_1,\ldots.n_k/\mu,\Phi) = \prod_{i=1}^{k} \frac{[\mu\, e^{-\Phi\tau_{i-1}}(1-e^{-\Phi t_i})]^{n_i}}{n_i!}$$

$$.\exp(-\mu\,(1 - e^{-\Phi(\tau_{k-1}+t_k)})) \qquad (3.7.5)$$

Taking the natural logarithms we get:

$$\ell(n_1,\ldots n_k/\mu,\Phi) = (\sum_{i=1}^{k} n_i)\log\mu - \Phi\sum_{i=1}^{k} n_i\tau_{i-1}$$

$$+\sum_{i=1}^{k} n_i \log(1 - e^{-\Phi t_i})$$

$$-\mu\,[1 - e^{-\Phi\tau_k}] - \sum_{i=1}^{k}\sum_{j=1}^{n_i}\log j \qquad (3.7.6)$$

The MLE $\hat{\mu}$ is obtained by successively differentiating (3.7.6) with respect to μ and Φ and equating to zero. This gives:

$$\hat{\mu} = \frac{\sum_{i=1}^{k} n_i}{1 - e^{-\Phi\tau_k}} \qquad (3.7.7)$$

Substituting (3.7.7) into (3.7.6) and solving numerically for $\hat{\Phi}$ we get the MLE of Φ.

The predictive probability function of N_{k+1} is obtained by substituting $\hat{\mu}$ and $\hat{\Phi}$ in (3.7.4). Also the reliability function R(t) is given by:

$$\hat{R}(t) = \exp(-\hat{\mu}\, e^{-\hat{\Phi}\tau_k}(1 - e^{-\hat{\Phi}t})) \qquad (3.7.8)$$

This is the reliability function of the JMNHPP model in section (2.9); the other software reliability growth mesures described in that section, such as Rate Function, ROCOF, IMTBF and the Median time to failure can be used here.

Finally, the expected number of failures in the phase k+1 is given by $M(\tau_k, \tau_k + t_{k+1})$ which is:

$$M(\tau_k, \tau_k + t_{k+1}) = \hat{\mu}\, \exp(-\hat{\Phi}\tau_k)(1 - \exp(-\hat{\Phi}t_{k+1})) \qquad (3.7.9)$$

and the predictive cdf is obtained by summing the probability mass function.

2.8. Discrete Non-Homogeneous Poisson Process for L Model (DNHPPL)

Littlewood (1984) assumes that the program starts life containing N faults, each of which will manifest itself in a failure after (independently) Pareto distributed time. Assuming that N has a Poisson distribution with mean μ, then by mixing over N, we get a NHPP with mean value function:

$$M(t) = \mu\,[1 - (\frac{\beta}{\beta + t})^{\alpha}] \quad (3.8.1)$$

The expected number of events in the interval $(\tau_{i-1},\tau_{i-1}+t_i)$ is given by:

$$M(\tau_{i-1},\tau_{i-1}+t_i) = M(\tau_{i-1}+t_i) - M(\tau_{i-1})$$

$$= \mu(\frac{\beta}{\beta+\tau_{i-1}})^{\alpha}\,[1 - (\frac{\beta - \tau_{i-1}}{\beta+\tau_{i-1}+t_i})^{\alpha}] \quad (3.8.2)$$

The model assumes each fix to be <u>perfect whenever a failure occurs.</u> The number of events in the i^{th} interval is Poisson distributed with probability mass function:

$$P(n_i/\mu,\alpha,\beta) = \frac{\{\mu(\frac{\beta}{\beta+\tau_{i-1}})^{\alpha}[1 - (\frac{\beta+\tau_{i-1}}{\beta+\tau_{i-1}+t_i})^{\alpha}]\}^{n_i}}{n_i!}$$

$$.\exp\{-\mu(\frac{\beta}{\beta+\tau_{i-1}})^{\alpha}[1 - (\frac{\beta+\tau_{i-1}}{\beta+\tau_{i-1}+t_i})^{\alpha}]\} \quad (3.8.3)$$

Hence for the data vector $(n_1t_1),(n_2,t_2),\ldots(n_k,t_k)$ the likelihood function is:

$$L(n_1,\ldots n_k/\mu,\alpha,\beta) = \prod_{i=1}^{k}\left[\frac{\{\mu(\frac{\beta}{\beta+\tau_{i-1}})^{\alpha}[1-(\frac{\beta+\tau_{i-1}}{\beta+\tau_{i-1}+t_i})^{\alpha}]\}^{n_i}}{n_i!}\right]$$

$$.\exp\{-\mu[1 - (\frac{\beta}{\beta+\tau_{k+1}+t_k})^{\alpha}]\} \quad (3.8.4)$$

Taking the natural logarithms we get:

$$\ell(n_1,\ldots n_k/\mu,\alpha,\beta) = (\sum_{i=1}^{k} n_i)\log\mu + \sum_{i=1}^{k} n_i\alpha\log(\frac{\beta}{\beta+\tau_{i-1}})$$

$$+\sum_{i=1}^{k} n_i\log\,[1 - (\frac{\beta + \tau_{i-1}}{\beta+\tau_{i-1}+t_i})^{\alpha}]$$

$$-\mu\,[1 - (\frac{\beta}{\beta+\tau_k})^{\alpha}] - \sum_{i=1}^{k}\ \sum_{j=1}^{n_i}\log j \quad (3.8.5)$$

and by successively differentiating with respect to μ, α and β and equating to zero, we get:

$$\hat{\mu} = \frac{\sum_{i=1}^{k} n_i}{1 - \left(\frac{\hat{\beta}}{\hat{\beta} + \tau_k}\right)^{\hat{\alpha}}} \tag{3.8.6}$$

The MLEs of α and β are obtained by maximizing (3.8.5) numerically with respect to α and β, after replacing μ by its estimate $\hat{\mu}$ (3.8.6).

By substituting $\hat{\mu}$, $\hat{\alpha}$, and $\hat{\beta}$ in (3.8.3) we can predict the probability function of the number of failures N_{k+1} which will occur during the $(k+1)^{th}$ test interval. The probability of zero failures during the interval of length t is given by:

$$\hat{R}(t) = \exp\left(-\hat{\mu}\left(\frac{\hat{\beta}}{\hat{\beta} + \tau_k}\right)^{\hat{\alpha}}\left[1 - \left(\frac{\hat{\beta} + \tau_k}{\hat{\beta} + \tau_k + t}\right)^{\hat{\alpha}}\right]\right) \tag{3.8.7}$$

This is the form of the reliability function for the LNHPP model (section (2.10)). The same expressions as obtained in (2.10) can be used for the rate function, ROCOF, IMTBF and the Median time to failure. The predictive cdf is obtained by summing the probability mass function. Also the predictive expected number of failures during this interval is given by:

$$M(\tau_k, \tau_k + t_{k+1}) = \hat{\mu}\left(\frac{\hat{\beta}}{\hat{\beta} + \tau_k}\right)^{\hat{\alpha}}\left(1 - \left(\frac{\hat{\beta} + \tau_k}{\hat{\beta} + \tau_k + t_{k+1}}\right)^{\hat{\alpha}}\right) \tag{3.8.8}$$

3.9. Discrete Weibull Model (DW)

Assume, as before, that the program starts life containing N faults and each of these faults will cause failure after independently Weibull distributed time. Let $n_1, n_2, \ldots$ be the number of failures which occurred in the test intervals $t_1, t_2, \ldots$ respectively. Assume that the fixing will take place at the end of each interval, and the numbers of faults removed at the end of each interval are equal to the numbers of observed failures during these intervals. As in section (3.2), and under the same conditions, it is easy to show that $N_i(t_i)$ has a Poisson distribution with mean value function $(N-C_{i-1})\alpha[(\tau_{i-1}+t_i)^{\beta}-\tau_{i-1}^{\beta}]$ i.e.

$$P(n_i/N,\alpha,\beta) = \frac{[(N-C_{i-1})\alpha(\tau_{i-1}+t_i)^{\beta}-\tau_{i-1}^{\beta}]^{n_i}}{n_i!}$$

$$.\exp(-(N-C_{i-1})\alpha[(\tau_{i-1}+t_i)^{\beta}-\tau_{i-1}^{\beta}]) \qquad (3.9.1)$$

where C_{i-1} is the total number of faults which have been fixed during the previous intervals, τ_{i-1} is the total time for the previous intervals.

The probability function in (3.9.1) depends on three parameters, N, α and β which can be estimated by using the ML method. The likelihood function for the data set $(n_1,t_1),\ldots.(n_k,t_k)$ is:

$$L(n_1,\ldots.n_k/N,\alpha,\beta) = \prod_{i=1}^{k}\left\{\frac{[(N-C_{i-1})\alpha((\tau_{i-1}+t_i)^{\beta}-\tau^{\beta}_{i-1})]^{n_i}}{n_i!}\right\}$$

$$.\exp(-\sum_{i=1}^{k}(N-C_{i-1})\alpha[(\tau_{i-1}+t_i)^{\beta}-\tau^{\beta}_{i-1}])) \qquad (3.9.2)$$

with natural logarithms:

$$\ell(n_1,\dots n_k/N,\alpha,\beta) = (\sum_{i=1}^{k} n_i)\log\alpha + \sum_{i=1}^{k} n_i \log(N-C_{i-1})$$

$$+ \sum_{i=1}^{k} n_i \log[(\tau_{i-1}+t_i)^{\beta} - \tau^{\beta}_{i-1}] - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log j$$

$$- \alpha \sum_{i=1}^{k} (N-C_{i-1})[(\tau_{i-1}+t_i)^{\beta} - \tau^{\beta}_{i-1}] \qquad (3.9.3)$$

This form can be maximized numerically to get the MLE $\hat{N}$, $\hat{\alpha}$, and $\hat{\beta}$ for the parameters N, α, and β respectively. By substituting for $\hat{N}$, $\hat{\alpha}$ and $\hat{\beta}$ in (3.9.1) we can predict the probability function for the number of failures N_{k+1} which will occur during the next test interval. The predicted mean value of N_{k+1} is:

$$M(\tau_k,\tau_k+t_{k+1}) = (\hat{N} - C_k)\hat{\alpha}[(\tau_k+t_{k+1})^{\hat{\beta}} - \tau_k^{\hat{\beta}}] \qquad (3.9.4)$$

Also the probability of zero failure during the interval of length t can be predicted by:

$$\hat{R}(t) = \exp(-(\hat{N} - C_{i-1})\hat{\alpha}[(\tau_k + t)^{\hat{\beta}} - \tau_k^{\hat{\beta}}]) \qquad (3.9.5)$$

which is the reliability function for the Weibull order statistic model (section (2.11)) evaluated at $t = t_{k+1}$. The rate function ROCOF, IMTBF and the Median time to failure are:

$$\lambda(t_{k+1}) = (\hat{N} - C_k)\hat{\alpha}\,\hat{\beta}\,(\tau_k + t_{k+1})^{\hat{\beta}-1}$$

$$\text{ROCOF} = (\hat{N} - C_k)\,\hat{\alpha}\,\hat{\beta}\,\tau_k^{\hat{\beta}-1}$$

$$\text{IMTBF} = 1/\text{ROCOF}$$

$$\text{Median} = \tau_k\left[\left\{1 - \frac{\log(.5)}{(\hat{N} - C_k)\hat{\alpha}\tau_k^{\hat{\beta}}}\right\}^{1/\hat{\beta}} - 1\right]$$

Finally, the predictive cdf can be obtained by summation

$$\hat{P}r\langle n_{k+1} \leqslant n\rangle = \sum_{n_{k+1}=0}^{n} P(n_{k+1})$$

Note that, if $\beta = 1$, this model becomes the DJM model. If $N \to \infty$ and $\alpha \to 0$, keeping γ=Nα, a non-zero constant, the form (3.9.1) becomes the Discrete Duane model.

CHAPTER IV

THE ANALYSIS OF PREDICTIVE QUALITY

4.1. Introduction

In the previous two chapters, several prediction systems have been considered. A potential user might reasonably ask which of the many available prediction systems is best for his data source, and which, if any, is close to reality. One way to answer these questions would be for the user to try more than one prediction system for a particular data source and compare the results with one another and, in some way, with reality.

In this chapter, some methods will be described to help the users of software reliability models to carry out such comparisons.

In section (4.2) the u-plot will be described, which is a method of detecting what might be informally termed "bias" in prediction. In section (4.3) the y-plot is introduced, which is a tool for detecting inadequacies of the prediction system in capturing the "trend" in the data.

In section (4.4) two informal measures of "noisiness" of predictions will be introducted. Also, in section (4.5) the Braun statistic will be introduced which measures the variation between each observation and its predicted mean value (MTTF or IMTBF).

Section (4.6) describes a Chi-square procedure for the discrete prediction problem.

Finally, in section (4.7) the prequential likelihood is introduced, which is a general procedure for the examination of prediction quality which embraces all of the above.

4.2. u-Plot

It is well known, Kendall and Stuart (1977), that if X is a random variable with cdf, F(X), then the transformation u = F(X) will follow the uniform distribution U(0,1). In software reliability prediction, the predictive cdf's involve certain parameters which must be estimated from the field data. It does not seem possible, for the software reliability models, to eliminate the effect of this estimation procedure in order to test the goodness of fit of the model alone. Accordingly, in the spirit of Braun and Paine (1977), Littlewood and Sofer (1981), Keiller et al (1983a), Keiller et al (1983b) and Keiller and Littlewood (1984), we shall deal with the predictive distribution directly.

Having observed $t_1, t_2, \ldots, t_{i-1}$, the user wants to predict the random variable T_i. More precisely, he wants a good estimate of:

$$F_i(t_i) \equiv P(T_i < t_i) \qquad (4.2.1)$$

From one of the prediction systems described before, a predictor of $F_i(t_i)$ can be calculated, say $\hat{F}_i(t_i)$.

The user is interested in the closeness of $\hat{F}_i(t_i)$ to the true $F_i(t_i)$. In fact the user may be only interested in summary statistics such as MTTF, the median time to failure, the ROCOF, etc. However, the quality of prediction of these statistics will depend upon the quality of $\hat{F}_i(t_i)$. Clearly, the difficulty of analysing the closeness of $\hat{F}_i(t_i)$ to $F_i(t_i)$ arises from our never knowing, even at later stages of analysis, the true $F_i(t_i)$. The only information which can be obtained is a single realisation of the random variable T_i when the software next fails.

After making the prediction $\hat{F}_i(t_i)$ based upon $t_1, t_2, \ldots, t_{i-1}$, the realisation t_i of T_i will be observed. This is a sample of size one from the true distribution $F_i(t_i)$, and all the analysis of the prediction quality will be based upon these pairs $(\hat{F}_i(t_i), t_i)$ only.

Consider the following sequence of transformations:

$$u_j = \hat{F}_j(t_j) \qquad j = i, i+1 \ldots n \qquad (4.2.2)$$

Each of these is a probability integral transform of the observed t_j, and if each $\hat{F}_j(t)$ were identical to the true $F_j(t)$, it is easy to see that u_j's would be realisations of independent uniform U(0,1) random variables. Thus the problem is transformed, from examining the closeness of $\hat{F}_j(t_j)$ to the true $F_j(t_j)$, to examining whether the sequence $\{u_j\}$ looks like a random sample from U(0.1). There are various aspects of this we can examine in detail. We shall consider first the uniformity. This can be done by comparing the empirical distribution function of $\{u_j, j=i,i+1,\ldots n\}$ with the cdf of the uniform distribution which is the line of unit slope through the origin.

The Kolmogorov distance (the maximum vertical difference between the two plots) can be used to examine the closeness between $\{\hat{F}_j\}$ and the true $\{F_j\}$.

Similarly, this procedure can be applied for the discrete software reliability prediction, since all the prediction systems in this case have the Poisson and negative binomial distributions. The cdf for both are given approximately by the incomplete gamma and beta function respectively [Kendall and Stuart,1977].

Using this procedure on the data given in Table 4.1 we find that Kolmogorov distances for JM and LV models are 0.1896 and 0.1437 respectively. The first figure is significant at level 1%, while the other figure is significant at 5%. The detailed plots, however, tell us more than this. The JM plots are everywhere above the line of unit slope (Figure 4.1), the LV plots almost everywhere below that line (Figure 4.3). This means that the u_j's from JM tend to be too small and those from LV too large. Now u_j represents the predicted probability that T_j will be less than t_j, so consistently too small u_j's suggest that the predictions are underestimating the chance of small t's. In other words the predictions are overestimating the reliability function, which means that the user will expect the system to continue in operation for longer time than is probable. Contrarily, too large u_j's suggest that the model is underestimating the reliability function, which implies that the system will fail in shorter time than is the case.

Thus, the JM plot tells us that the user will face the first situation and LV the second. That is, the JM predictions are too optimistic, the LV predictions are too pessimistic. So the truth, as evidenced from this simple analysis, might be expected to lie somewhere between the predictions of JM and LV, but perhaps closer to LV.

This discussion will be continued later on in Chapter V. At this stage we shall merely note that a conservative position would be to use LV for the next prediction, in the reasonable belief that this will not overestimate the reliability of the product.

TEST SYSTEM : SYS 1

INTER FAILURE TIME

3.	30.	113.	81.	115.
9.	2.	91.	112.	15.
138.	50.	77.	24.	108.
88.	670.	120.	26.	114.
325.	55.	242.	68.	422.
180.	10.	1146.	600.	15.
36.	4.	0.	8.	227.
65.	176.	58.	457.	300.
97.	263.	452.	255.	197.
193.	6.	79.	816.	1351.
148.	21.	233.	134.	357.
193.	236.	31.	369.	748.
0.	232.	330.	365.	1222.
543.	10.	16.	529.	379.
44.	129.	810.	290.	300.
529.	281.	160.	828.	1011.
445.	296.	1755.	1064.	1783.
860.	933.	707.	33.	868.
724.	2323.	2930.	1461.	843.
12.	261.	1800.	865.	1435.
30.	143.	108.	0.	3110.
1247.	943.	700.	875.	245.
729.	1897.	447.	386.	446.
122.	990.	948.	1082.	22.
75.	482.	5509.	100.	10.
1071.	371.	790.	6150.	3321.
1045.	648.	5485.	1160.	1864.
4116.				

TABLE 4.1. Execution times in seconds between successive failures, Musa (1981). Read left to right in rows.

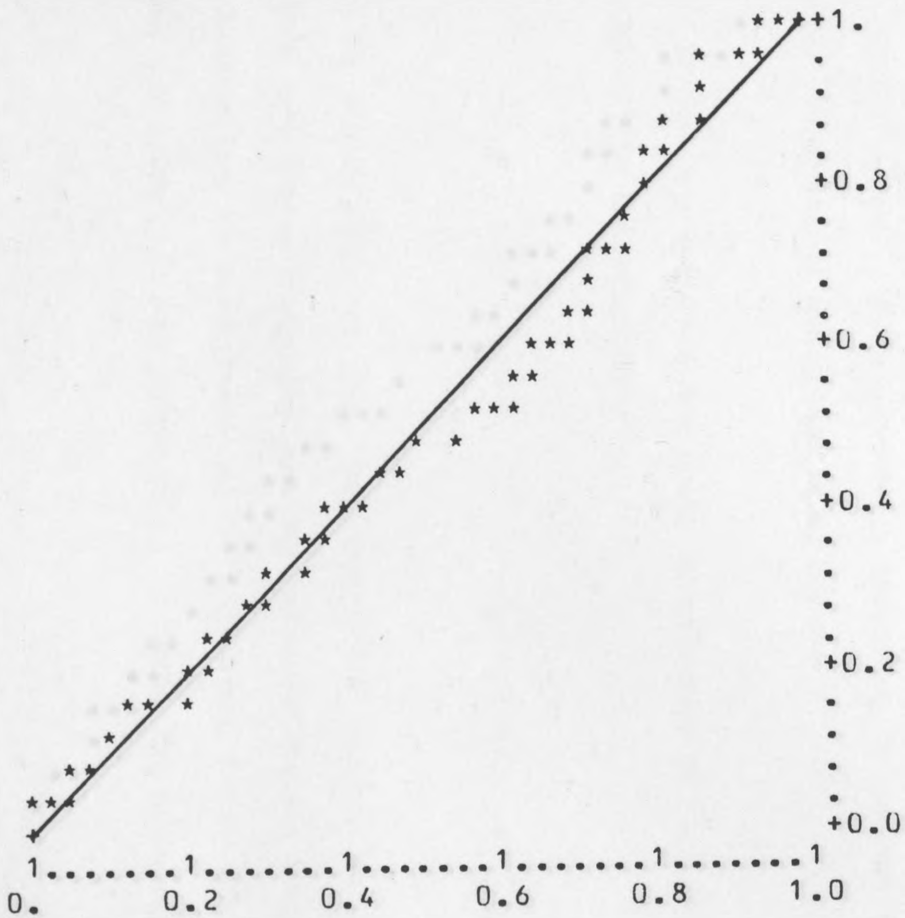


FIG.4.2. JM y-Plots, data of Table 4.1, the plots based on the line printer output.

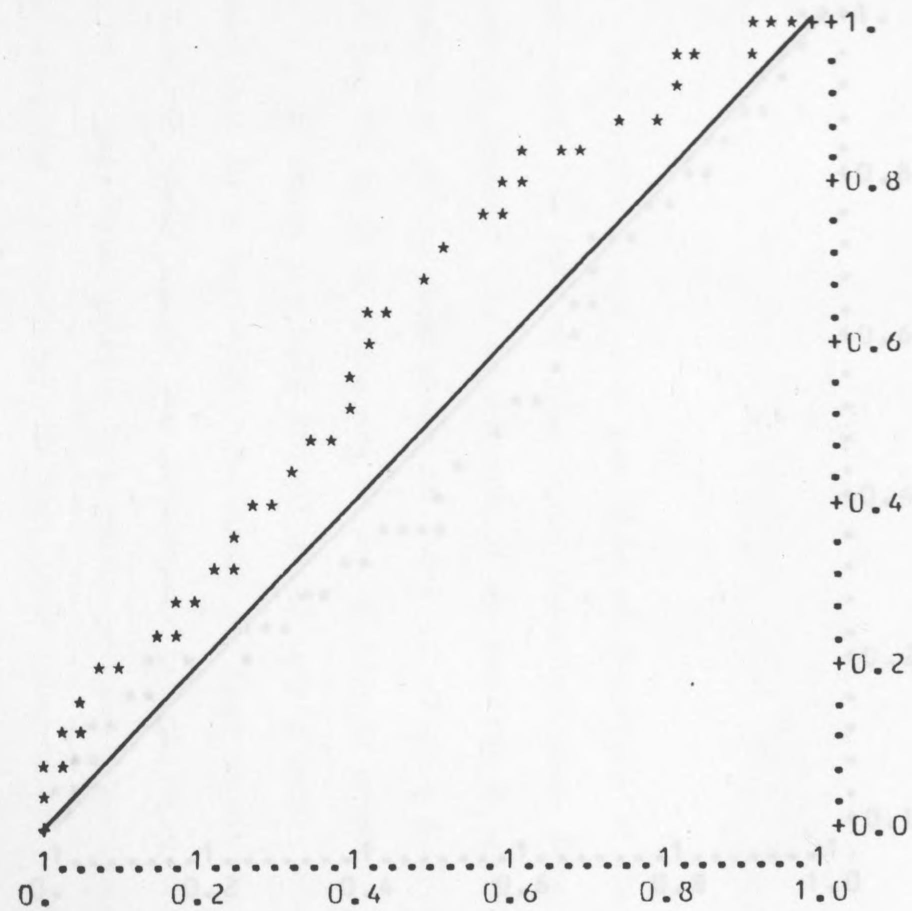


FIG.4.1. JM u-Plots, data of Table 4.1, the plots based on the line printer output.

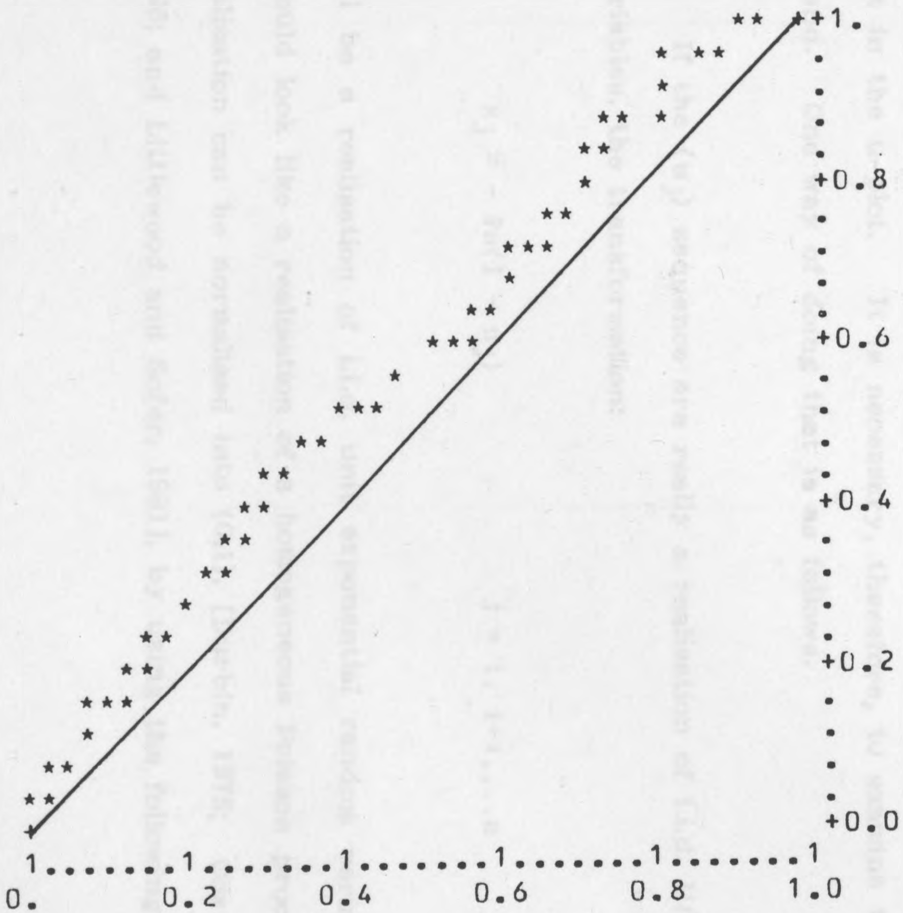


FIG.4.4. LV y-Plots, data of Table 4.1, the plots based on the line printer output.

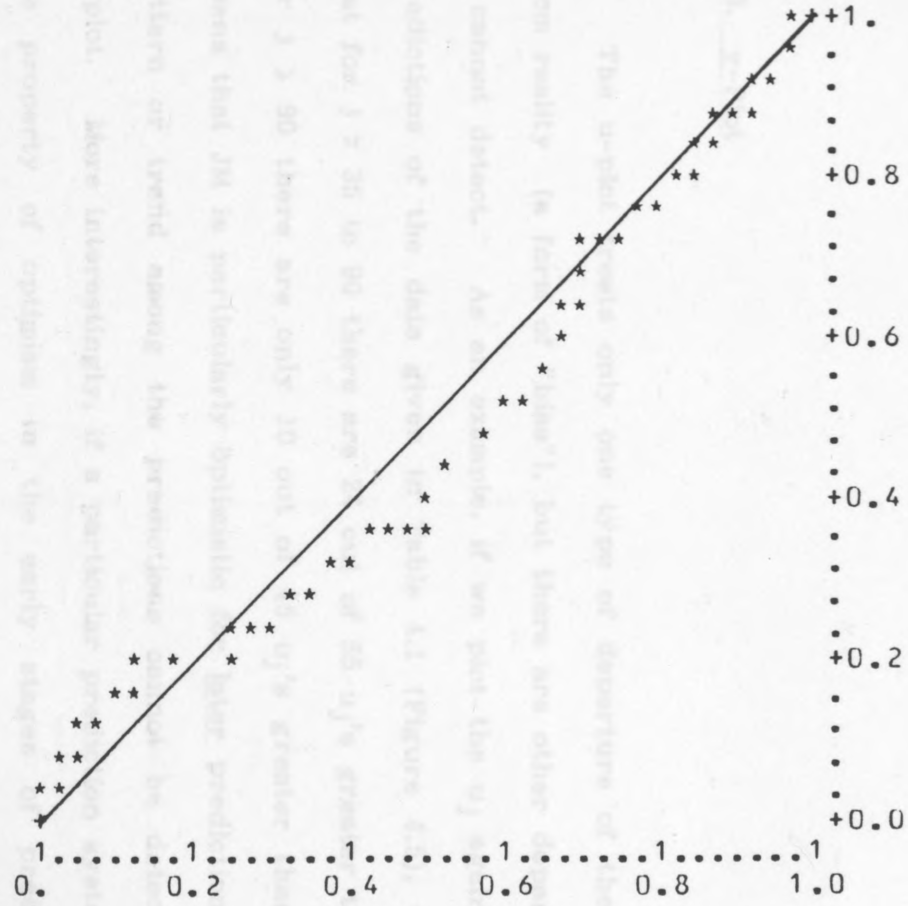


FIG.4.3. LV u-Plots, data of Table 4.1, the plots based on the line printer output.

4.3. y-Plot

The u-plot treats only one type of departure of the predictions from reality (a form of "bias"), but there are other departures which it cannot detect. As an example, if we plot the u_j against j for JM predictions of the data given in Table 4.1 (Figure 4.5), we will find that for j = 35 to 90 there are 24 out of 55 u_j's greater than 0.5, but for j $\geqslant$ 90 there are only 10 out of 46 u_j's greater than 0.5, which means that JM is particularly optimistic for later predictions. Such a pattern or trend among the predictions cannot be detected by the u-plot. More interestingly, if a particular prediction system has, say, the property of optimism in the early stages of predictions and pessimism in the later predictions, these deviations will be averaged out in the u-plot. It is necessary, therefore, to examine the u_j's for trend. One way of doing that is as follows.

If the $\langle u_j \rangle$ sequence are really a realisation of i.i.d. U(0,1) random variables, the transformation:

$$x_j = -\ell n(1 - u_j) \qquad j = i, i+1, \ldots n \qquad (4.3.1)$$

will be a realisation of i.i.d. unit exponential random variables which should look like a realisation of a homogeneous Poisson process. This realisation can be normalised into (0,1), [Durbin, 1975; Cox and Lewis, 1966; and Littlewood and Sofer, 1981], by using the following form:

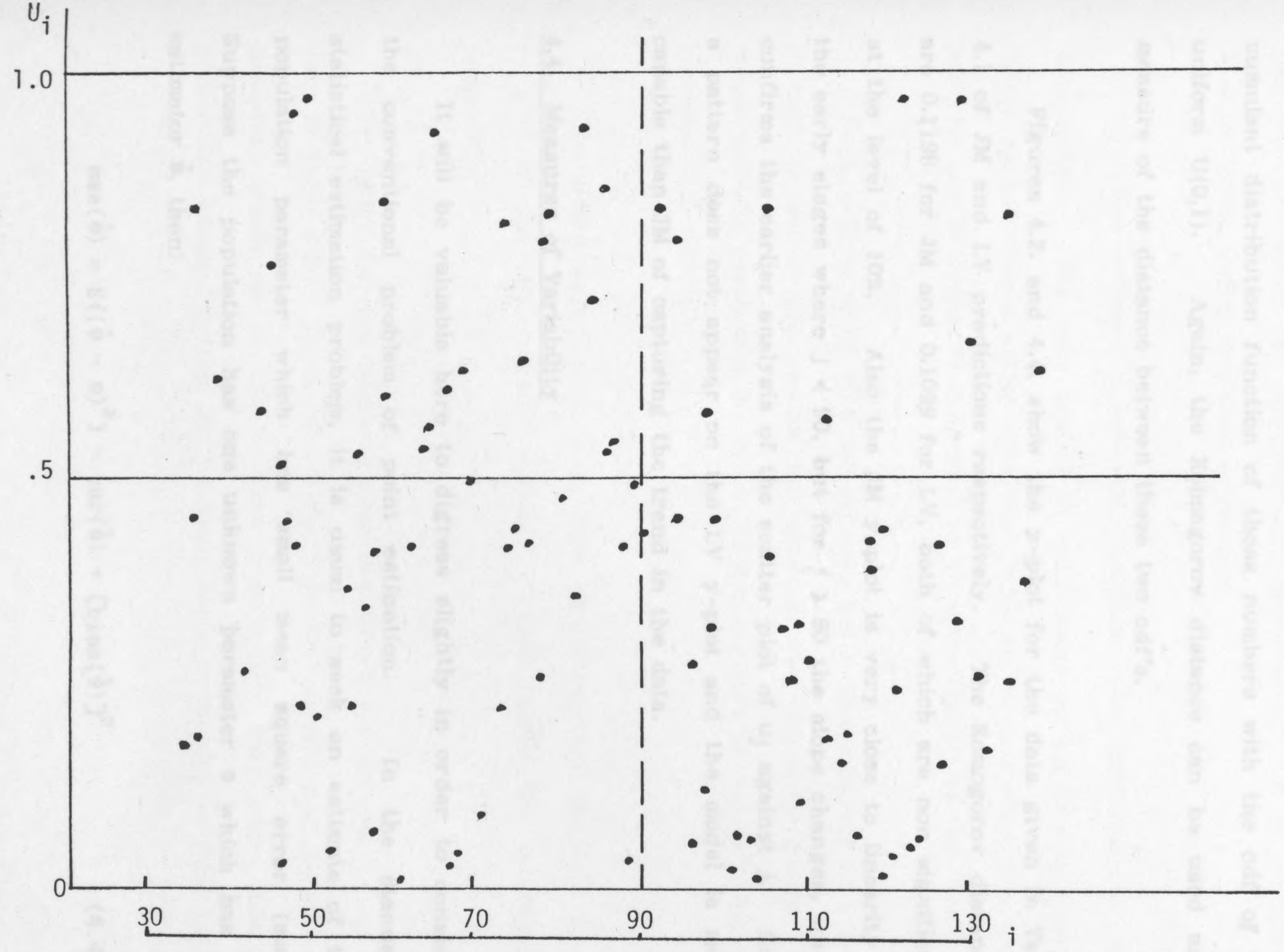


FIG.4.5. Scatter plot of U_i against i for JM predictions from Table 4.1 data.

$$y_j = \sum_{r=1}^{j} x_r / \sum_{r=1}^{n} x_r \qquad j = i, i+1, \ldots n \qquad (4.3.2)$$

The $\{y_j\}$ sequence should be compared with the cdf of the uniforn U(0,1) distribution, and this is carried out by comparing the empirical cumulant distribution function of those numbers with the cdf of the uniform U(0,1). Again, the Kolmogorov distance can be used as a measure of the distance between these two cdf's.

Figures 4.2. and 4.4. show the y-plot for the data given in Table 4.1 of JM and LV predictions respectively. The Komogorov distances are 0.1198 for JM and 0.1099 for LV, both of which are non significant at the level of 10%. Also the JM y-plot is very close to linearity at the early stages where $j < 90$, but for $j \geq 90$ the slope changes, which confirms the earlier analysis of the scatter plot of u_j against j. Such a pattern does not appear on the LV y-plot and the model is more capable than JM of capturing the trend in the data.

4.4. Measures of Variability

It will be valuable here to digress slightly in order to consider the conventional problem of point estimation. In the standard statistical estimation problem, it is usual to seek an estimate of the population parameter which has small mean square error (mse). Suppose the population has one unknown parameter θ which has an estimator $\hat{\theta}$, then:

$$\mathrm{mse}(\hat{\theta}) = E\{(\hat{\theta} - \theta)^2\} = \mathrm{var}(\hat{\theta}) + [\mathrm{bias}(\hat{\theta})]^2 \qquad (4.4.1)$$

Now, suppose there are two estimators, say, $\hat{\theta}_1$, $\hat{\theta}_2$, both having the same mse, but not the same variance. There is clearly a trade-off between the bias of the estimator and its variance. It is not obvious how to choose among two such estimators.

In the software prediction problem the situation is more complicated. The user wishes at each stage to estimate a function, $F_j(t_j)$, rather than a scalar (the true value of θ). Howver, the u-plot analysis can be taken as analogous to an investigation of the "bias".

From this point of view we can see the y-plot analysis as a crude attempt to detect changes in the bias as the reliability grows. It would only be reasonable to expect this procedure to detect quite slow changes.

The u-plot and y-plot, then, give us information which is similar to bias in the conventional context. This leaves the question of variability.

In Figure 4.6. the medians plots are shown for JM and LV on the data of Table 4.1. These reveal the (relative) optimism of JM and the (relative) pessimism of LV. More importantly, they show that the predictions emanating from JM are more variable than those from LV.

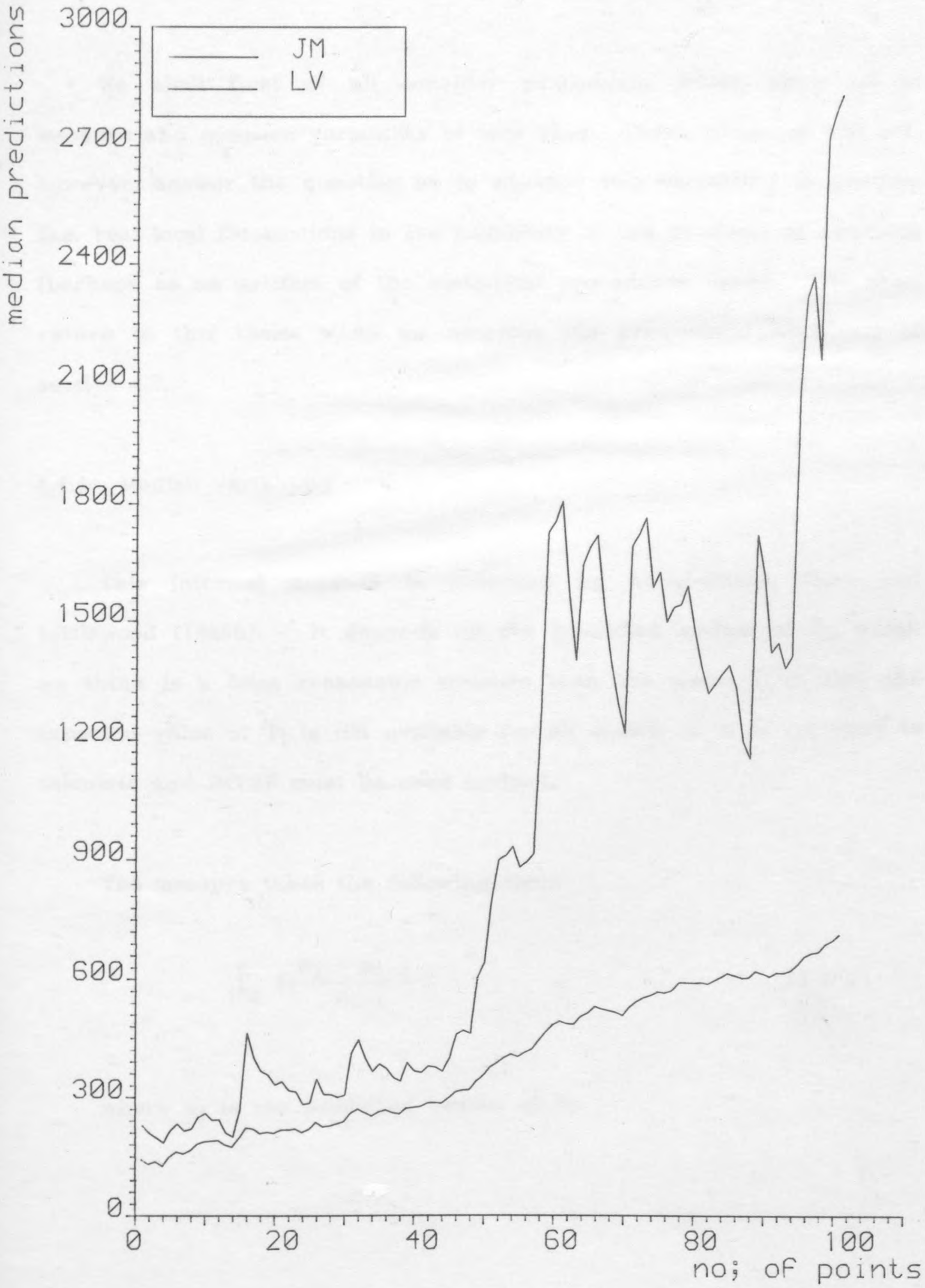


FIG.4.6. The plots of predictive medians of JM&LV for the data in Table 4.1

We shall first of all consider procedures which allow us to measure and compare variability of this kind. These measures will not, however, answer the question as to whether this variability is genuine (i.e. real local fluctuations in the reliability of the product) or spurious (perhaps as an artifact of the statistical procedures used). We shall return to this theme when we consider the prequential likelihood in section 4.7.

4.4.1. Median variability

This informal measure is proposed by Abdel-Ghaly, Chan and Littlewood (1985b). It depends on the predicted median of T_i, which we think is a more reasonable measure than the mean. In fact the expected value of T_i is not available for all models or it is not easy to calculate and IMTBF must be used instead.

The measure takes the following form:

$$\sum_{i=2}^{n} \left| \frac{m_i - m_{i-1}}{m_{i-1}} \right| \qquad (4.4.2)$$

where m_i is the predicted median of T_i.

We expect the system of prediction with most variable predictions will give a greater value for the above form. For example, the values for JM and LV models are 9.5725 and 2.9610 (Table 4.2) for the data in Table 4.1.

No formal test can be carried out for this statistic. Thus only comparisons will be carried out between the value of the measure for different models on the same data set (Table 4.2).

4.4.2. Rate variability

Abdel-Ghaly, Chan and Littlewood (1985b) proposed this statistic as a different criterion to help the user in comparing between different prediction systems. The statistic has the following form:

$$\sum_{i=2}^{n} \left| \frac{r_i - r_{i-1}}{r_{i-1}} \right| \tag{4.4.3}$$

where r_i is the predicted ROCOF of T_i evaluated at $t_i = 0$

Again, this statistic will only be used to compare different prediction systems on the same data set. We believe that the better prediction systems will give smaller values of the statistic. For example, the value of JM and LV models are 8.3659 and 3.1844 respectively (Table 4.2). The values of the statistic will be ranked where the smaller values will get the lower ranks.

Test Continuous Data System 1

No.of Observations: = 136
Starting Sample Size : = 35

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	9.5725	10	8.3659	10	1.3118	10	.1896 1%	10	.1198 20%	9	5.9914 770.2707	9
BJM	8.6087	8	7.2078	8	1.1122	9	.1702 2%	9	.1161 20%	8	3.9588 770.6942	10
L	4.1417	4	3.8293	4	0.8216	2	.0805 N.S.	2	.0642 N.S.	1	5.95499 761.3816	1
BL	4.3043	6	3.9653	6	0.8361	3	.1361 5%	5	.0751 N.S.	4	6.1570 763.9245	4
LV	2.9610	2	3.1844	3	0.8957	7	.1437 5%	6	.1099 20%	6	5.6750 764.8684	6
KL	2.7293	1	3.0126	2	0.8881	5	.1336 5%	4	.1128 20%	7	5.6735 764.7652	5
D	3.1084	3	2.9219	1	0.8910	6	.1590 20%	8	.0931 N.S.	5	5.7480 765.2992	7
JMNHPP	8.6235	9	7.3428	9	1.1093	8	.1525 2%	7	.1245 20%	10	5.9016 768.5679	8
LNHPP	4.1557	5	3.8445	5	0.8193	1	.0805 N.S.	2	.0642 N.S.	1	5.9664 761.3930	2
W	7.3796	7	6.6088	7	0.8621	4	.0747 N.S.	1	.0750 N.S.	3	5.8059 763.0100	3

TABLE 4.2. The Analysis of data in Table 4.1.

4.5. Braun Statistic

In the previous section 4.4., the median and rate variabilities are considered. These two measures are detecting the local variation between adjacent predictions, rather than the variation between predictions and observations.

The measure considered here is a global statistic to examine the closeness of the predictive mean value (MTTF or IMTBF) of T_i to the observation t_i. Braun and Paine (1977) proposed this simple statistic to measure the quality of predictions of reliability growth and to compare between models. The statistic takes the following form:

$$\frac{(n-1)\sum_{i=1}^{n}(t_i - \hat{E}(T_i))^2}{(n-2)\sum_{i=1}^{n}(t_i - \bar{t})^2} \qquad (4.5.1)$$

where $\hat{E}(T_i)$ is the estimated mean of T_i, i.e. the expectation of the predictive distribution $\hat{F}_i(t_i)$, and n is the number of predictions.

Braun and Paine claimed that this statistic will be small (certainly less than one) if the model is of any use. In fact, the normalising denominater is not strictly necessary here, since it will be the same for all prediction systems on a particular data set. The values of this statistic are compared and the prediction system with smaller value is preferred. For example, Table 4.2 shows the value of the statistic for the different systems of predictions on the data set given in Table 4.1. The values for JM and LV models are 1.3118 and 0.8957, which suggests that LV is performing better for this data set than JM.

The form (4.5.1) can be modified as follows:

$$\frac{\sum_{i=1}^{k}(n_i - \hat{E}(N_i))^2 t_i}{\sum_{i=1}^{k}(n_i - \bar{n})^2 t_i} \tag{4.5.2}$$

to be used for the discrete software data, where $\hat{E}(N_i)$ is the estimated mean of $N_i(t_i)$ and k is the number of intervals to be predicted.

The data of Table 4.3. were obtained by "discretising" the data of Table 4.1. in an obvious way. The Braun statistic (4.5.2) is calculated for the predictions of the various discrete models for this data, giving the values shown in Table 4.4. Those values for DJM and DLV models are 1.3231 and 1.2911 respectively.

There is no obvious way of carrying out a formal test to see whether a particular realisation of the statistic is "too large", so this should be seen as a comparative procedure.

I	NR(I)	TN(I)	TP(I)
1	15	0.	1000.00
2	5	15.00	1000.00
3	4	20.00	1000.00
4	3	24.00	1000.00
5	1	27.00	1000.00
6	10	28.00	1000.00
7	4	38.00	1000.00
8	6	42.00	1000.00
9	1	48.00	1000.00
10	0	49.00	1000.00
11	6	49.00	1000.00
12	4	55.00	1000.00
13	3	59.00	1000.00
14	2	62.00	1000.00
15	1	64.00	1000.00
16	4	65.00	1000.00
17	3	69.00	1000.00
18	3	72.00	1000.00
19	3	75.00	1000.00
20	1	78.00	1000.00
21	1	79.00	1000.00
22	2	80.00	2000.00
23	2	82.00	2000.00
24	2	84.00	2000.00
25	3	86.00	2000.00
26	2	89.00	2000.00
27	1	91.00	2000.00
28	0	92.00	2000.00
29	2	92.00	2000.00
30	3	94.00	2000.00
31	2	97.00	2000.00
32	5	99.00	2000.00
33	0	104.00	2000.00
34	2	104.00	2000.00
35	2	106.00	2000.00
36	3	108.00	2000.00
37	3	111.00	2000.00
38	3	114.00	2000.00
39	4	117.00	2000.00
40	1	121.00	2000.00
41	0	122.00	2000.00
42	3	122.00	2000.00
43	3	125.00	2000.00
44	0	128.00	2000.00
45	0	128.00	2000.00
46	0	128.00	2000.00
47	1	128.00	2000.00
48	1	129.00	2000.00
49	2	130.00	2000.00
50	0	132.00	2000.00
51	0	132.00	2000.00
52	2	132.00	2000.00
53	1	134.00	2000.00
54	0	135.00	2000.00
55	1	135.00	2000.00

TABLE 4.3. Discretized data of the data set in Table 4.1. Read from left to right in rows.

Test Discrete Data System: AD-D18

No.of Observations: = 55
Starting Sample Size: = 13

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	4.9273	8	4.3498	8	1.3231	8	109.101 .1%	43.505 .1%(8)	.4467 1%	8	1.3097 79.6583	8
DL	1.8329	4	1.7110	4	.9736	2	42.800 N.S.		.2263 5%	4	1.3473 67.4904	2
DLV	1.1151	2	1.0804	2	1.2911	6	35.877 N.S.		.1006 N.S.	2	1.4867 71.7529	5
DKL	1.0693	1	1.0137	1	1.3139	7	36.199 N.S.		.1166 N.S.	3	1.5732 71.9156	6
DD	1.5875	3	1.5001	3	1.2160	4	35.492 N.S.		.0886 N.S.	1	1.4414 70.2690	4
DJMNHPP	4.3993	7	3.8239	7	1.2186	5	87.188 .1%	25.971 .5%(9)	.3934 1%	7	1.3233 75.3720	7
DLNHPP	1.8738	5	1.7482	5	.9700	1	42.678 N.S.		.2288 5%	5	1.3295 67.3376	1
DW	4.1581	6	3.7995	6	1.0836	3	46.737 N.S.		.2599 1%	6	1.4414 69.4048	3

TABLE 4.4. The analysis of Table 4.3 data (The discretized data for the data in Table 4.1)

4.6. Chi-Square Test

This test is intended to be used in comparing the performance solely of the discrete software prediction systems. In the standard application of this test, the n observations in a random sample from a population are classified into k mutually exclusive classes. The test statistic is:

$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{4.6.1}$$

where

O_i = the observed number in the class i

E_i = the expected number in the class i under the null hypothesis.

In the conventional context, the E_i depend on some unknown parameters which have to be estimated from the sample itself. The distribution of the statistic under the null hypothesis is then chi-square with degrees of freedom dependent on the number of unknown parameters. Here the estimation of parameters is carried out once using the full data set.

In the software reliability prediction problem the situation is different since we based our expectation for the number of failures during phase i on $n_1(t_1)$, $n_2(t_2)$,....$n_{i-1}(t_{i-1})$. The form (4.6.1) becomes:

$$Q = \sum_{j=1}^{k} \frac{[n_j(t_j) - E(\hat{N}_j(t_j))]^2}{E(\hat{N}_j(t_j))} \tag{4.6.2}$$

where $n_j(t_j)$ is the observed number during the phase j and $E(\hat{N}_j(t_j)$ is the estimated expected number of failure for this phase, i.e. the expected value of the predictive probability distribution of the random variable $N_j(t_j)$. The form (4.6.2) will follow the chi-square distribution with (k-i+1) degrees of freedom if the predictive distributions are the true distribution. Thus the statistic measures the closeness of prediction to reality.

Applying this test on the data in Table 4.3. gives the results shown in Table 4.4. For example, the values of the test for DJM and DLV models are 109.101 and 35.877 respectively. For 42 degrees of freedom, the first figure is significant at 0.1% while the second figure is non-significant even at large levels (Table 4.4). There is thus very strong evidence against the DJM predictions.

4.7. The Prequential Likelihood Ratio

The predictive cdf, $\hat{F}_i(t_i)$, for T_i based on $t_1, t_2, \ldots t_{i-1}$ will be assumed to have a predictive probability density function:

$$\hat{f}_i(t_i) = \frac{d}{dt}\hat{F}_i(t_i) \tag{4.7.1}$$

Then for the sequence $T_i, T_{i+1}, \ldots T_n$ the predictive joint density function is given by:

$$PL_{(n-i+1)} = \hat{f}(t_i, t_{i+1}, \ldots t_n) = \prod_{j=1}^{n} \hat{f}_j(t_j) \tag{4.7.2}$$

This function is called the prequential likelihood (PL). Dawid (1982, 1984a, 1984b, 1985) has treated theoretical issues concerned with the validity of forecasting systems using this notion. Dawid's discussion of calibration, in partiular, is relevant to the software reliability prediction problem [Abdel-Ghaly, Chan and Littlewood, 1985b]. We shall use the prequential likelihood ratio (PLR) in our analysis of the prediction system performance. If there are two prediction systems A with PL:1

$$PL^A_{(n-i+1)} = \prod_{j=i}^{n} \hat{f}^A_j(t_j) \tag{4.7.3}$$

and B with PL:

$$PL^B_{(n=i+1)} = \prod_{j=i}^{n} \hat{f}^B_j(t_j) \tag{4.7.4}$$

the prequential likelihood ratio will be:

$$PLR_{(n-i+1)} = \frac{PL^A_{(n-i+1)}}{PL^B_{(n-i+1)}} = \prod_{j=i}^{n} \frac{\hat{f}^A_j(t_j)}{\hat{f}^B_j(t_j)} \qquad (4.7.5)$$

Dawid (1984a) shows that if the realised sequence of $\{PLR_{(n-i+1)}\}$ prequential likelihood ratios tends to infinity as $n \to \infty$, then the prediction system B will be discredited in favour of prediction system A. Conversely, if $PLR_{(n-i+1)}$ tends to zero as $n \to \infty$, B discredits A.

To get an intuitive feel for the behaviour of the prequential likelihood, let us consider, for simplicity, the problem of predicting a sequence of identically distributed random variables, i.e. $F_i(t) = F(t)$ and $f_i(t) = f(t)$ for all i. Suppose, we have two sequences of predictor densities. The first sequence is "biased" to the left of the true distribution (Figure 4.7), but is not excessively noisy. The observations, which tend to fall in the body of the true distribution, will tend to be in the right hand tails of the predictor densities. It follows that the prequential likelihood will tend to be small.

The second sequence of predictions (Figure 4.8) are very "noisy" but have an expectation close to the true distribution (low "bias"). Again, the observations will tend to lie in the body of the true distribution and so have a tendency to appear in (right or left) tails of the predictor distributions. The prequential likelihood will again tend to be small.

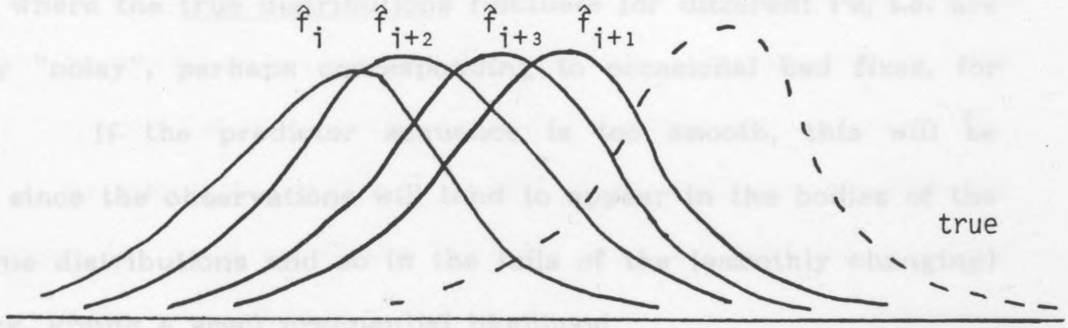


FIG.4.7. The predictions have low "noise" and high "bias".

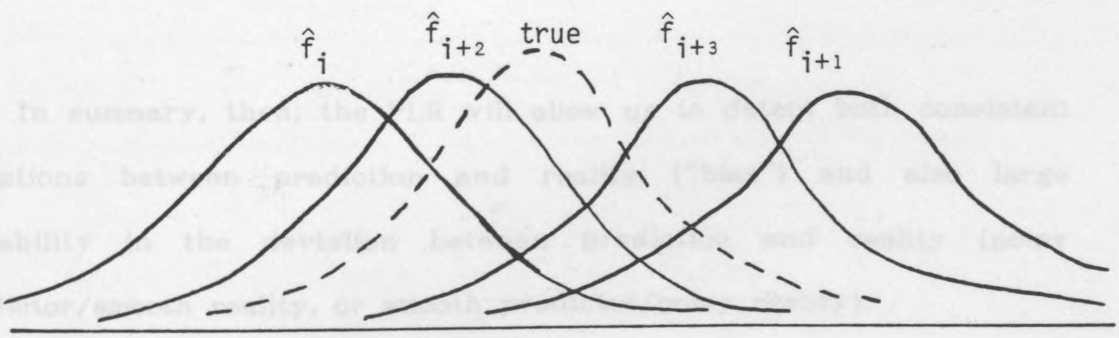


FIG.4.8. The predictions have high "noise" and low "bias".

These arguments extend to our non-stationary case. Consider the case where the true distributions fluctuate for different i's, i.e. are genuinely "noisy", perhaps corresponding to occasional bad fixes, for example. If the predictor sequence is too smooth, this will be detected since the observations will tend to appear in the bodies of the noisy, true distributions and so in the tails of the (smoothly changing) predictors, giving a small prequential likelihood.

Thus the prequential likelihood can, in principle, detect predictors which are too noisy (when the true distributions are not variable) and predictors which are too smooth (when the true distributions are variable). This is a significant improvement over our previous noise statistics, which could detect noisy predictors but not tell whether they reflected actual noise in reliability.

In summary, then, the PLR will allow us to detect both consistent deviations between prediction and reality ("bias") and also large variability in the deviation between prediction and reality (noisy predictor/smooth reality, or smooth predictor/noisy reality).

It is worth emphasising that, although our informal discussion of bias only treated shifts between prediction and reality, it can detect consistent differences of a more complicated nature. Elimination of these generally biased predictors is the subject of recent research on adaptive procedures: Keiller and Littlewood (1984).

As an example of the power of the PLR consider Table 4.5. To emphasise the inadequacy of the u-plot and y-plot procedures when used on their own, Miller (1983) proposed an unbiased predictor based only upon the previous one or two observations. He assumed that the sequence $\langle T_i \rangle$ was of exponential random variables and estimated the mean of T_i by using t_{i-1} or $(t_{i-1} + t_{i-2})/2$, from which can be obtained approximately unbiased predictors $\hat{F}_i(t)$ for T_i. In fact, his intention was to produce an estimator which had a good u-plot ("unbiased"). For the data of Table 4.1 this procedure gives u-plot and y-plot distances of 0.078 and 0.069 respectively which are not significant at the 10% level. Both distances are much better than those for JM and LV (Table 4.2), although the predictions are clearly worthless.

Applying the PLR procedure in comparing JM and LV predictions with Miller's second predictor gives the results shown in Table 4.5.

n	JM $PLR_{(n-i+1)}$	LV $PLR_{(n-i+1)}$
45	4.00	3.26
55	30.8	82.1
65	158.0	517.0
75	8.92×10^4	7.18×10^5
85	9.32×10^5	1.01×10^6
95	4.91×10^6	5.72×10^5
105	2.48×10^6	2.53×10^7
115	6.01×10^5	2.63×10^8
125	3.67×10^6	3.37×10^{10}
135	6.34×10^6	3.90×10^{11}

TABLE 4.5. The PLR of JM versus Miller predictor and LV versus Miller predictor using the data in Table 4.1. The starting sample size i = 35.

In both cases the Miller predictions are clearly being discredited, even in favour of JM, which is known to be a biased prediction system for this data from the previous analysis.

Table 4.6. gives the PLR of JM against LV for the same data set. There is no evidence, for n up to about 90, that one prediction system is preferable to the other. Thereafter, however, there is strong evidence that LV is outperforming JM.

n	$PLR_{(n-i+1)}$
45	1.1935
55	0.3176
65	0.2518
75	0.0962
85	0.7452
95	6.5021
105	0.0884
115	0.001773
125	0.000814
135	0.001196

TABLE 4.6. The PLR of JM versus LV for the data in Table 4.1. The starting sample size i = 35.

CHAPTER V

DATA ANALYSIS AND RESULTS

5.1. Introduction

In this chapter several real data sets will be analysed using the models and tools mentioned in the previous chapters.

The main computational problem in all cases is the successive re-estimation of the ML estimates of the parameters in those models (non-Bayesian) for which this technique is used. In all cases the Nelder-Mead (1965) simplex algorithm is used for the optimisation. This algorithm was chosen, rather than more recent methods for unconstrained optimisation, because of its well-known robustness. This robustness is, however, purchased at the price of less efficiency, and it is probably worth considering the use of Nelder-Mead in association with other techniques for possible commercial application of these models.

All calculations for the thesis are conducted in the same manner. If the data file contains n inter-failure times , $m(\leqslant n-1)$ is chosen as a starting sample size. The MLE's are obtained for that sample and then used to predict log pdf, cdf, rate, ROCOF, MTTF, IMTBF and median for the next failure-time, T_{m+1}. The sample is then increased by the observed t_{m+1} the process repeated for T_{m+2}, and so on. A similar procedure is used for sequential prediction in the discrete case.

The analysis has been carried out for many sets of inter-failure time data and of discrete type. The data were obtained from different sources such as Musa (1979), Braun and Paine (1977), Moek (1984), Ohba (1984), Misra, (1983), and through private communication.

5.2. The Analysis of the inter-failure time data

It is not possible to present the complete output of all the prediction systems on all data sets. Indeed, this is not desirable: one of the objectives of this thesis is to give guidelines to a user as to how he might use summaries of the information to arrive at choices among models. The majority of the results, then, will use a particular summary table. The more extensive analysis of the first three data sets will show how this can be augmented as a result of the questions which arise from the summaries. The important point is that the summary tables steer us through the large number of results available on predictive performance.

The first three data sets were chosen from those which clearly exhibited reliability growth and were of a reasonable size.

5.2.1. Musa System 1 Data

This data set is shown in Table 4.1 which gives the execution time in seconds between successive failures (read from left to right in rows). The set contains 136 points of data. The analysis begins with a sample size of 35 observations to estimate the model's parameters and then make predictions about the 36^{th} failure; successive one-step-ahead predictions are then made.

Table 4.2. summarises the results concerning the quality of performance of the various prediction systems on this data.

Considering the JM model, Table 5.1 gives the parameter estimates at different sample size. It is clear that $\hat{N}$ is close to n (the sample size) at all times and $\hat{\Phi}$ is decreasing as n increases. This implies that the fault rates are not equal as assumed in this model [Littlewood, 1981]. It is noticeable that for n equal 90 or more, the JM predictions for the reliability function are mostly greater than .5 (Figure 4.5). This evidence of over-optimistic prediction is supported by the u-plot (Figure 4.1) where most points lie above the line of unit

Sample size n	$\hat{N}$	$\hat{\Phi}$
40	55	.000199
60	77	.000118
80	99	.000079
90	100	.000077
100	106	.000067
110	119	.000057
120	133	.000040
130	138	.000037

Table 5.1. The MLE's of JM model at different sample size for the data in Table 4.1.

slope, with Kolmogorov distance of 0.1896 which is significant at 1% level. This means that the u_j's from JM tend to be too small, suggesting that the predictions are under-estimating the change of small t's.

The median prediction plot (Figure 5.1) shows that JM, BJM and JMNHPP predict larger medians than the remaining predictions systems. The u-plots for the latter two are similar to the u-plot of JM (Figure 5.2 and 5.4) with Kolmogorov distances significant at the 2% level. These models have the largest ranks on all the statistics shown in Table 4.2.

In particular, these three prediction systems give the worst results on the PL ranks: there is a significant difference between the 7^{th} ranking model, D, and the 8^{th}, JMNHPP.

The y-plots in Figures 4.2, 5.3 and 5.5 for JM, BJM and JMNHPP are very close to linearity in the early stages. This suggests that the too optimistic predictions from these models are occurring mainly at the later stages. This is apparent from the median plots (Figure 5.1).

LV, KL and D models were the least variable prediction systems, since these three models have the smallest median and rate variabilities (Table 4.2). This result is confirmed by visual inspection of the median plots (Figure 5.1) which show also that the predictions of these

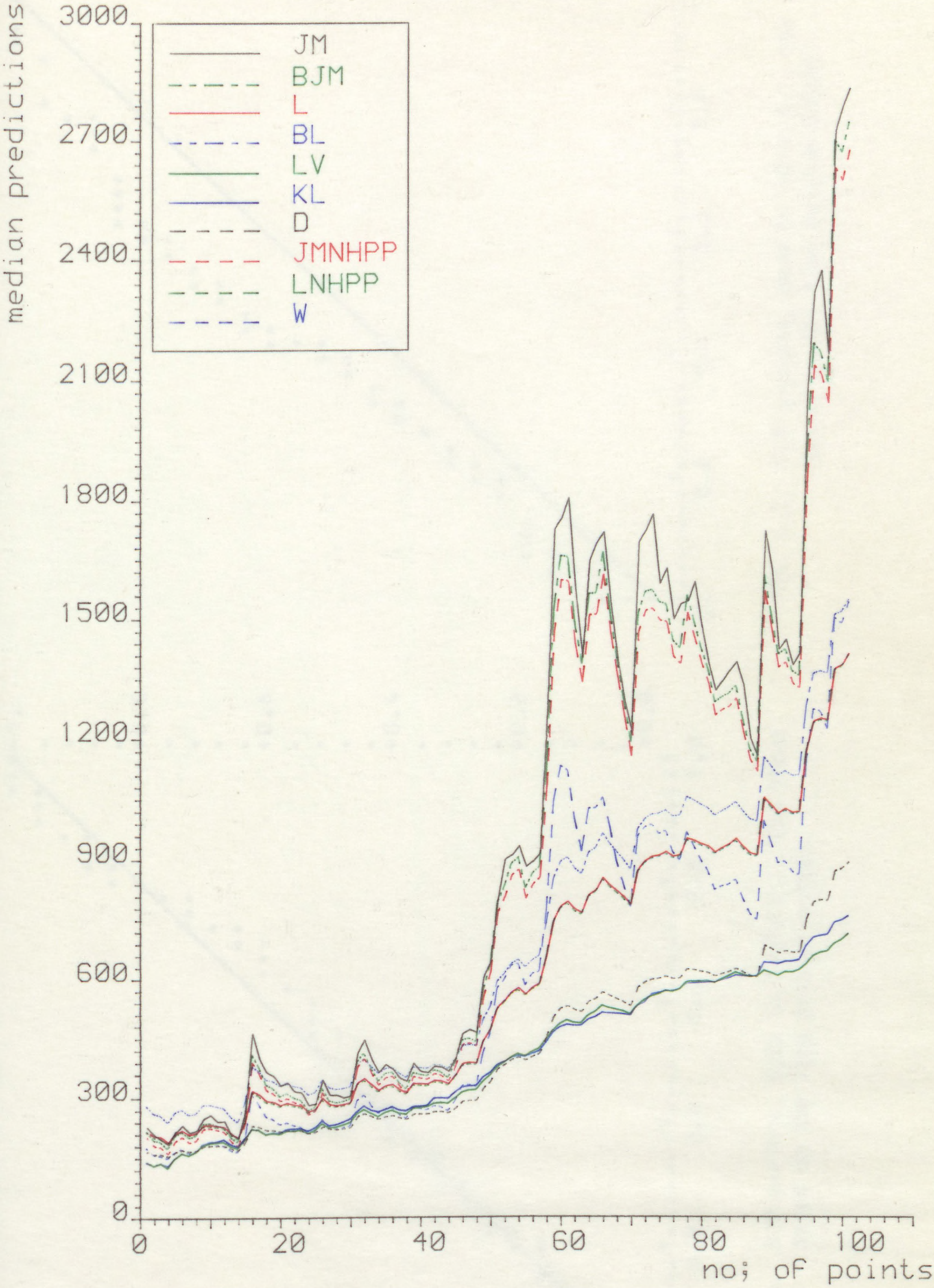


FIG.5.1. The plots of predictive medians for the data in Table 4.1

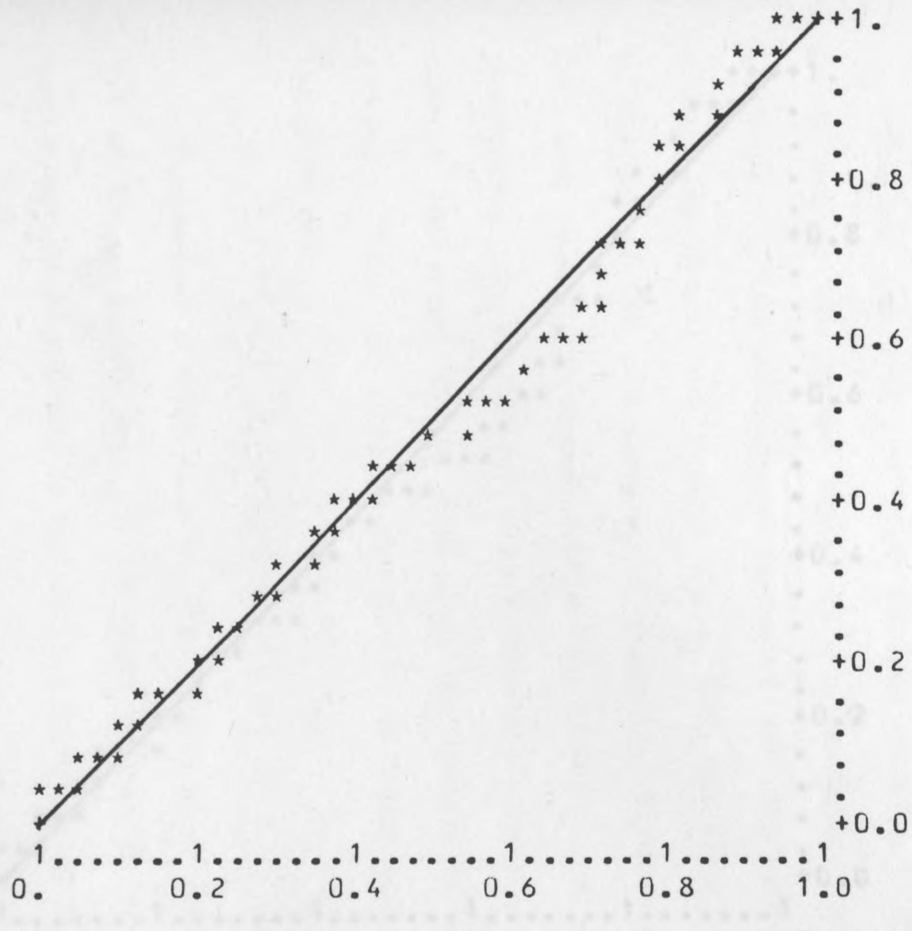


FIG.5.3. BJM y-plots, data in Table 4.1, the plots based on the line printer output.

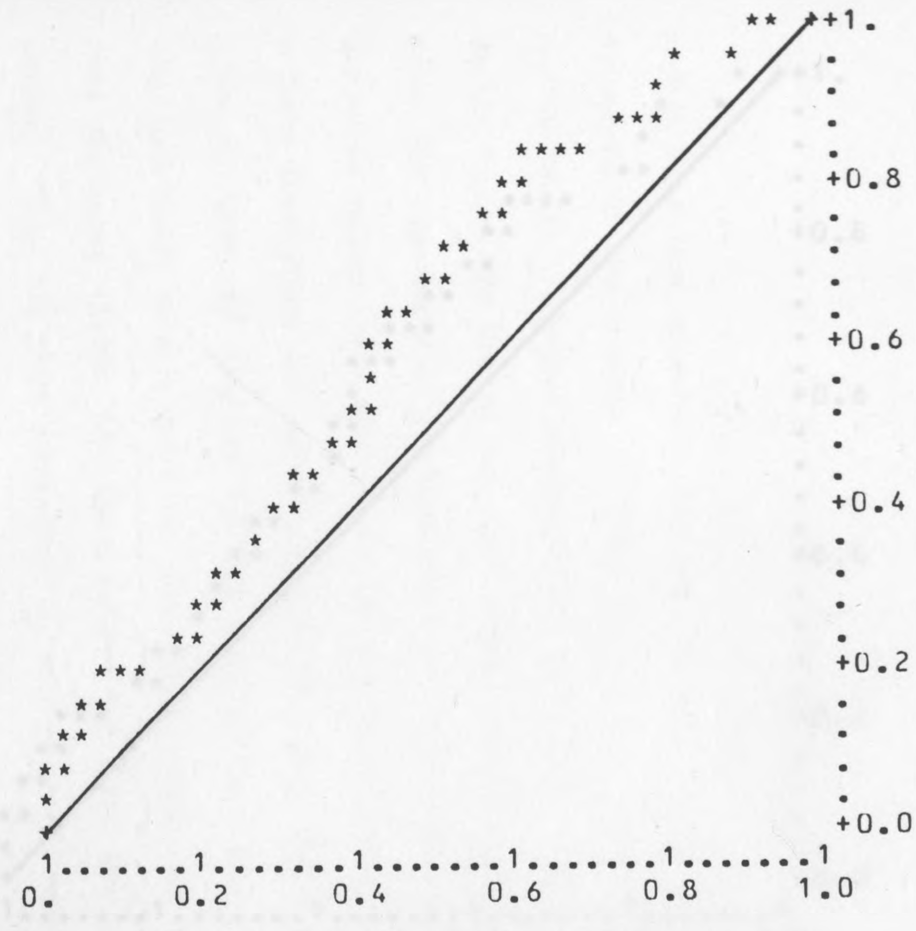


FIG. 5.2. BJM u-plots, data in Table 4.1, the plots based on the line printer output.

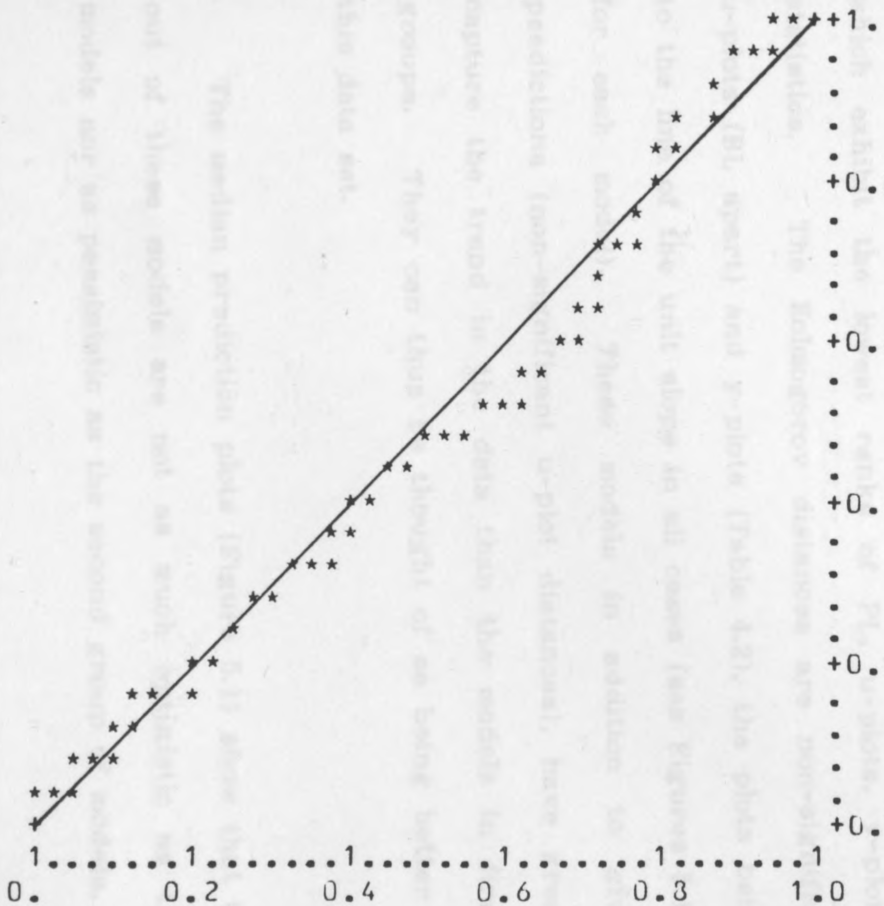
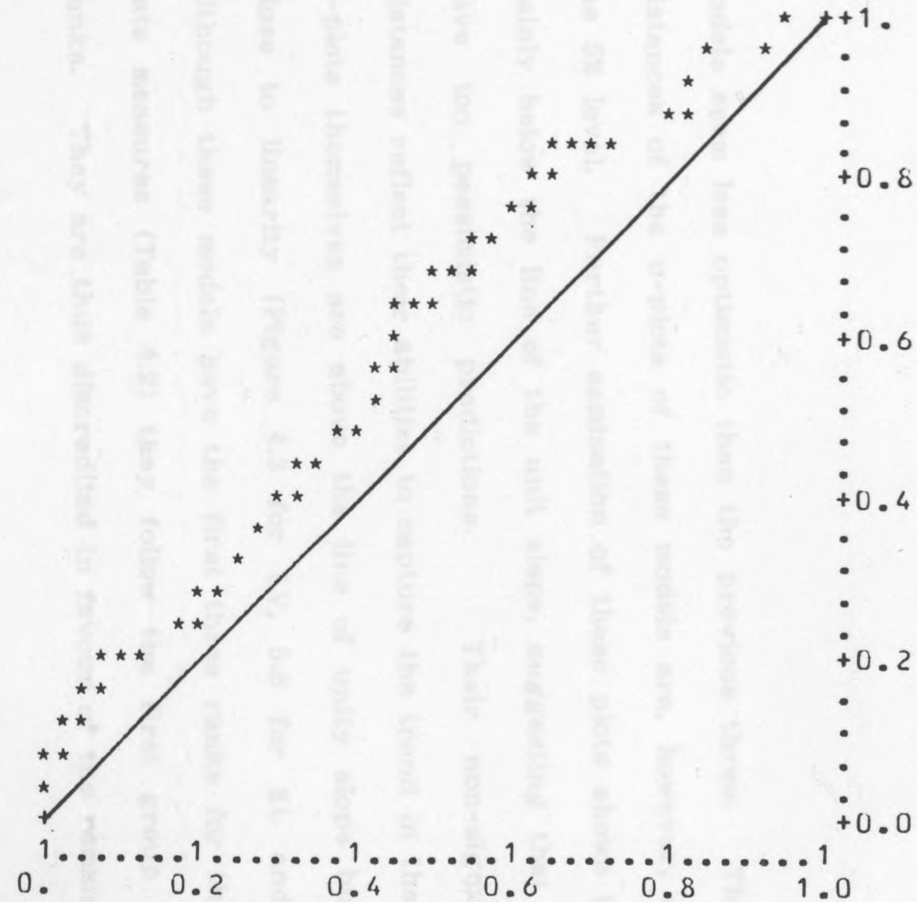


FIG. 5.5. JMNHPP y-plots, data in Table 4.1., the plots based on the line printer output.



FIG, 5.4, JMNHPP u-plots, data in Table 4.1, the plots based on the line printer output.

models seem less optimistic than the previous three. The Kolmogorov distances of the u-plots of these models are, however, significant at the 5% level. Further exmination of these plots shows that they are mainly below the line of the unit slope, suggesting that these models gave too pessimistic predictions. Their non-significant y-plot distances reflect their abilities to capture the trend in the data. The y-plots themselves are above the line of unity slope but with shape close to linearity (Figure 4.3 for LV, 5.6 for KL and 5.8 for D). Although these models have the first three ranks for the median and rate measures (Table 4.2) they follow the first group on the other ranks. They are thus discredited in favour of the remaining models.

The last group of prediction systems contains L, BL, LNHPP and W which exhibit the lowest ranks of PL, u-plots, y-plots and Braun statistics. The Kolmogorov distances are non-significant for both u-plots (BL apart) and y-plots (Table 4.2), the plots being very close to the line of the unit slope in all cases (see Figures 5.10 to 5.17, two for each model). These models in addition to giving unbiased predictions (non-significant u-plot distances), have greater ability to capture the trend in the data than the models in first and second groups. They can thus be thought of as being better calibrated for this data set.

The median prediction plots (Figure 5.1) show that the predictions out of these models are not as much optimistic as the first three models nor as pessimistic as the second group of models, but they are

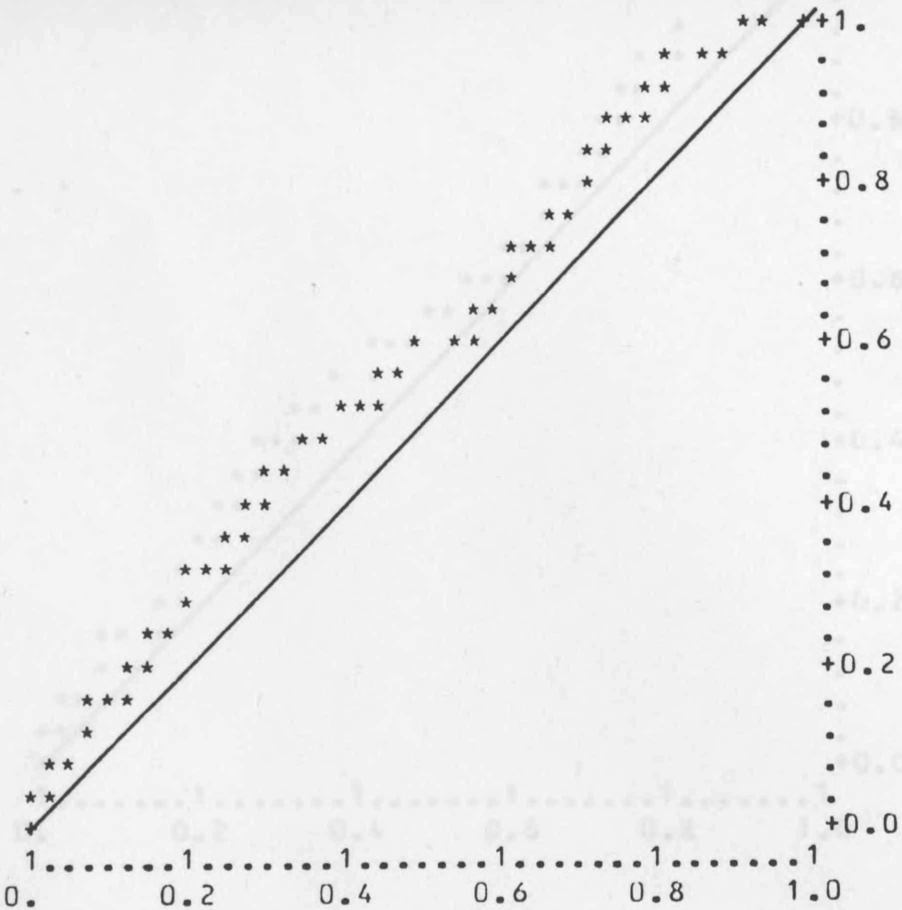


FIG. 5.7. KL y-plots, data in Table 4.1, the plots based on line printer output.

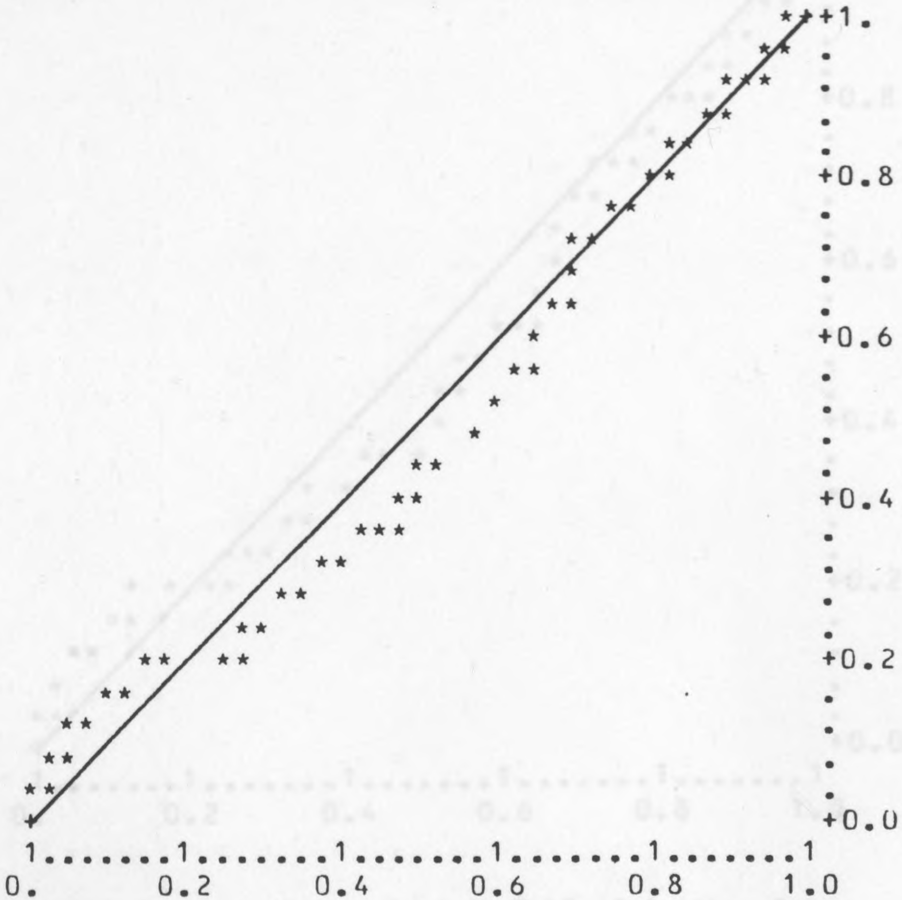


FIG.5.6. KL u-plots, data in Table 4.1, the plots based on line printer output.

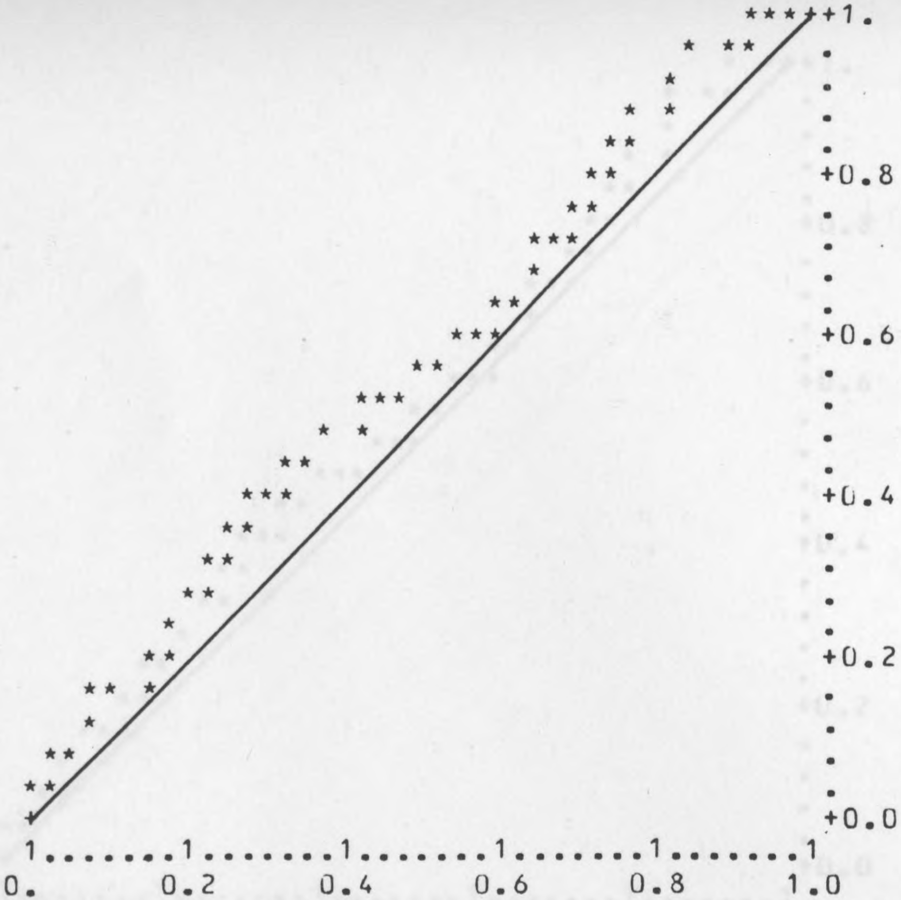


FIG.5.9 D y-plots, data in Table 4.1, the plots. based on the line printer output.

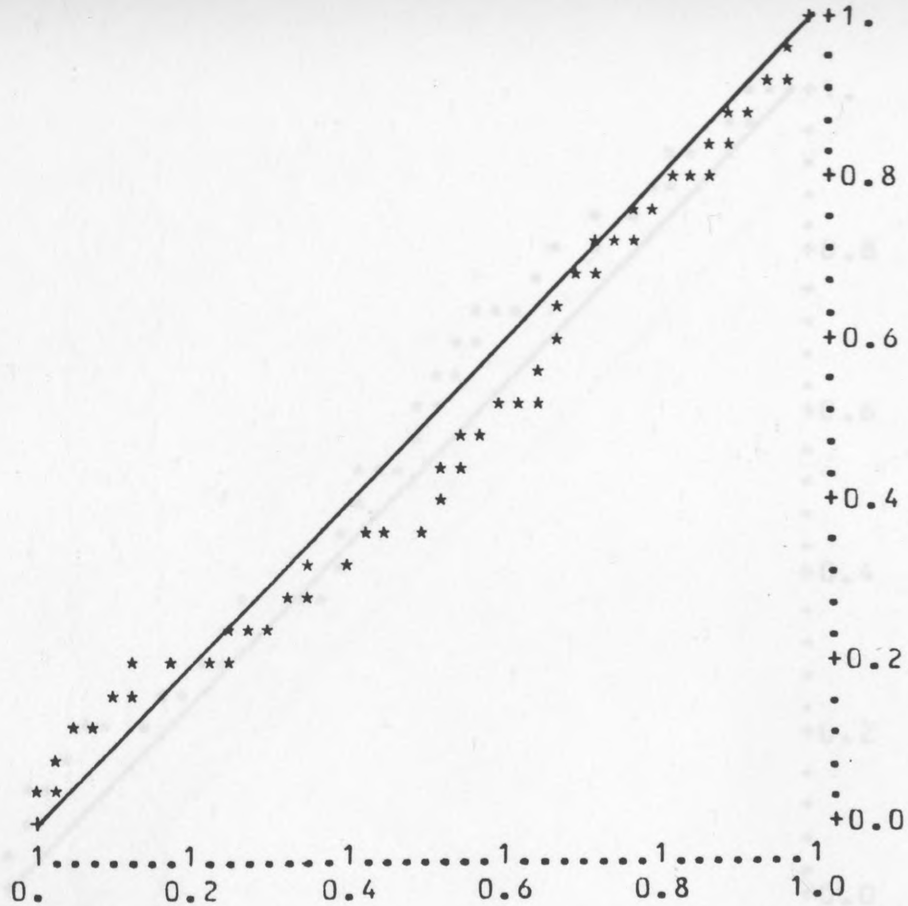


FIG.5.8. D u-plots, data in Table 4.1, the plots based on the line printer output.

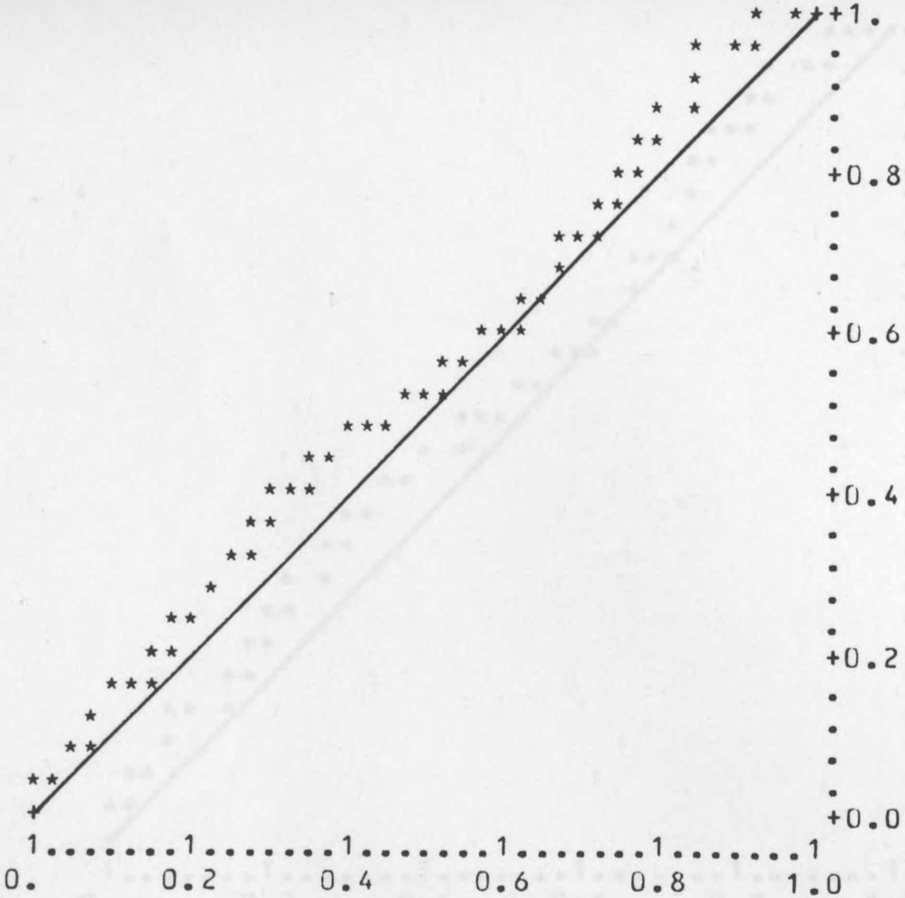


FIG.5.11. L y-plots, data in Table 4.1, the plots based on the line printer output.

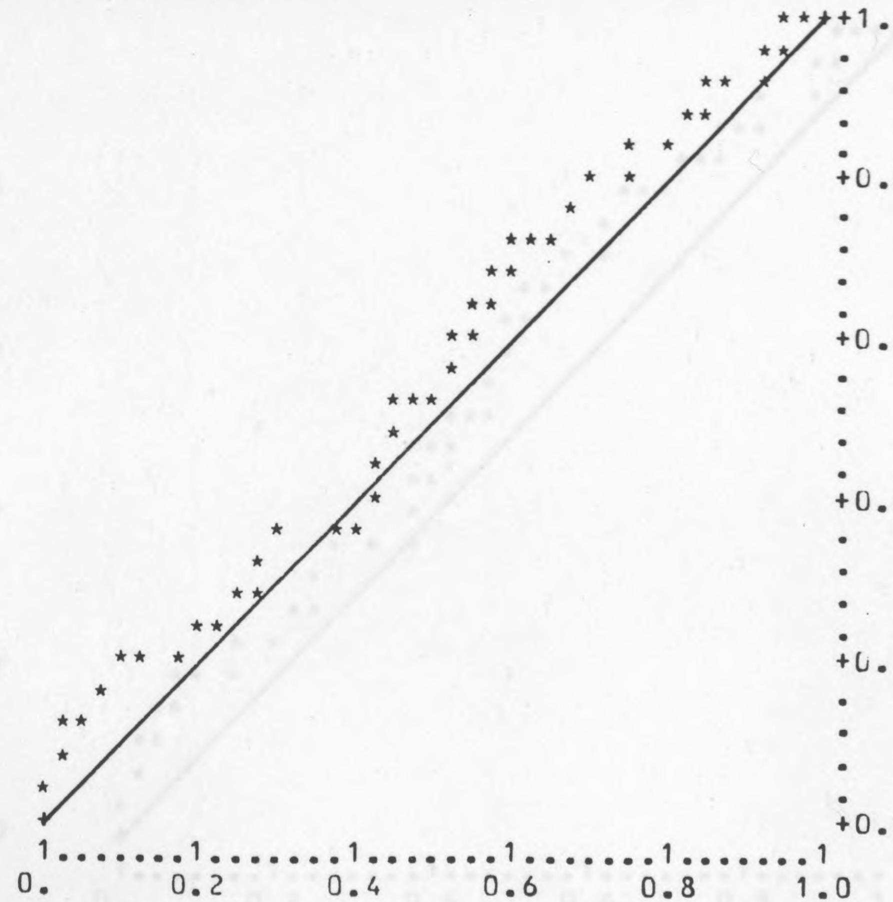


FIG. 5.10. L u-plots, data in Table 4.1, the plots based on the line printer output.

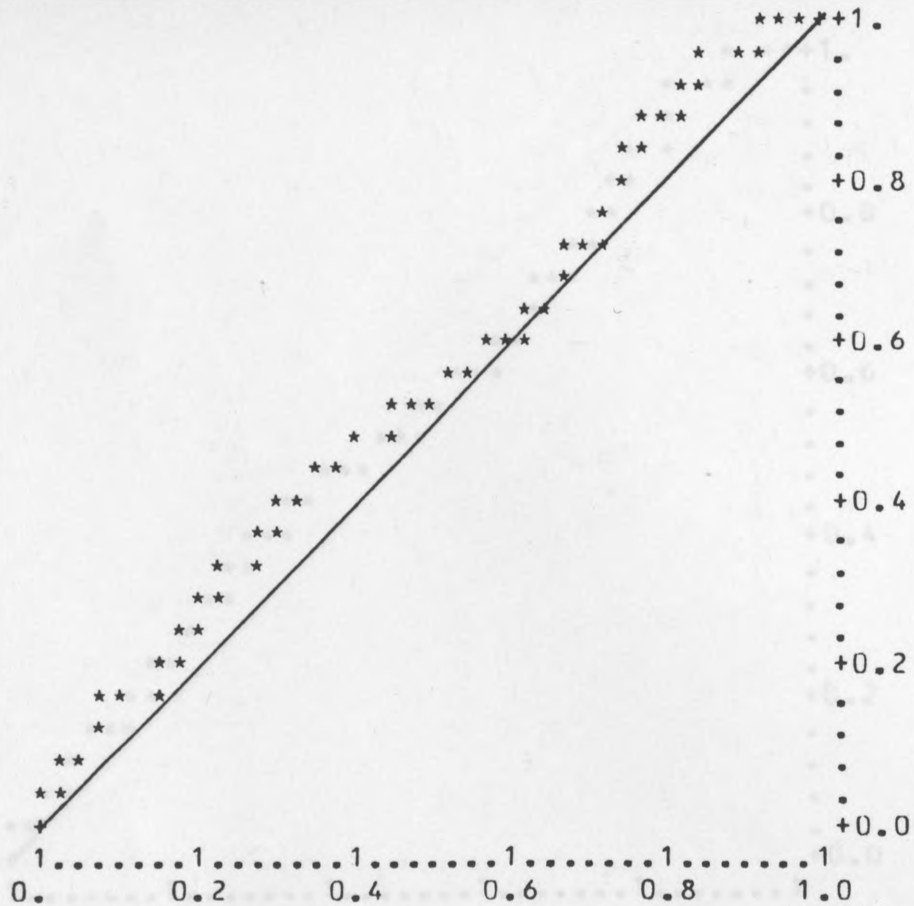


FIG.5.13 BL y-plot, data in Table 4.1, the plots based on the line printer output.

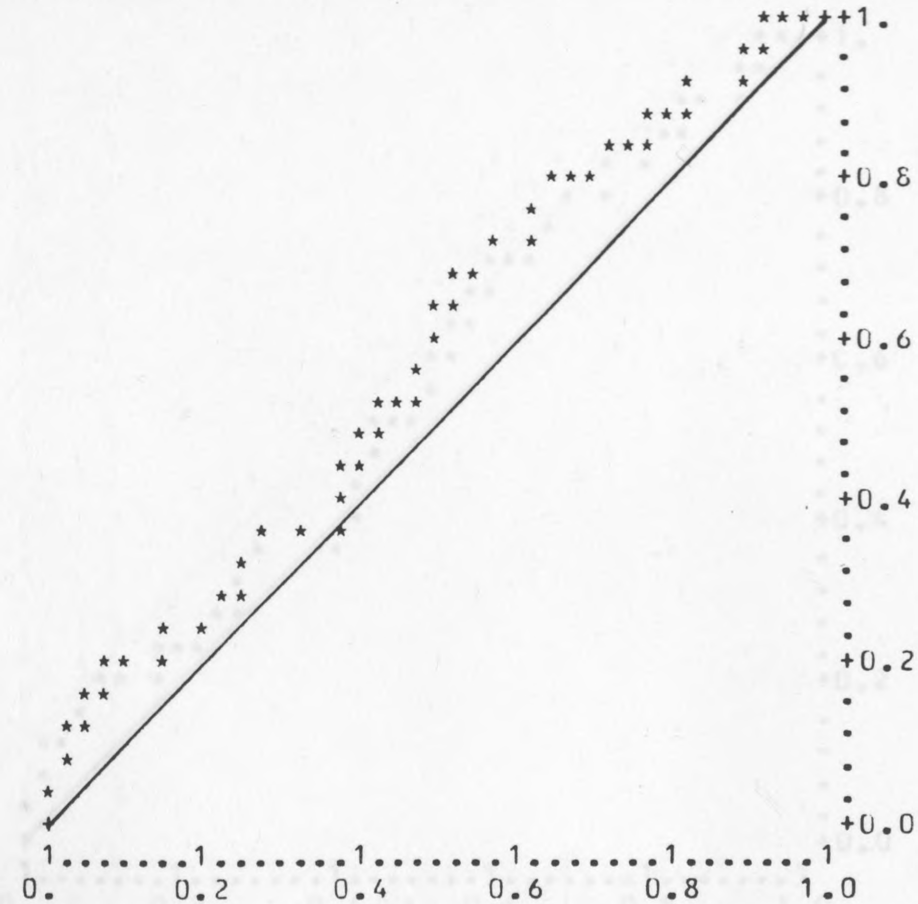


FIG.5.12. BL u-plots, data in Table 4.1., the plots based on the line printer output.

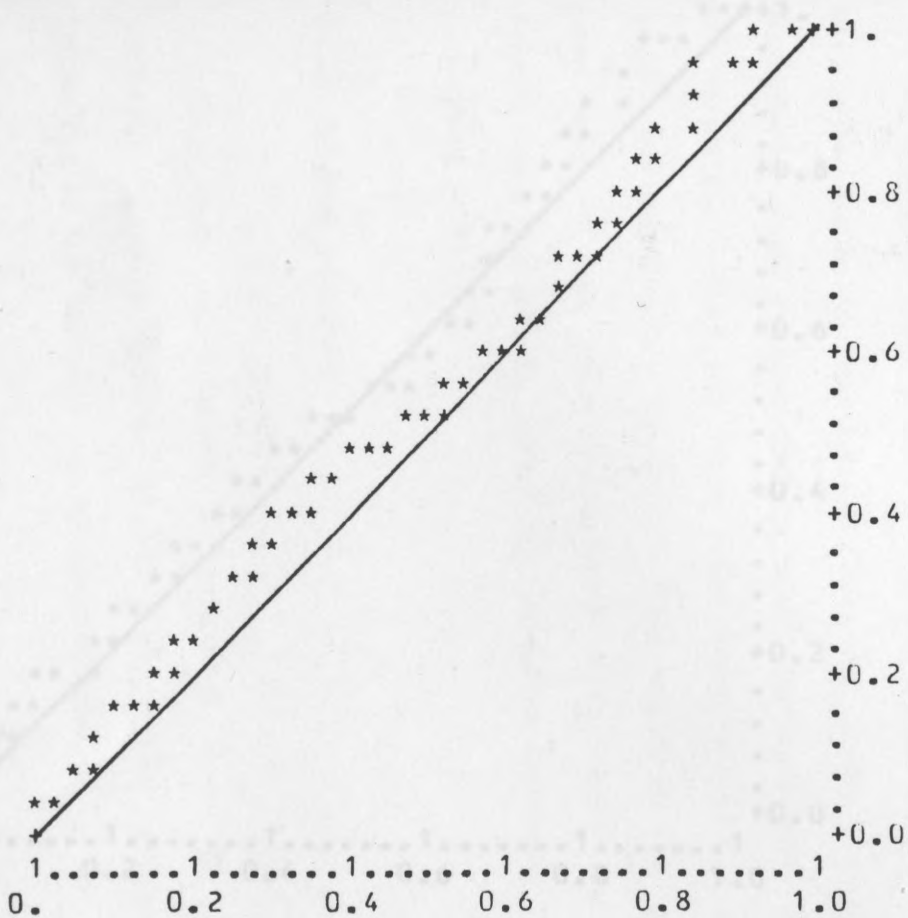


FIG.5.15. LNHPP y-plots, data in Table 4.1, the plots based on the line printer output.

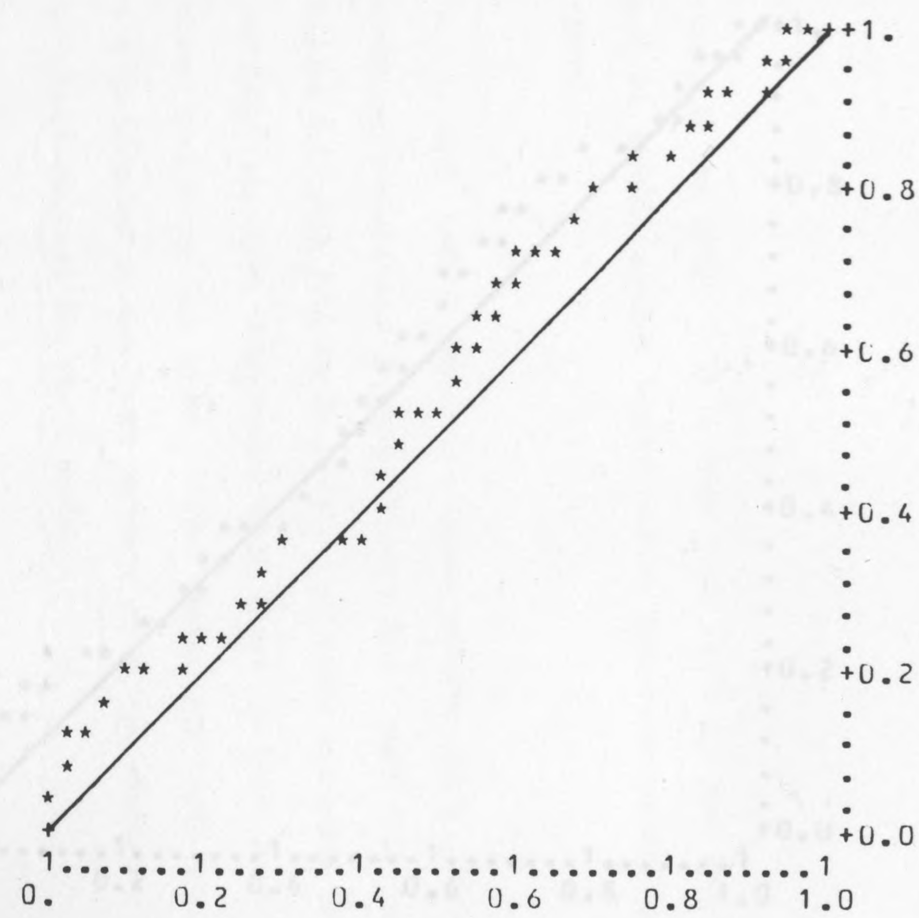


FIG.5.14. LNHPP u-plots, data in Table 4.1, the plots based on the line printer output.

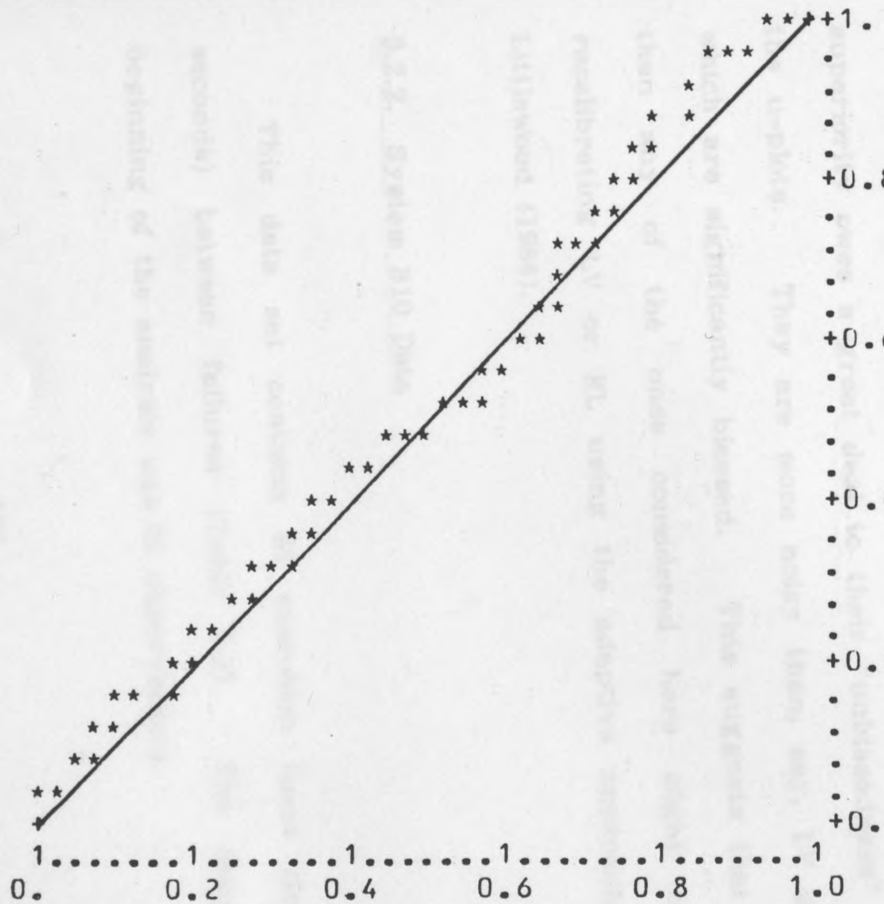


FIG.5.17. W y-plots, data in Table 4.1., the plots based on the line printer output.

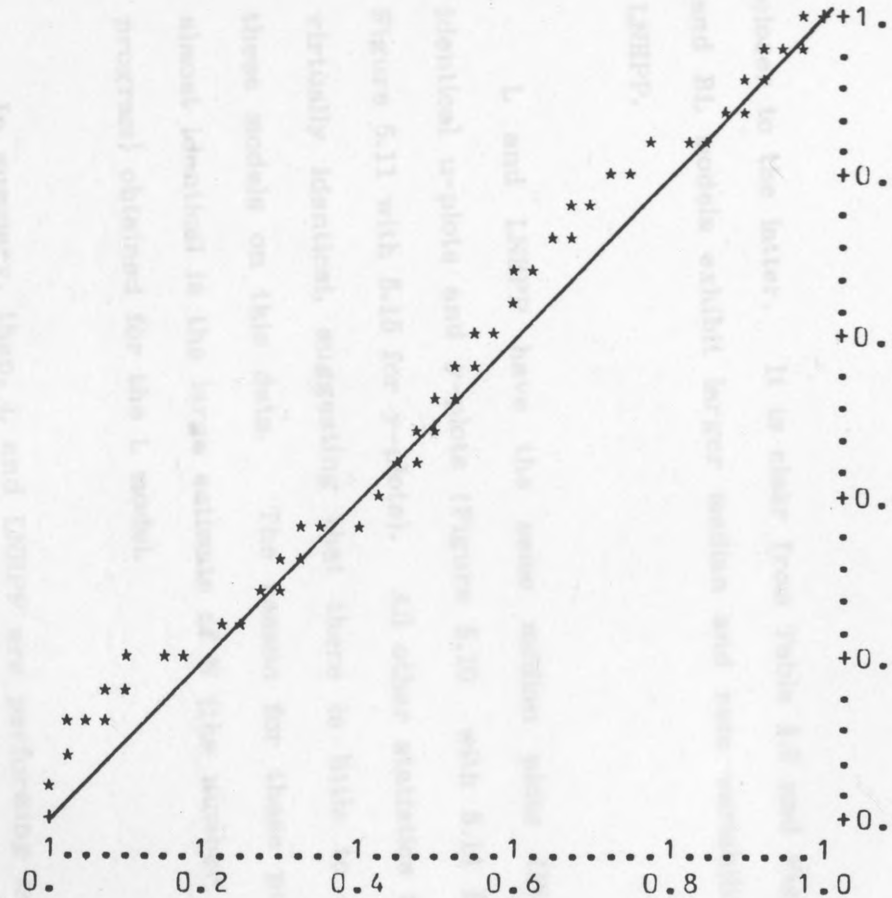


FIG.5.16. W u-plots, data in Table 4.1., the plots based on the line printer output.

closer to the latter. It is clear from Table 4.2 and Figure 5.1 that W and BL models exhibit larger median and rate variability than L and LNHPP.

L and LNHPP have the same median plots (Figure 5.1) and identical u-plots and y-plots (Figure 5.10 with 5.14 for u-plots and Figure 5.11 with 5.15 for y-plots). All other statistics in Table 4.2 are virtually identical, suggesting that there is little to choose between these models on this data. The reason for these predictions being almost identical is the large estimate of N (the number of faults in the program) obtained for the L model.

In summary, then, L and LNHPP are performing best on this data set, with little to distinguish between them. However, their superiority owes a great deal to their "unbiasedness" as revealed by the u-plots. They are more noisy than, say, LV and KL, both of which are significantly biassed. This suggests that a better model than any of the ones considered here might be obtained by recalibrating LV or KL using the adaptive approach of Keiller and Littlewood (1984).

5.2.2. System B10 Data

This data set contains 86 execution times (in hundredths of seconds) between failures (Table 5.2). The sample size at the beginning of the analysis was 30 observations.

TEST B DATA 1+2+3 FNS MIXED DATA (D-B10)

INTER-FAILURE TIMES

479.	266.	277.	554.	1034.
949.	693.	597.	117.	170.
117.	1274.	469.	1174.	693.
1908.	135.	277.	596.	757.
437.	2230.	437.	340.	405.
575.	277.	363.	522.	618.
277.	1300.	821.	213.	1620.
1601.	298.	874.	618.	2640.
5.	149.	1034.	2441.	460.
565.	1119.	437.	927.	4462.
714.	181.	1485.	757.	3154.
2115.	884.	2037.	1481.	559.
490.	593.	1769.	85.	2836.
213.	1866.	490.	1487.	4322.
1418.	1023.	5490.	1520.	3281.
2716.	2175.	3505.	725.	1963.
3979.	1090.	245.	1194.	994.
3902.				

TABLE 5.2. Execution time in hundredths of seconds between successive failures. Read left to right in rows.

Table 5.3. shows the summarised results of the various prediction systems on this data set.

Apart from JMNHPP, it is noticeable that all prediction systems give similar PL. All give non-significant y-plot distances. However, closer inspection suggest that there are differences in noise and bias.

Thus LV and KL have good PL performance but quite poor u-plot distances. The detailed plots, Figures 5.18 and 5.20, are below the line of unit slope, suggesting that these predictions tend to be too pessimistic. Thus, the median plots for LV and KL in Figure 5.22 are too low.

The median plots for JMNHPP and D are even more pessimistic than those of LV and KL. That they are too pessimistic is confirmed by examined of their u-plots (Figures 5.23 and 5.25). Each of these is, however, giving low noise values.

There is strong evidence, then, that the lowest five median plots (LV, KL, D, JMNHPP and LNHPP), Figure 5.22, are too low. The remaining median plots, however, show more noise.

It therefore seems to be the case that there is clear trade-off operating between noise and bias. This suggests that the more noisy predictions (JM, BJM, L, BL and W) are in fact too noisy.

Test Continuous Data System DATA B10

No.of Observations: = 86
Starting Sample Size : = 30

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	4.2304	8	3.8051	8	1.1104	8	.1206 N.S.	4	.1147 N.S.	10	6.9029 466.2226	6
BJM	3.7347	7	3.2738	7	1.0392	6	.1096 N.S.	2	.0771 N.S.	5	7.0444 466.6472	7
L	5.2873	9	4.6731	9	1.0374	5	.1298 N.S.	5	.0897 N.S.	8	6.8807 465.3674	1
BL	3.1012	6	2.7750	6	0.9768	4	.1197 N.S.	3	.0621 N.S.	4	7.0444 465.8074	4
LV	2.3276	4	2.1809	3	0.9675	2	.1677 10%	6	.0510 N.S.	2	6.9114 465.5275	2
KL	2.3052	3	2.2122	4	0.9660	1	.1690 10%	7	.0507 N.S.	1	6.9087 465.6919	3
D	1.9589	2	1.8458	2	1.0422	7	.2089 2%	9	.0520 N.S.	3	6.8819 467.7761	9
JMNHPP	1.0631	1	1.0284	1	1.2129	10	.2710 1%	10	.0846 N.S.	7	6.8807 473.8700	10
LNHPP	3.0507	5	2.7602	5	0.9692	3	.1692 10%	8	.0817 N.S.	6	6.8807 465.8452	5
W	6.1607	10	5.4777	10	1.1748	9	.1002 N.S.	1	.1107 N.S.	9	6.9128 466.8940	8

TABLE 5.3. The analysis results of data in Table 5.2.
The ML routine did not terminate normally for L, LV, and KL.
The BL routine shows overflo.

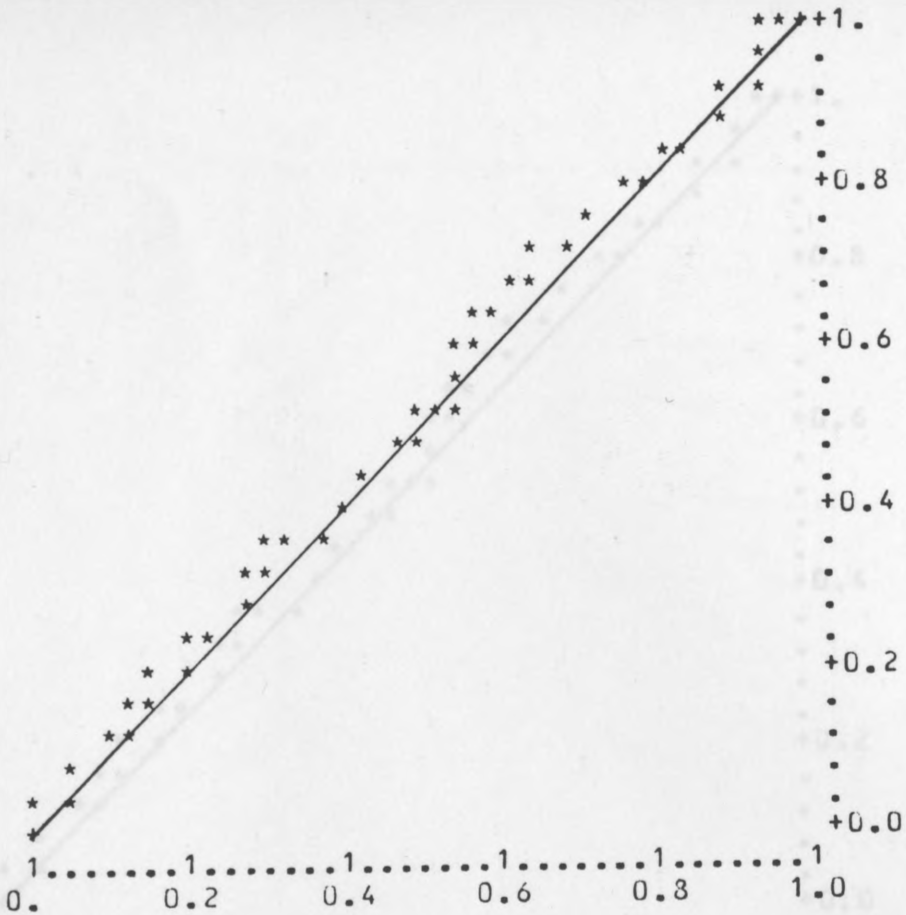


FIG.5.19. LV y-plots, data in Table 5.2, the plots based on the line printer output.

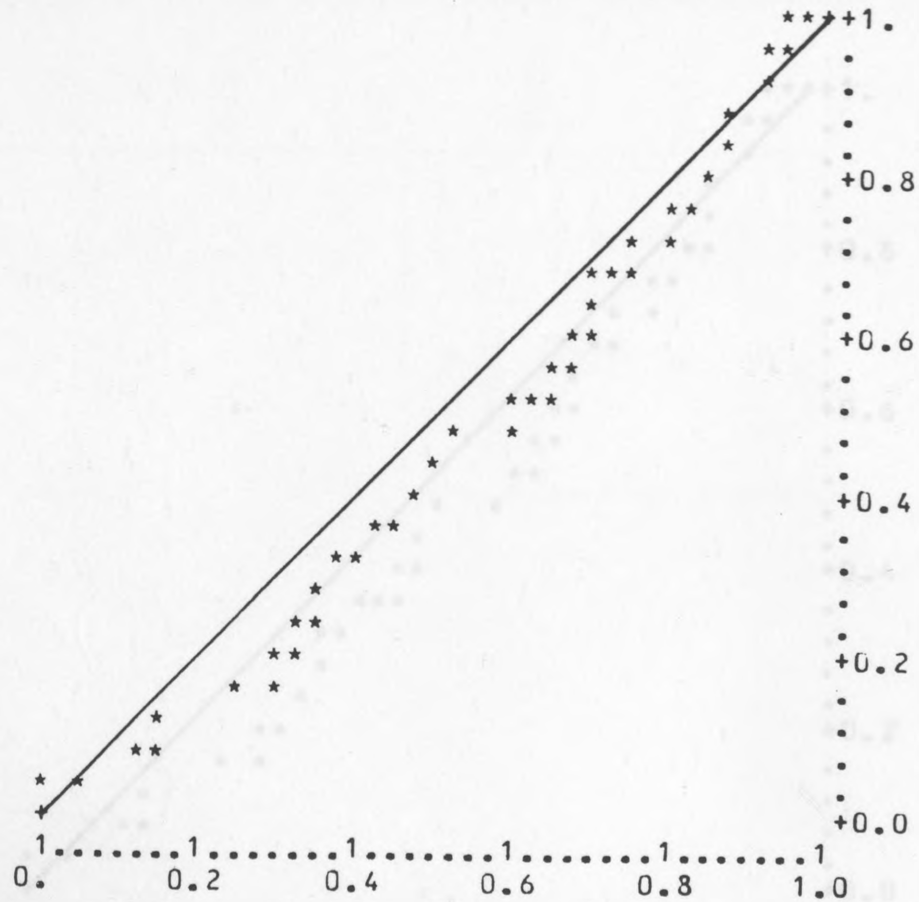


FIG.5.18. LV u-plots, data in Table 5.2, the plots based on the line printer output.

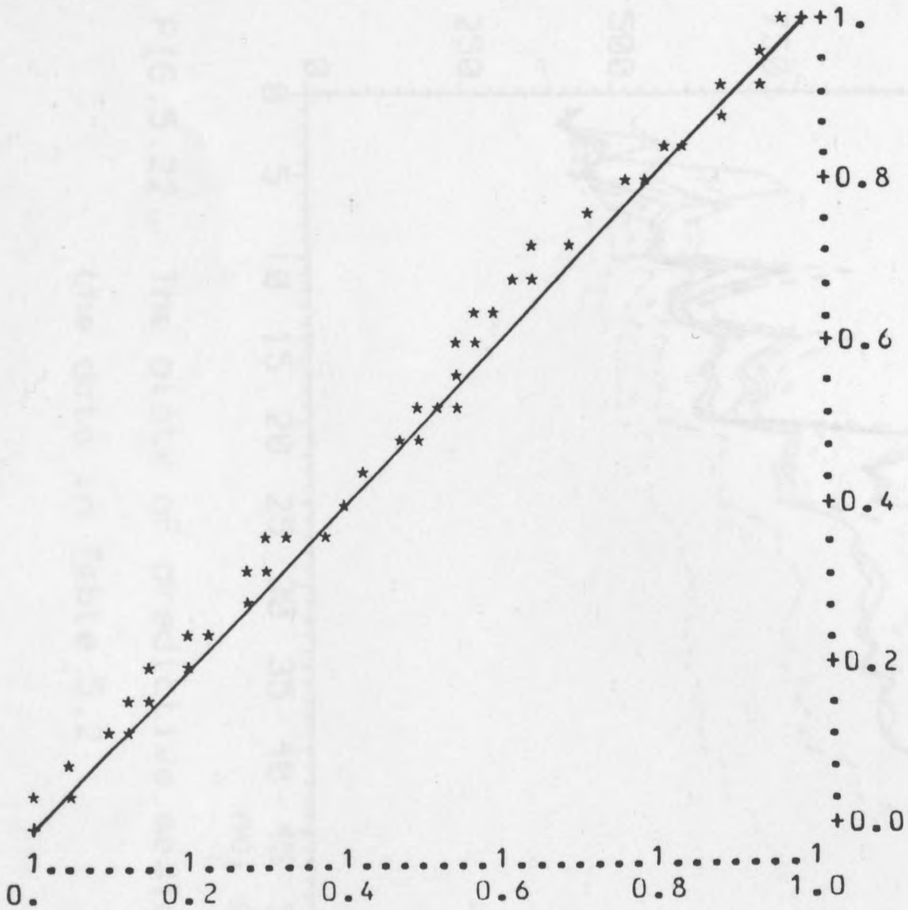


FIG.5.21. KL y-plots, data in Table 5.2, the plots based on the line printer output.

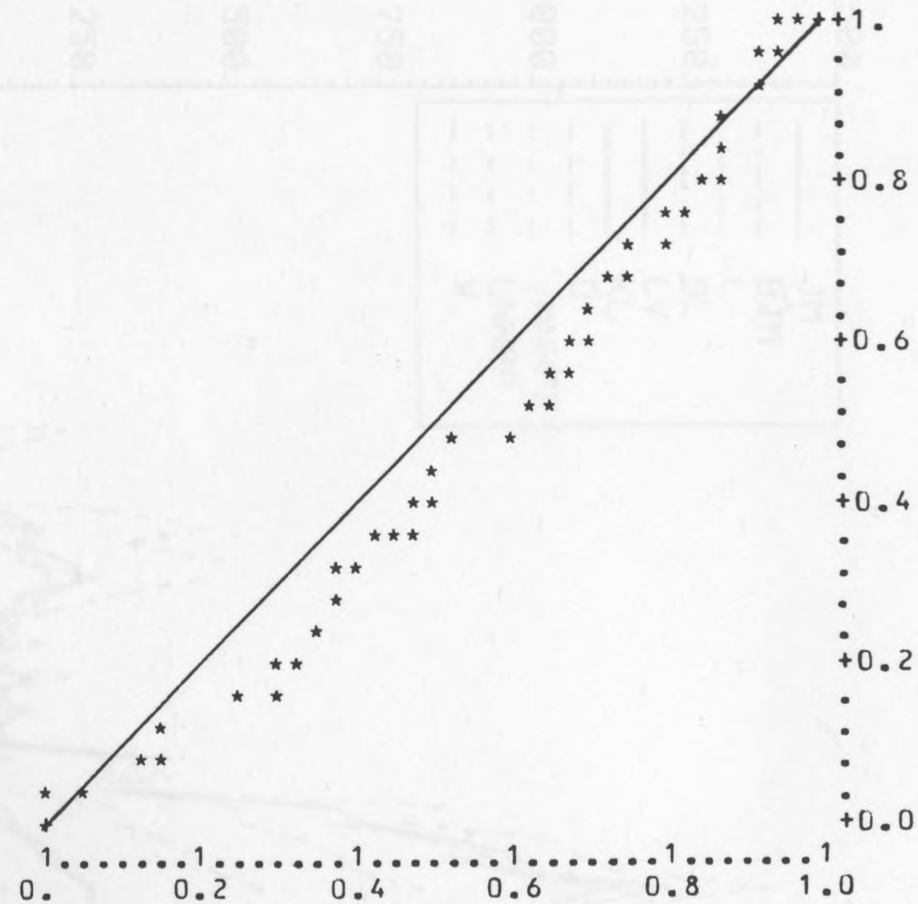


FIG.5.20. KL u-plots, data in Table 5.2, the plots based on the line printer output.

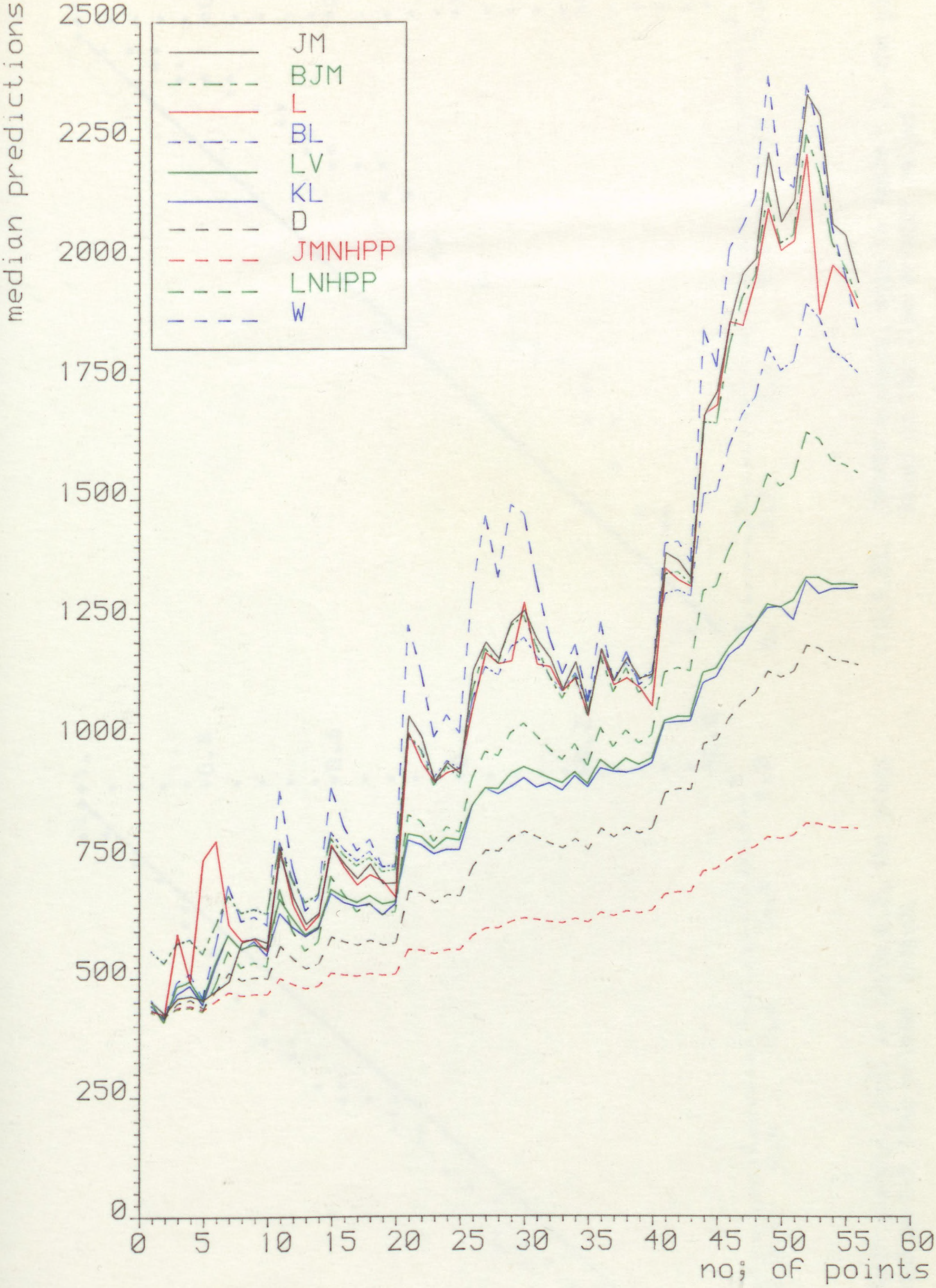


FIG.5.22. The plots of predictive medians for the data in Table 5.2

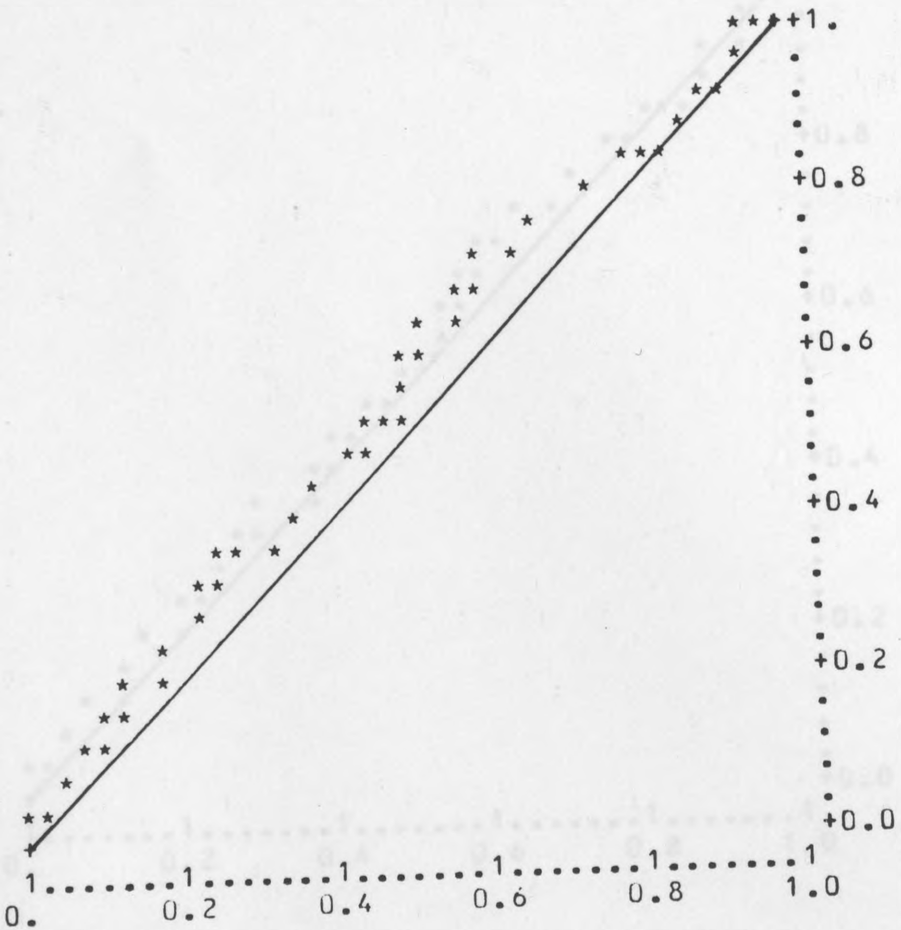


FIG.5.24. JMNHPP y-plots, data in Table 5.2, the plots based on the line printer output.

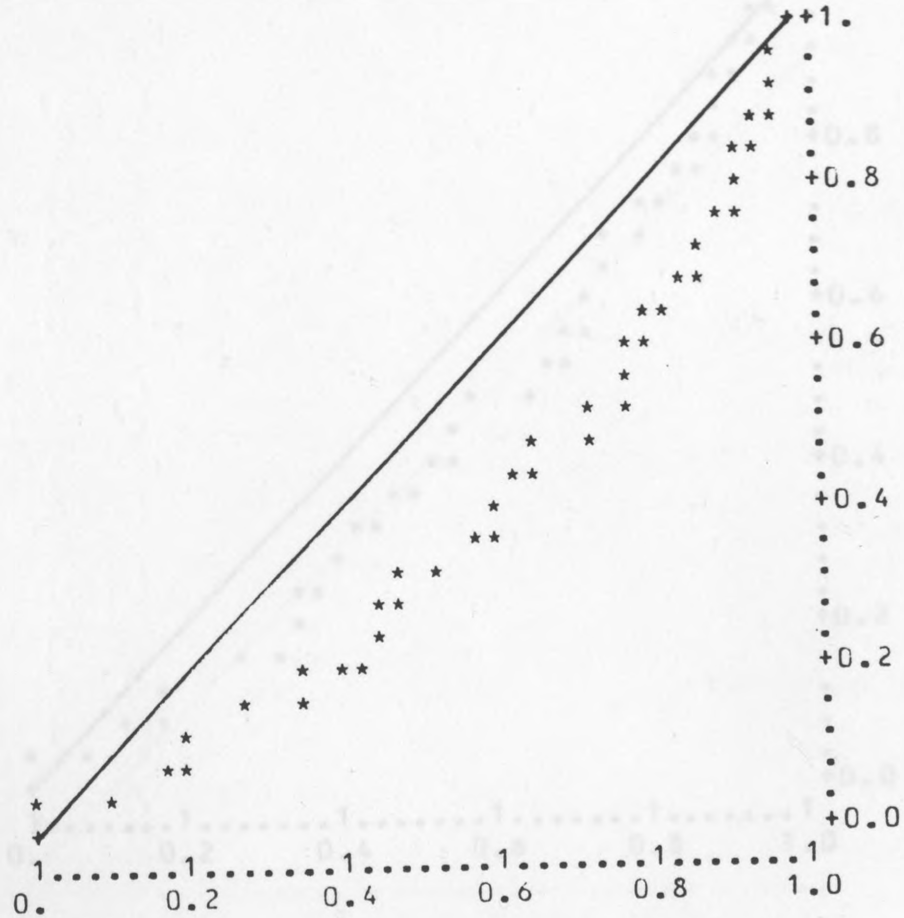


FIG.5.23. JMNHPP u-plots, data in Table 5.2, the plots based on the line printer output.

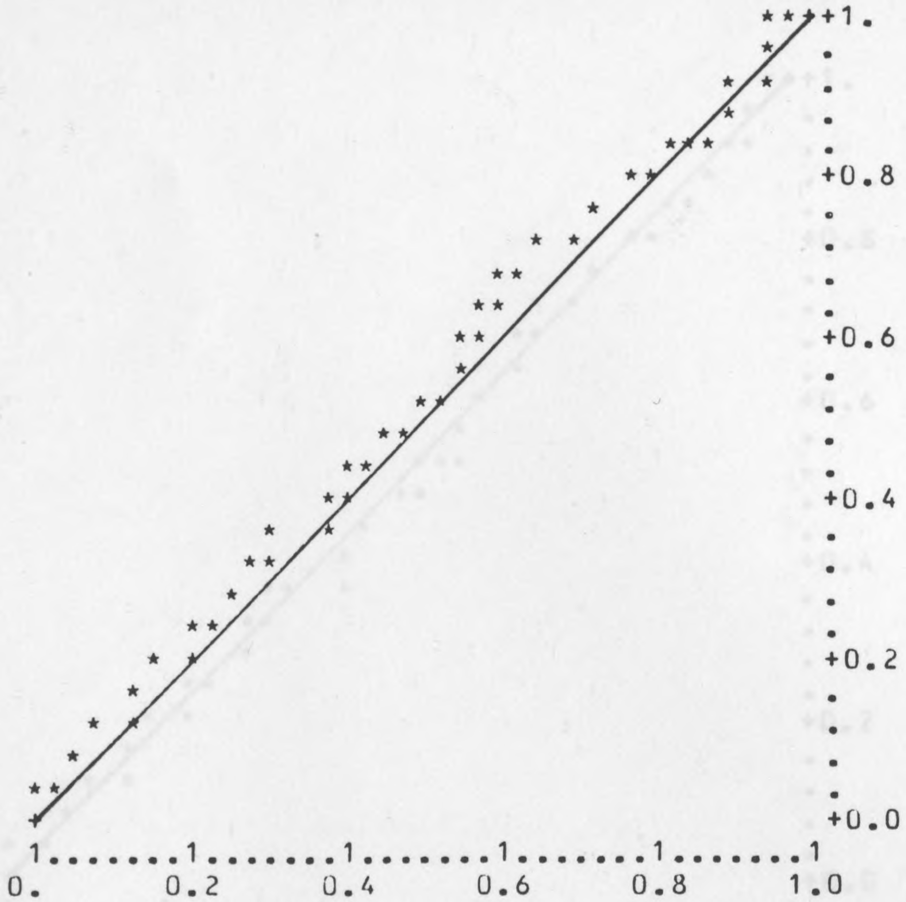


FIG.5.26. D y-plots, data in Table 5.2, the plots based on the line printer output.

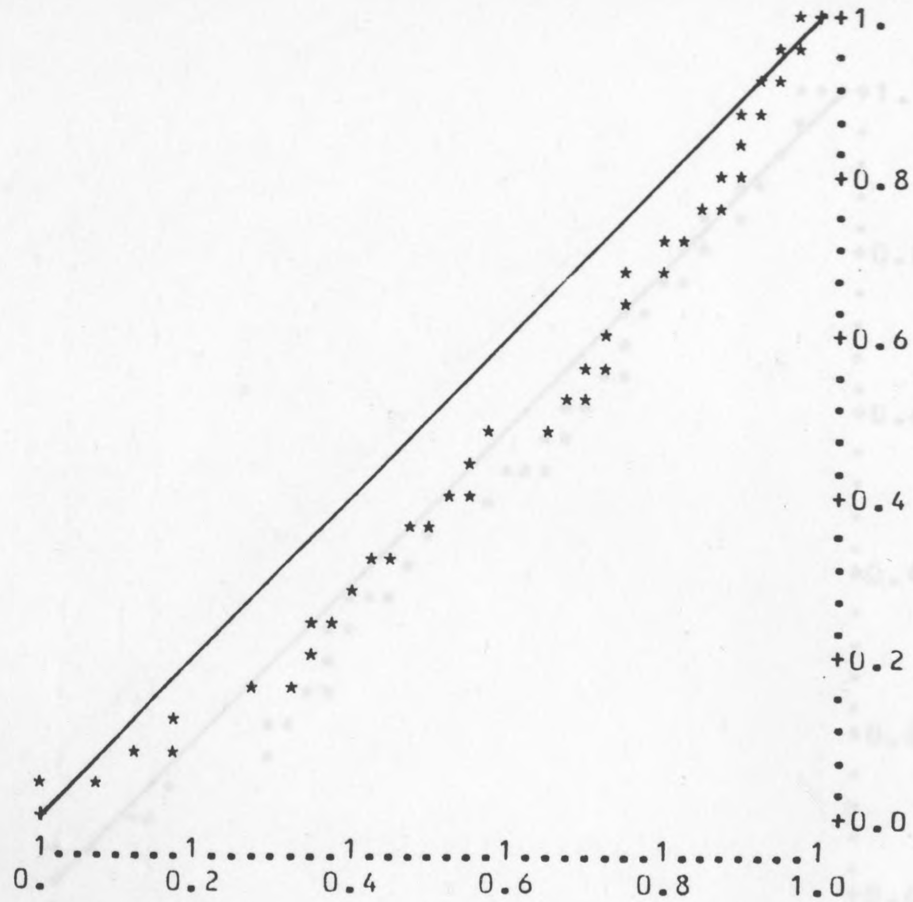


FIG.5.25. D u-plots, data in Table 5.2, the plots based on the line printer output.

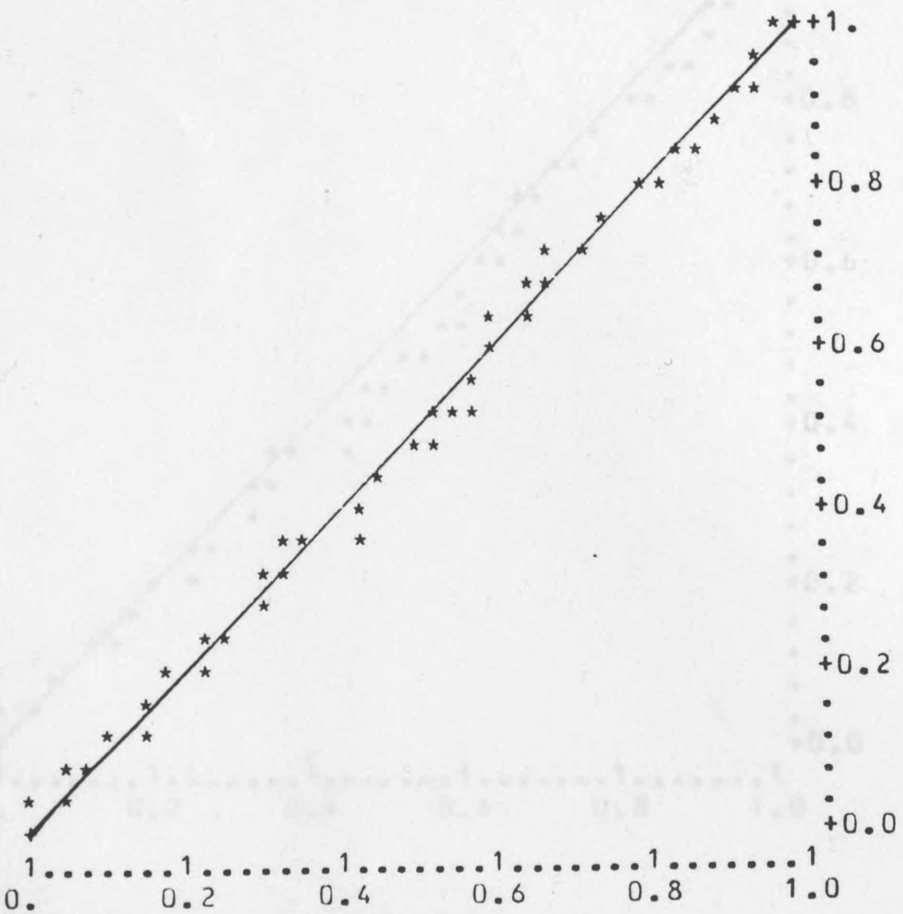


FIG.5.28. LNHPP y-plots, data in Table 5.2, the plots based on the line printer output.

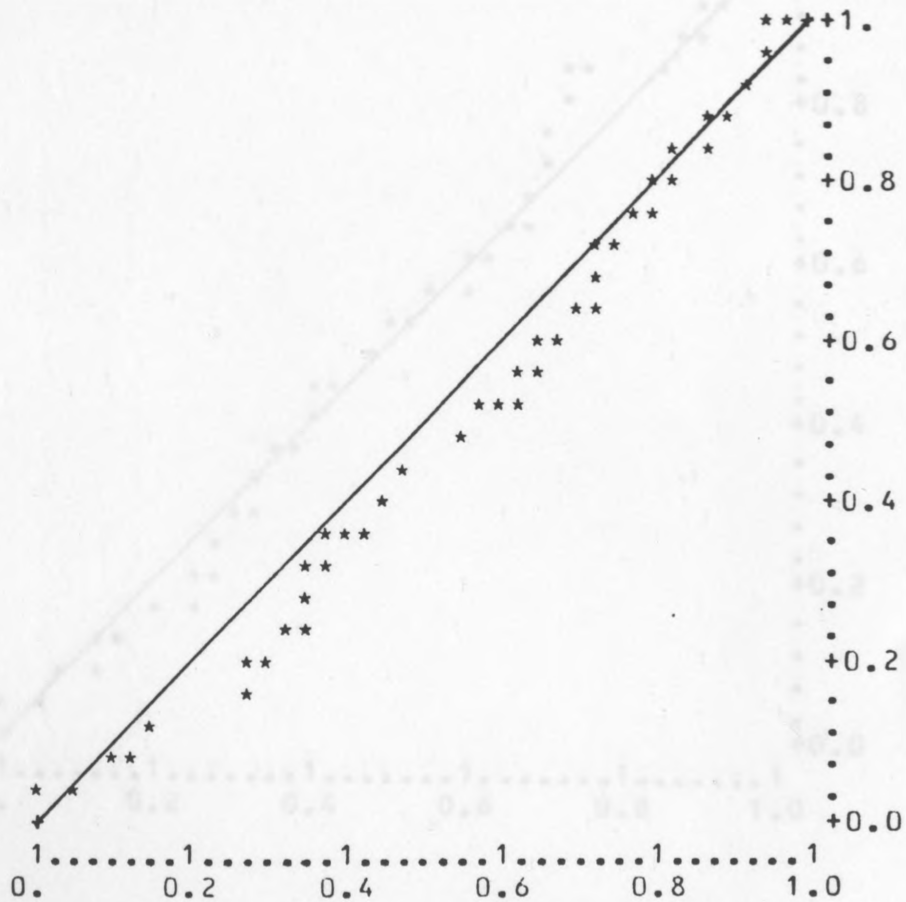


FIG.5.27. LNHPP u-plots, data in Table 5.2, the plots based on the line printer output.

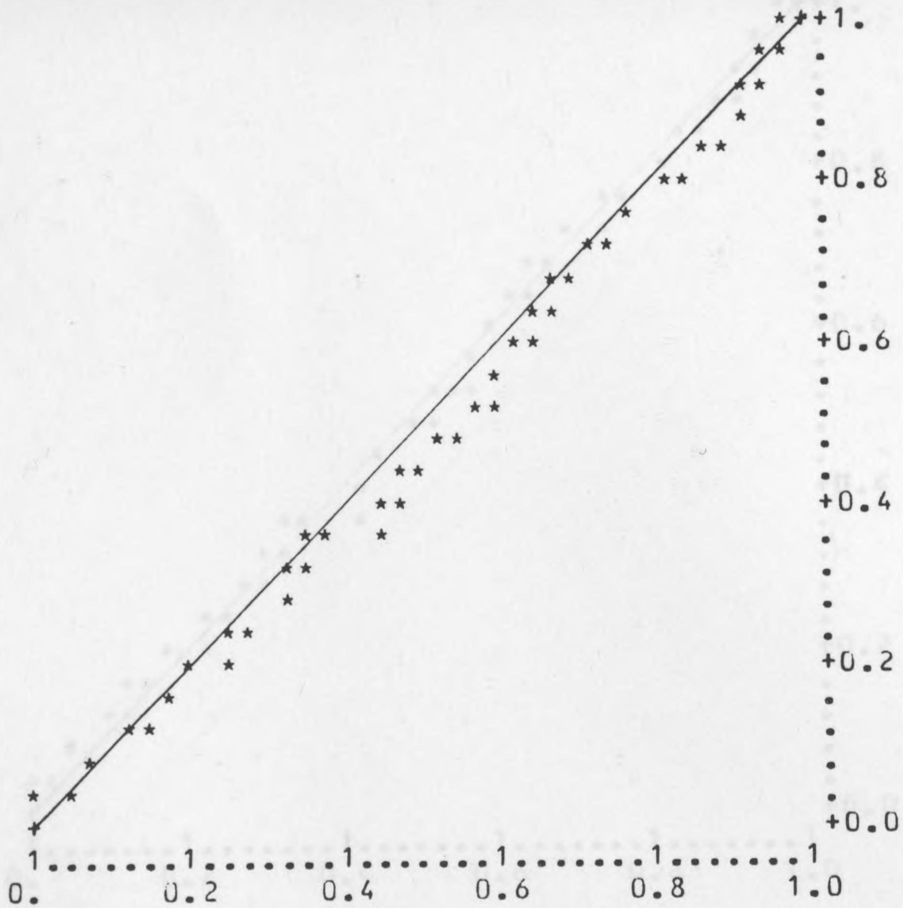


FIG.5.30 JM y-plots, data in Table 5.2, the plots based on the line printer output.

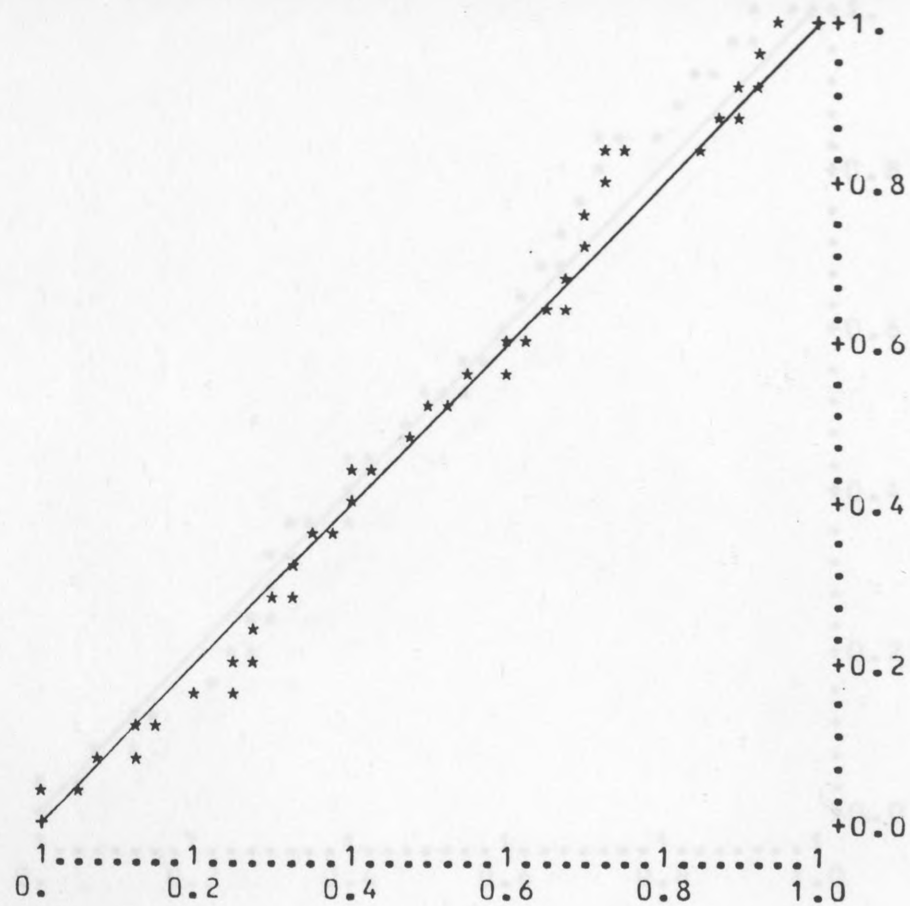


FIG. 5.29. JM u-plots, data in Table 5.2, the plots based on the line printer output.

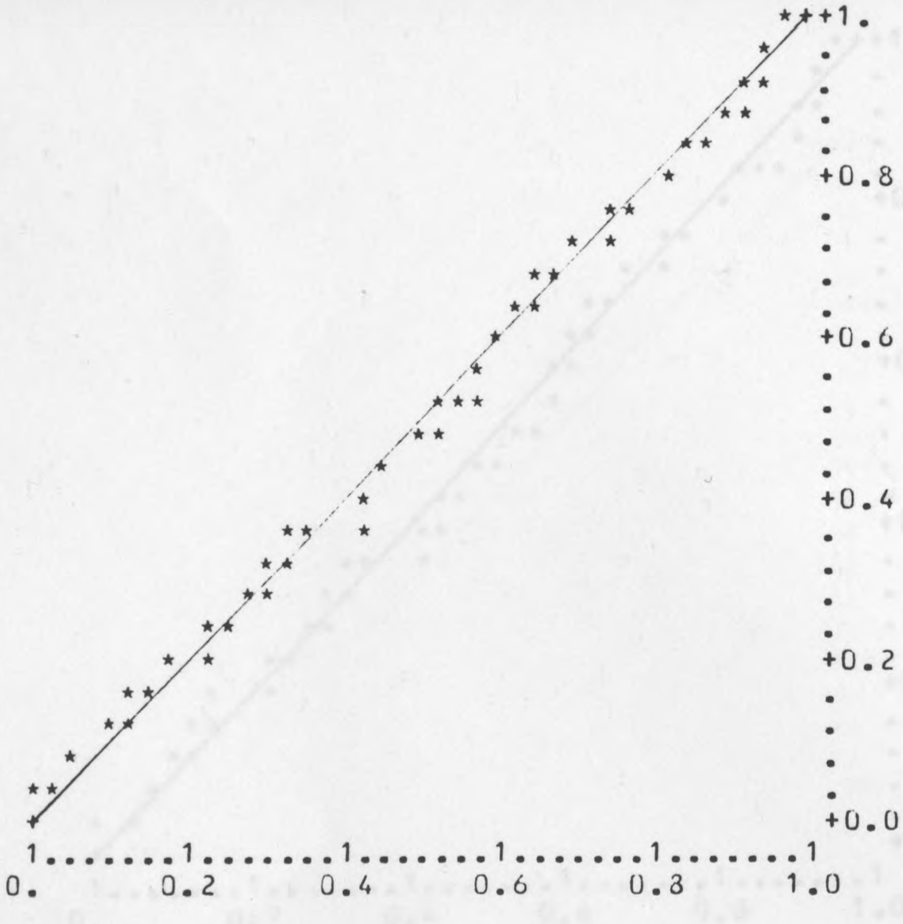


FIG.5.32. BJM y-plots, data in Table 5.2, the plots based on the line printer output.

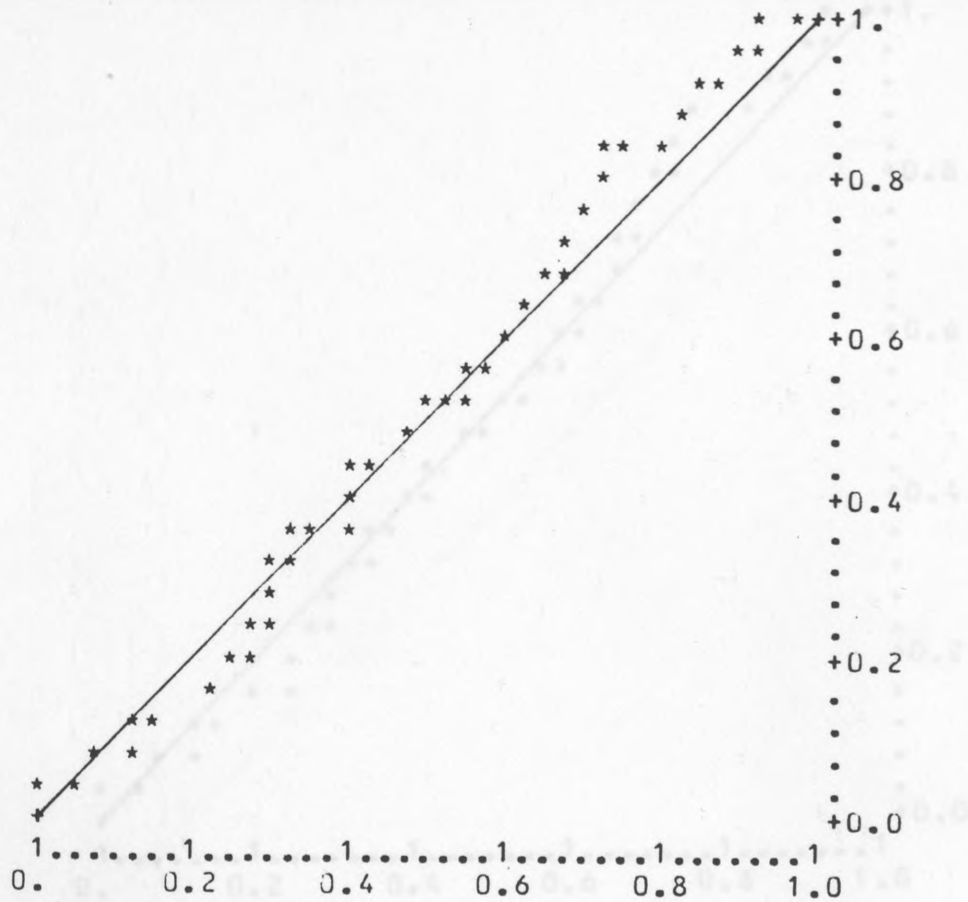


FIG. 5.31. BJM u-plots, data in Table 5.2, the plots based on the line printer output.

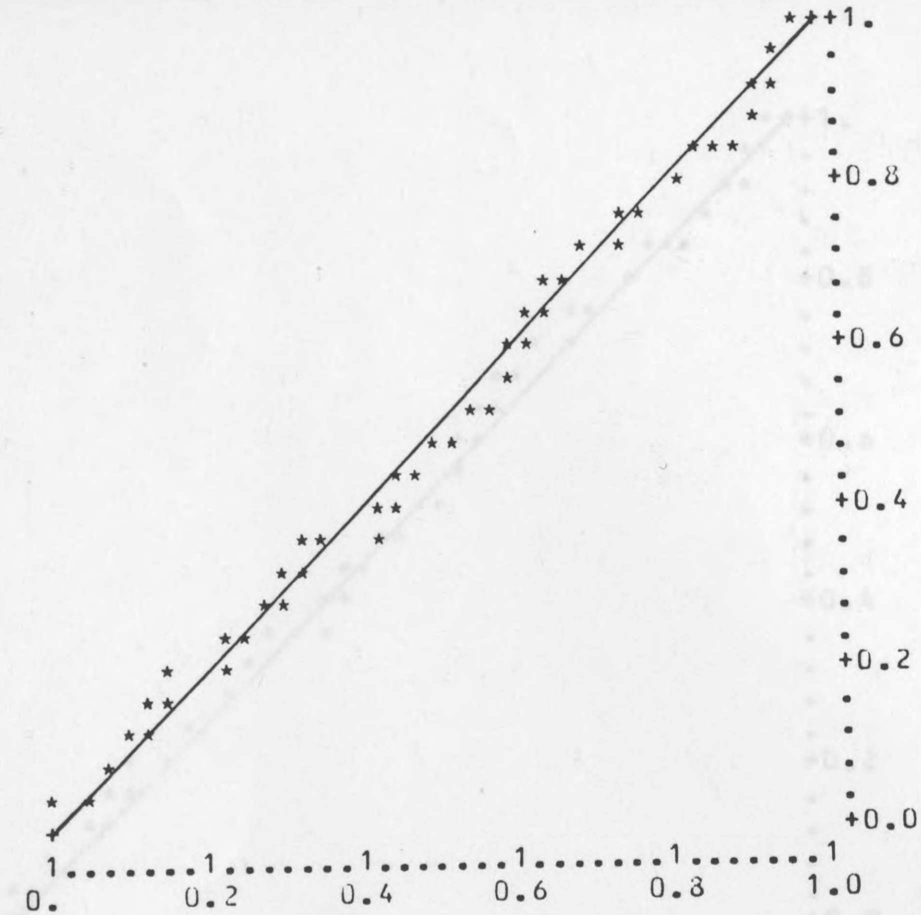


FIG.5.34. L y-plots, data in Table 5.2, the plots based on the line printer output.

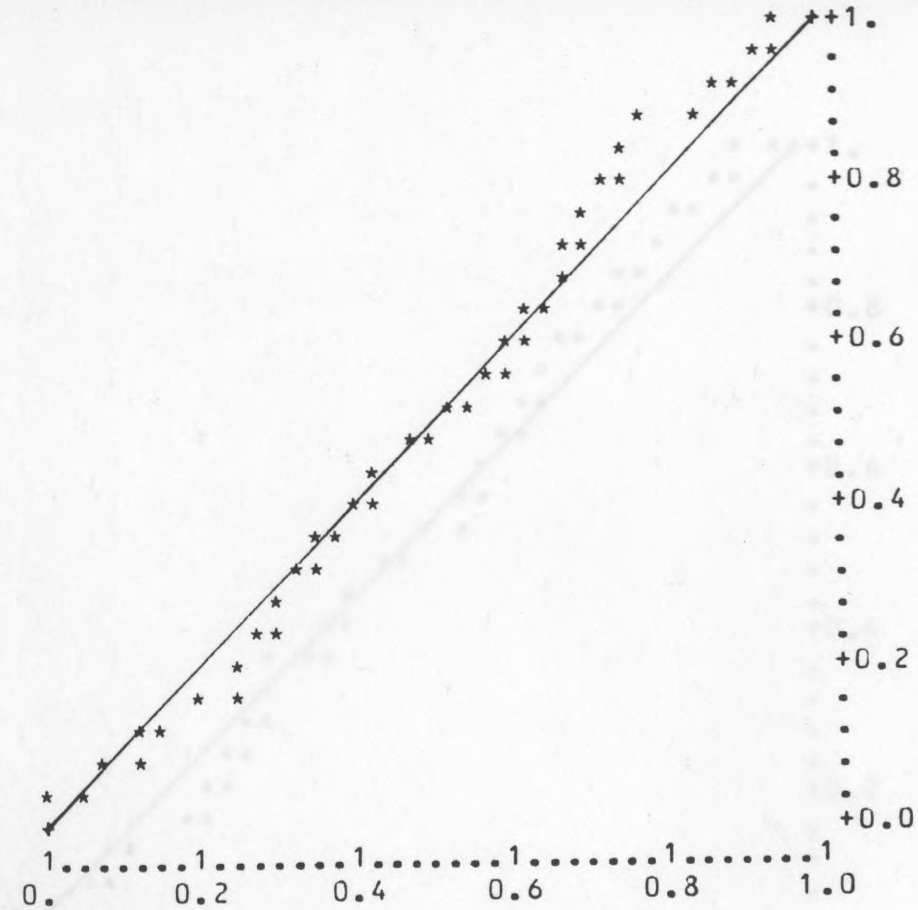


FIG.5.33. L u-plots, data in Table 5.2, the plots based on the line printer output.

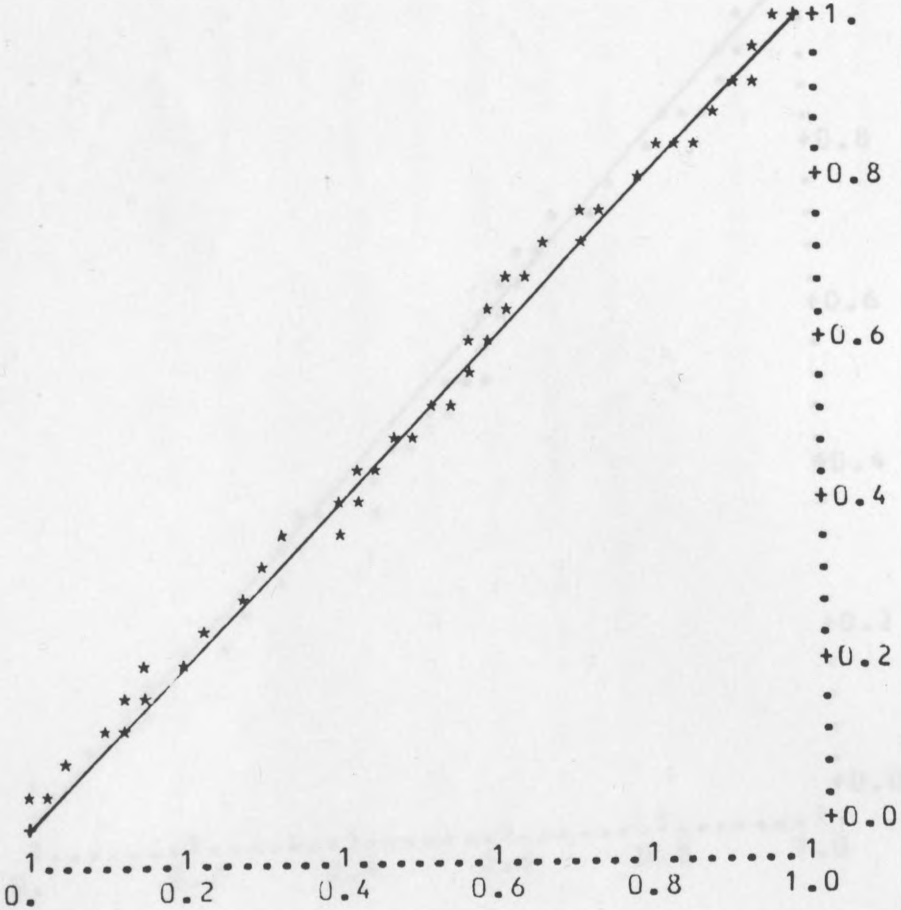


FIG.5.36. BL y-plots, data in Table 5.2, the plots based on the line printer output.

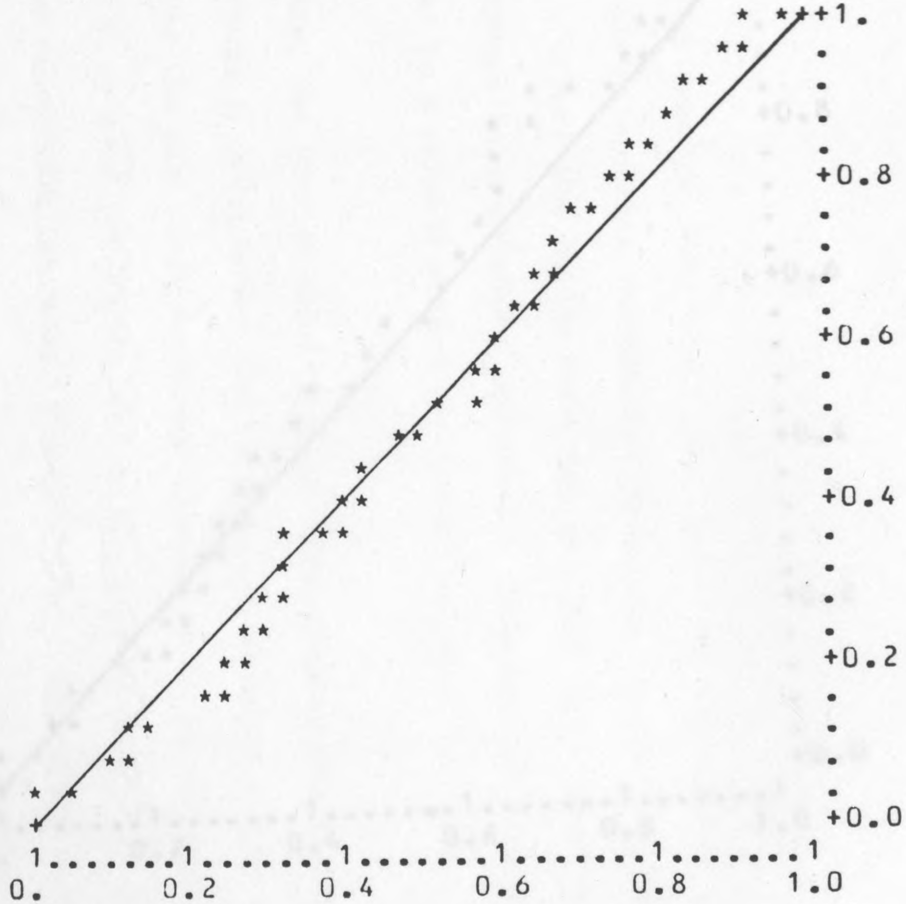


FIG.5.35. BL u-plots, data in Table 5.2, the plots based on the line printer output.

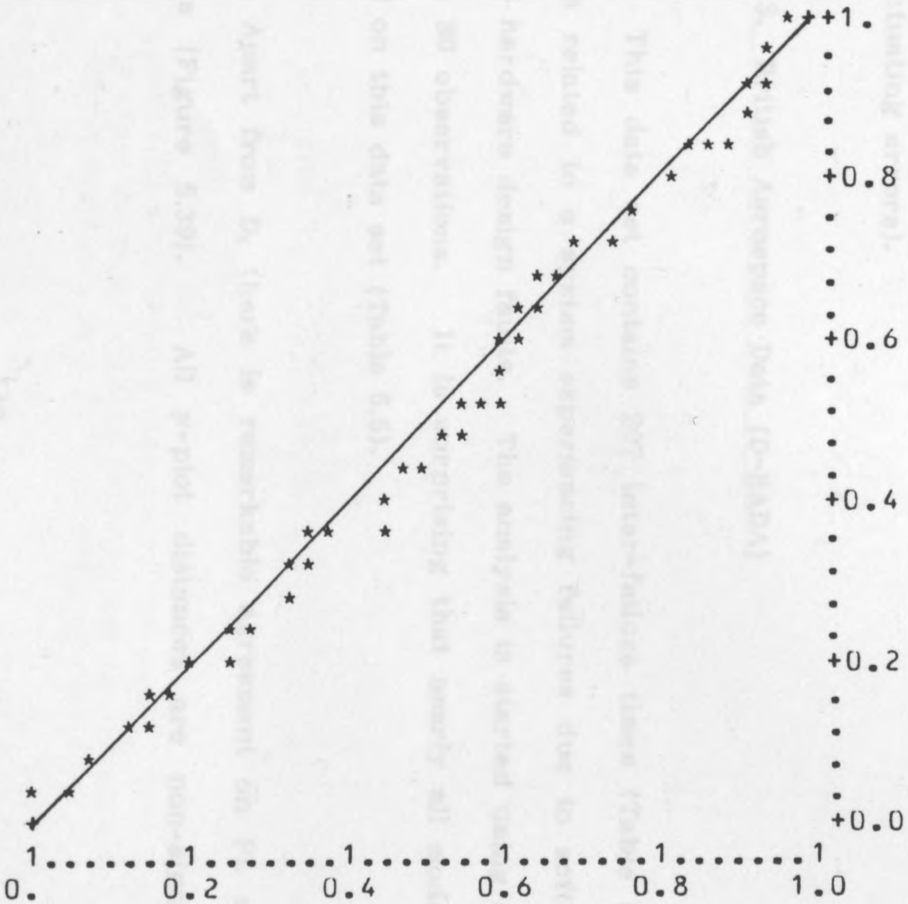


FIG.5.38. W y-plots, data in Table 5.2, the plots based on the line printer output.

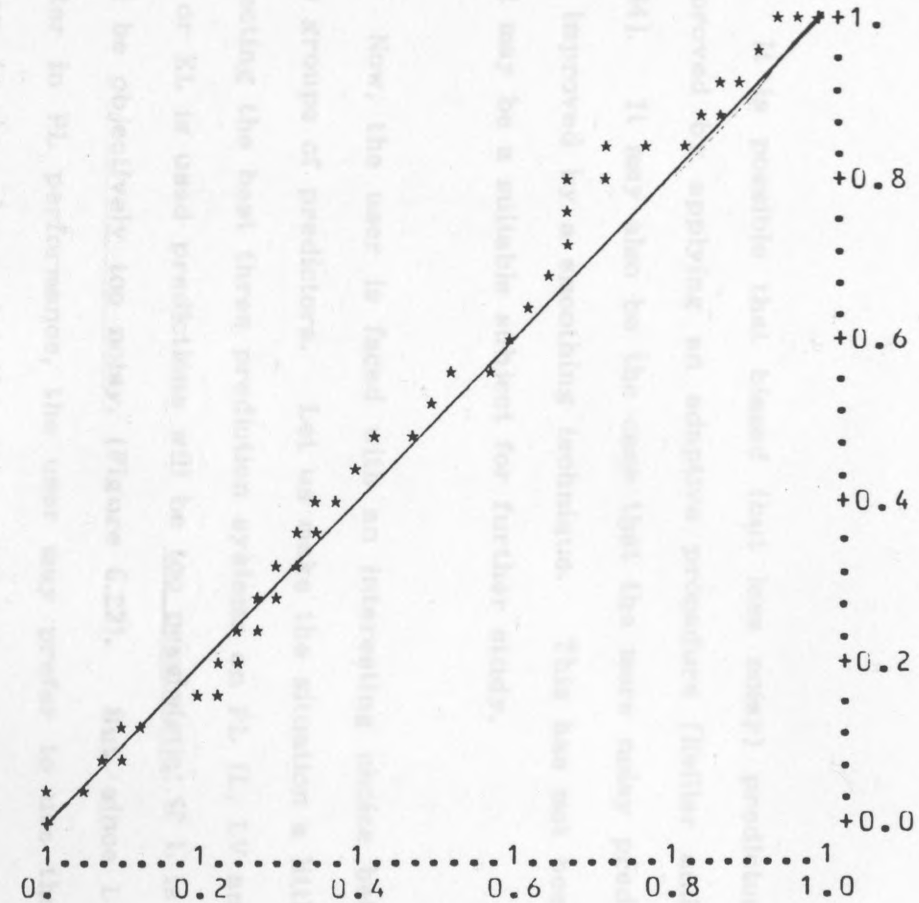


FIG. 5.37. W u-plots, data in Table 5.2, the plots based on the line printer output.

It is possible that biased (but less noisy) predictors could be improved by applying an adaptive procedure [Keiller and Littlewood, 1984]. It may also be the case that the more noisy predictors could be improved by a smoothing technique. This has not been attempted but may be a suitable subject for further study.

Now, the user is faced with an interesting choice between these two groups of predictors. Let us make the situation a little easier by selecting the best three prediction systems on PL (L, LV and KL). If LV or KL is used predictions will be too pessimistic; if L is used, they will be objectively too noisy, (Figure 5.22). But, since L is slightly better in PL performance, the user may prefer to use that model in order to be closer to the true reliability on average (but with fluctuating errors).

5.2.3. British Aerospace Data (D-BADA)

This data set contains 207 inter-failure times (Table 5.4). The data related to a system experiencing failures due to software faults and hardware design faults. The analysis is started using a sample of size 80 observations. It is surprising that nearly all models perform well on this data set (Table 5.5).

Apart from D, there is remarkable agreement on PL and median plots (Figure 5.39). All y-plot distances are non-significant, all

BRITISH AEROSPACE DATA

INTER-FAILURE TIMES

39.	10.	4.	36.	4.
5.	4.	91.	49.	1.
25.	1.	4.	30.	42.
9.	49.	44.	32.	3.
78.	1.	30.	205.	5.
129.	103.	224.	186.	53.
14.	9.	2.	10.	1.
34.	170.	129.	4.	4.
35.	5.	5.	22.	36.
35.	121.	23.	33.	48.
32.	21.	4.	23.	9.
13.	165.	14.	22.	41.
12.	138.	95.	49.	62.
2.	35.	89.	90.	69.
22.	15.	19.	42.	14.
11.	41.	210.	16.	30.
37.	66.	9.	16.	14.
24.	12.	159.	89.	118.
29.	21.	18.	2.	114.
37.	46.	17.	1.	150.
332.	160.	66.	206.	9.
26.	62.	239.	13.	4.
85.	85.	240.	178.	34.
102.	9.	146.	59.	48.
25.	25.	111.	5.	31.
51.	6.	193.	27.	25.
96.	26.	30.	30.	17.
320.	73.	39.	13.	13.
19.	128.	34.	84.	40.
177.	349.	274.	32.	58.
31.	114.	39.	88.	84.
232.	108.	36.	36.	7.
22.	30.	239.	3.	39.
63.	152.	63.	30.	245.
196.	46.	152.	102.	9.
228.	220.	208.	73.	3.
33.	6.	212.	91.	3.
10.	172.	21.	173.	371.
40.	48.	126.	90.	149.
30.	317.	500.	673.	432.
66.	168.	66.	66.	120.
49.	332.			

TABLE 5.4. Operating time between successive failures. This data relates to a system experienceinq failures due to software and hardware desiqn faults. Read left to riqht in rows.

Test Continuous Data System D-BADA

No.of Observations: = 207
Starting Sample Size : = 80

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	4.77464	10	4.51571	10	.94099	6	.0833 N.S.	6	.0723 N.S.	8	4.68239 711.03607	6
BJM	4.20910	6	3.92470	6	.91588	1	.0835 N.S.	7	.0712 N.S.	7	4.71441 710.90424	3
L	4.68990	9	4.47386	9	.92432	2	.0703 N.S.	2	.0648 N.S.	4	4.68096 710.944	4
BL	3.41688	5	3.13034	4	.93298	3	.0741 N.S.	4	.0547 N.S.	1	4.71441 710.43318	1
LV	2.64111	2	2.74784	2	.94939	8	.0874 N.S.	8	.0652 N.S.	5	4.76011 711.83459	8
KL	2.75691	3	2.88578	3	.95333	9	.1020 20%	9	.0664 N.S.	6	4.76750 712.58476	9
D	1.98090	1	1.92384	1	.99347	10	.1168 10%	10	.0736 N.S.	9	4.67301 714.49132	10
JMNHPP	4.58110	7	4.29811	7	.93700	4	.0752 N.S.	5	.0750 N.S.	10	4.67015 711.02128	5
LNHPP	3.29779	4	3.14030	5	.93897	5	.0717 N.S.	3	.0583 N.S.	3	4.67625 710.89510	2
W	4.67634	8	4.43484	8	.94514	7	.0636 N.S.	1	.0564 N.S.	2	4.67301 411.25569	7

TABLE 5.5. Analysis of data in Table 5.4.
ML terminated abnormally for many L predictions.
BJM and BL calculations involved overflo and underflo on some predictions.

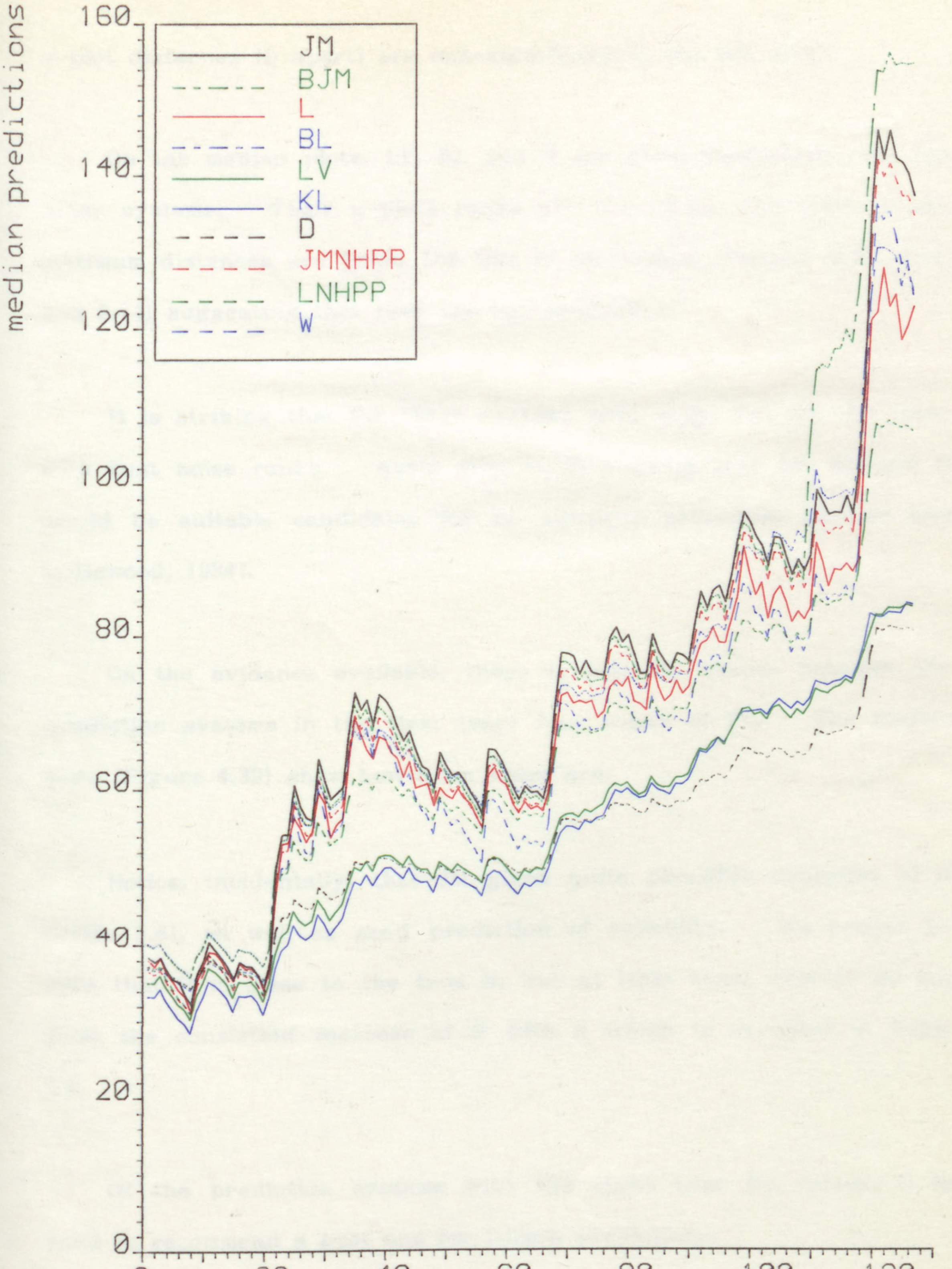


FIG.5.39. The plots of predictive medians for the data in Table 5.4

u-plot distances (D apart) are non-significant at the 10% level.

On the median plots, LV, KL and D are more pessimistic than the other systems. Their u-plots ranks are the worst, and these u-plot maximum distances are <u>below</u> the line of unit slope (Figures 5.40, 5.42, and 5.44) suggesting that they are <u>too</u> pessimistic.

It is striking that the three systems with <u>worst</u> PL, are the three with best noise ranks. Again this might suggest that LV, KL and D would be suitable candidates for an adaptive procedure [Keiller and Littlewood, 1984].

On the evidence available, there is little to choose between the prediction systems in the first (say) four ranks of PL. The median plots (Figure 4.39) show how close these are.

Notice, incidentally, that JM gives quite plausible estimates of N (Table 5.6), as well as good prediction of reliablity. We cannot be sure that $\hat{N}$ is close to the true N, but at least these results do not show the consistent increase of $\hat{N}$ with n which is revealed in Table 5.1.

Of the prediction systems with the eight best PL values, it is hard to recommend a best one for future predictions.

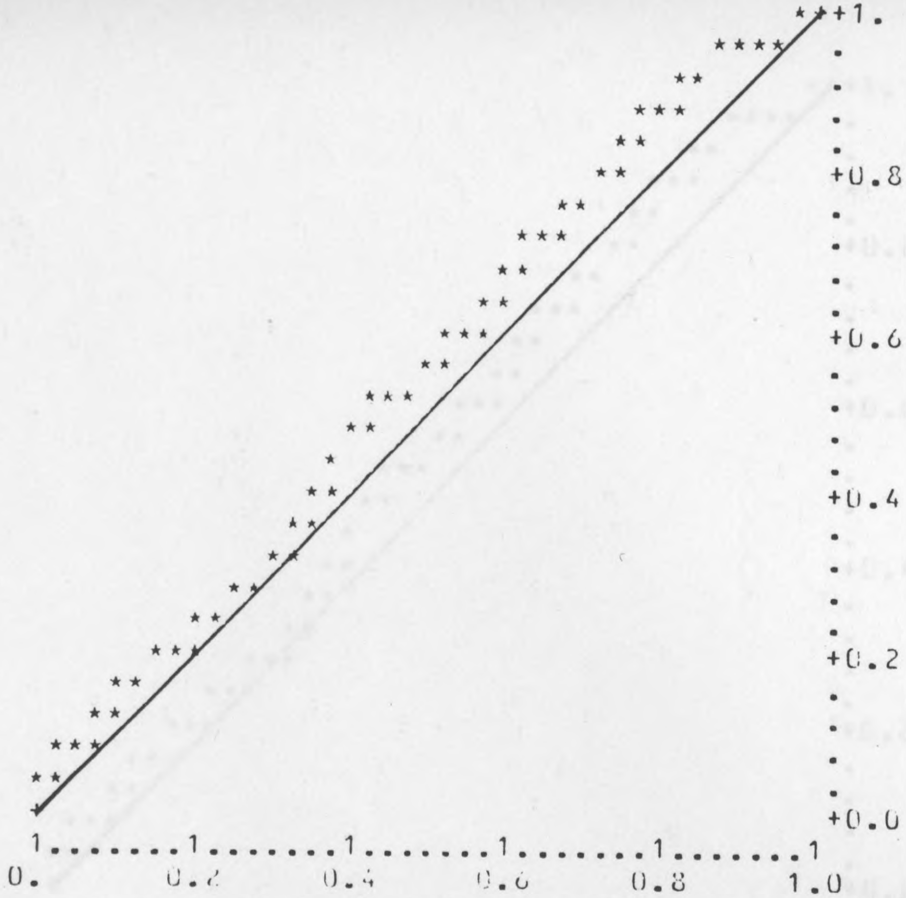


FIG. 5.41. LV y-plots, data in Table 5.4., the plots based on the line printer output.

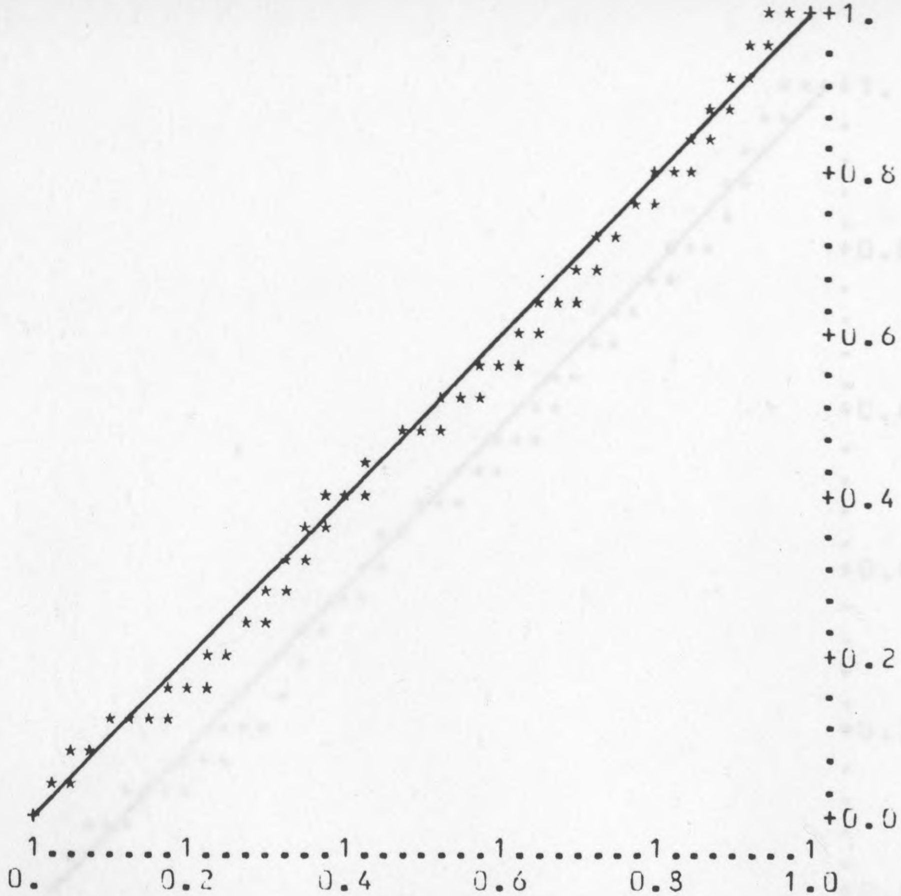


FIG.5.40. LV u-plots, data in Table 5.4, the plots based on the line printer output.

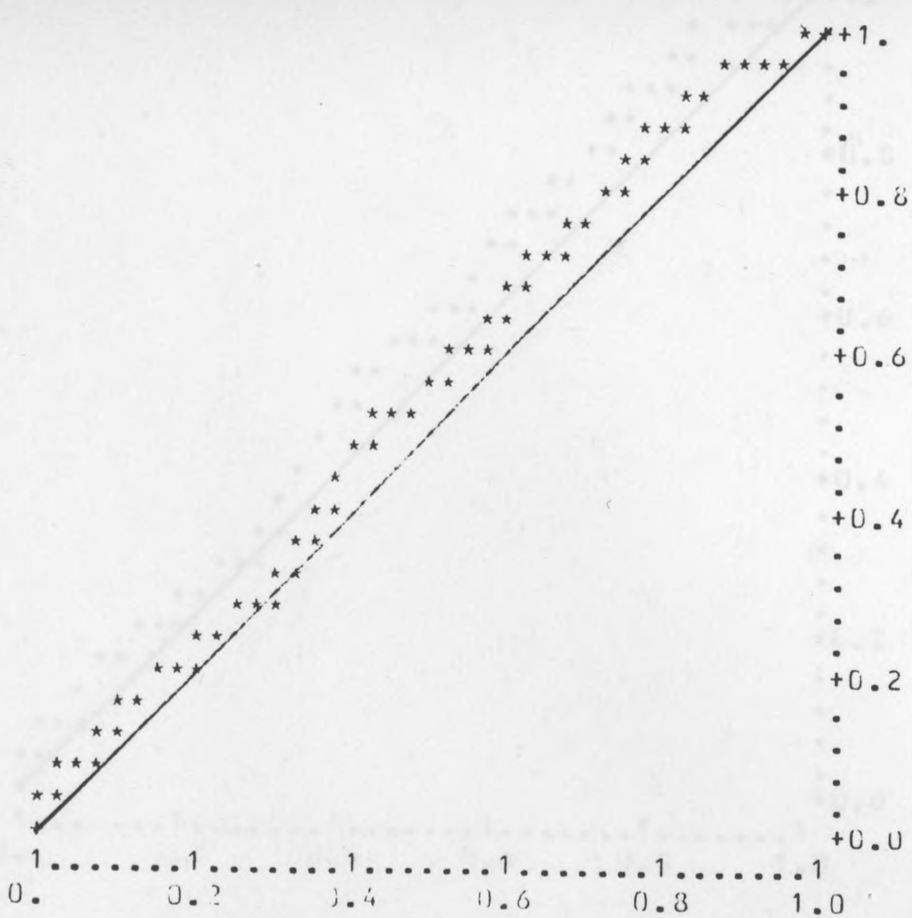


FIG.5.43. KL y-plots, data in Table 5.4, the plots based on the line printer output.

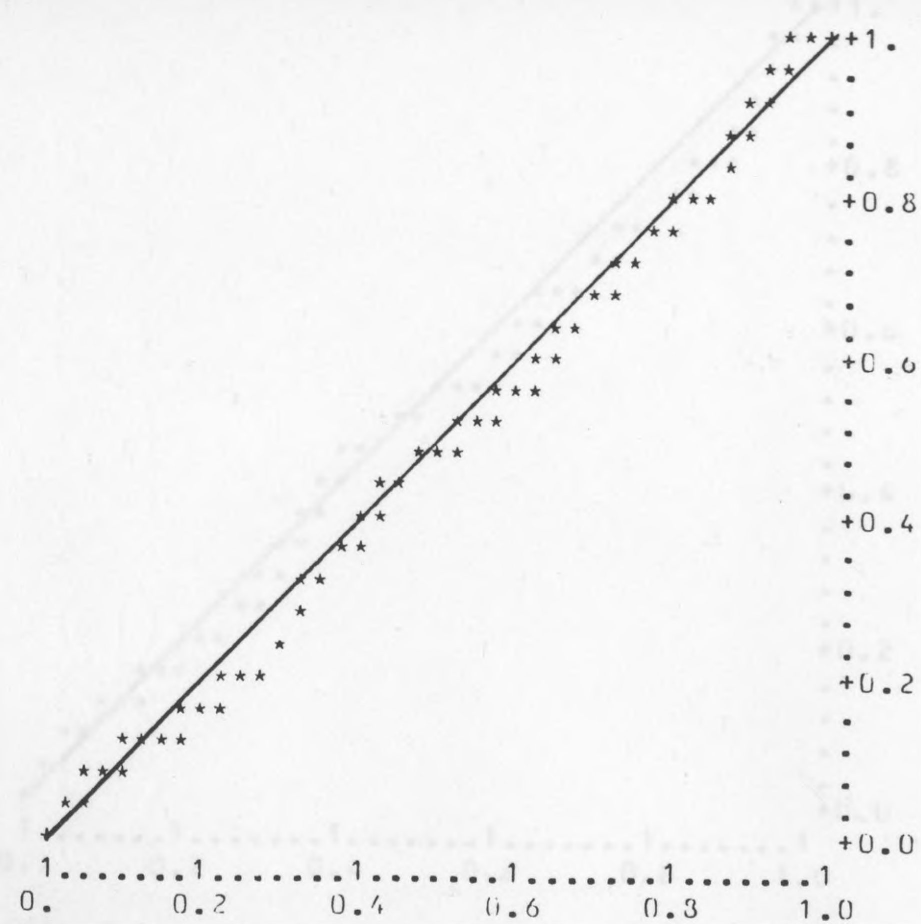
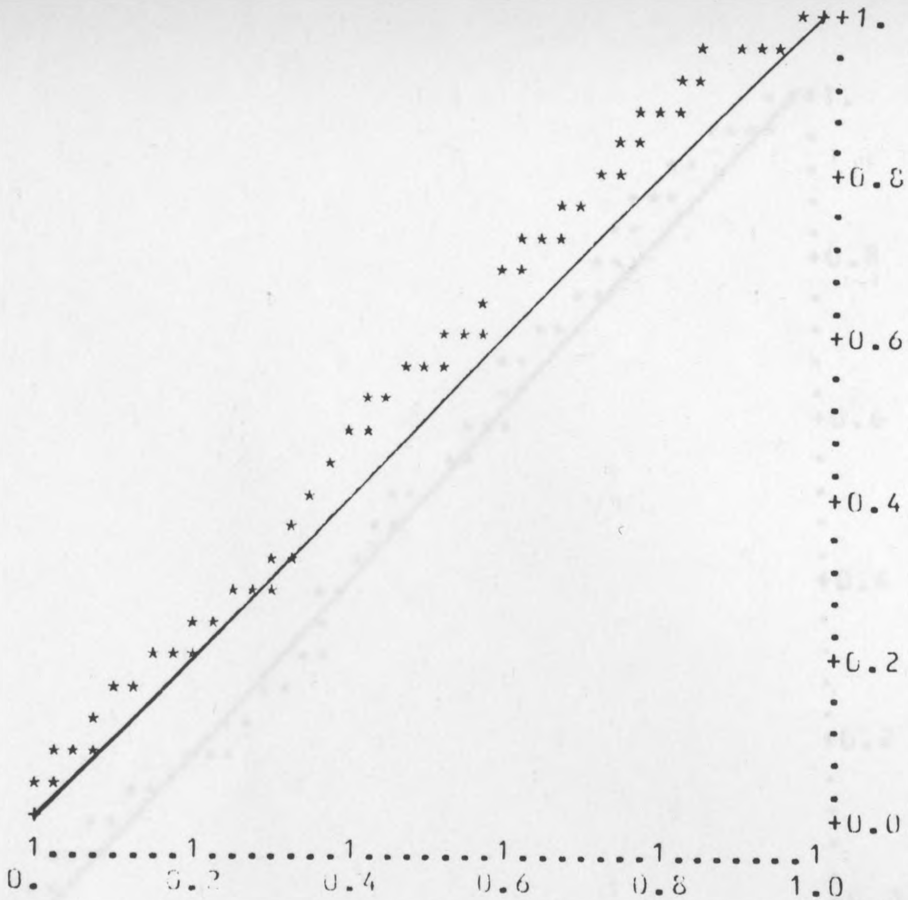
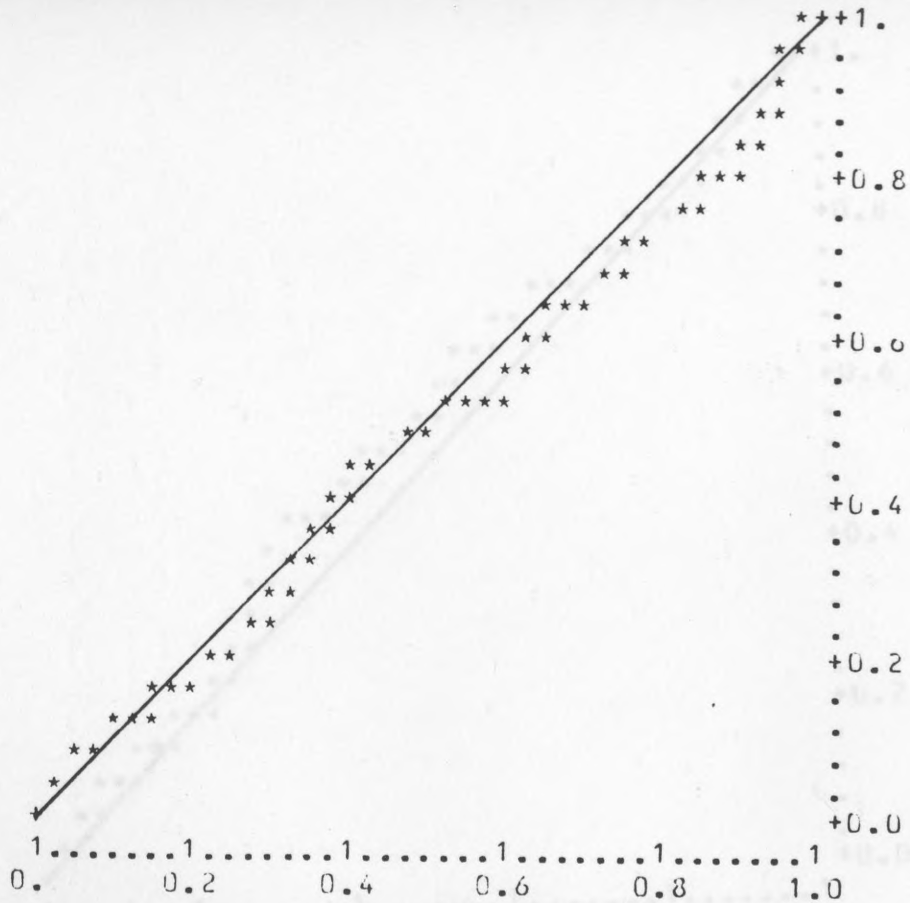


FIG. 5.42. KL u-plots, data in Table 5.4, the plots based on the line printer output.



FIG, 5.45. D y-plots, data in Table 5.4, the plots based on the line printer output.



FIG, 5.44. D u-plots, data in Table 5.4, the plots based on the line printer output.

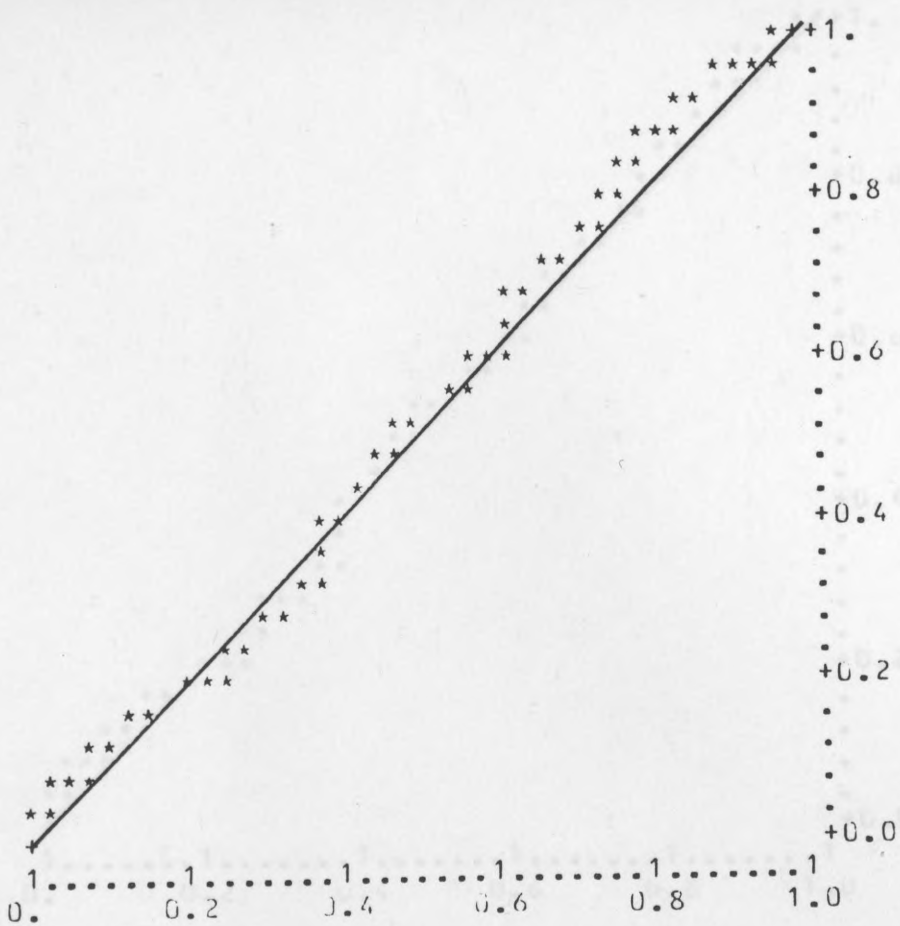


FIG. 5.47. JM y-plots, data in Table 5.4, the plots based on the line printer output.

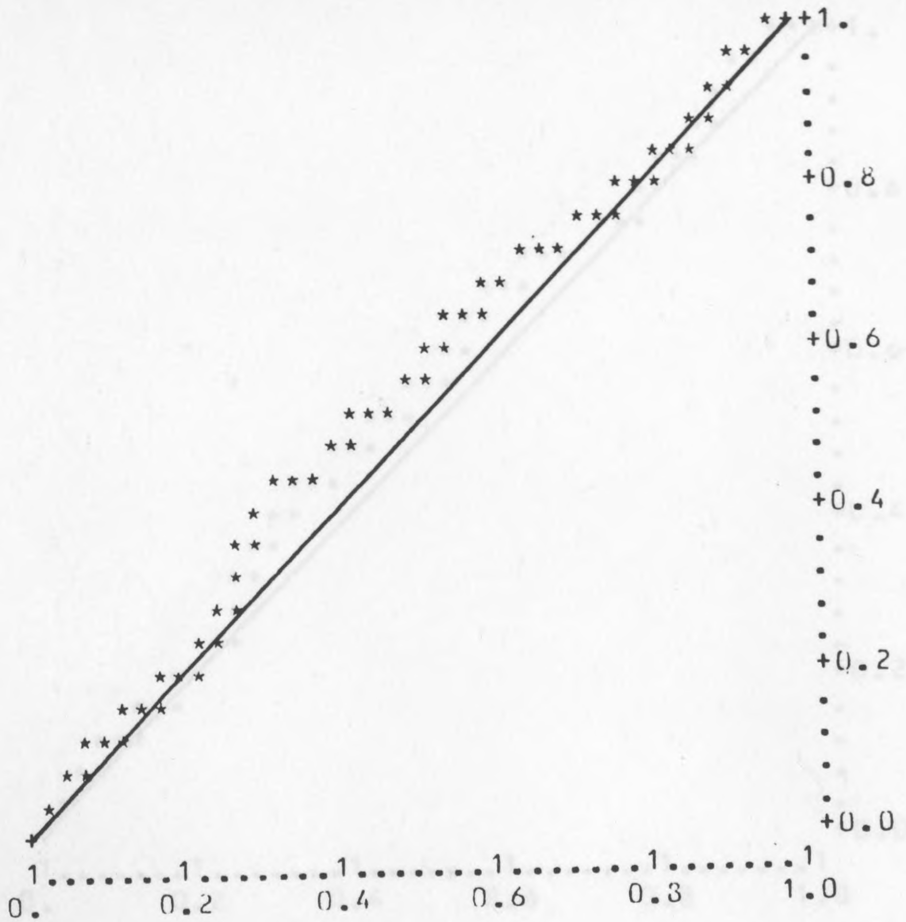


FIG. 5.46. JM u-plots, data in Table 5.4, the plots based on the line printer output.

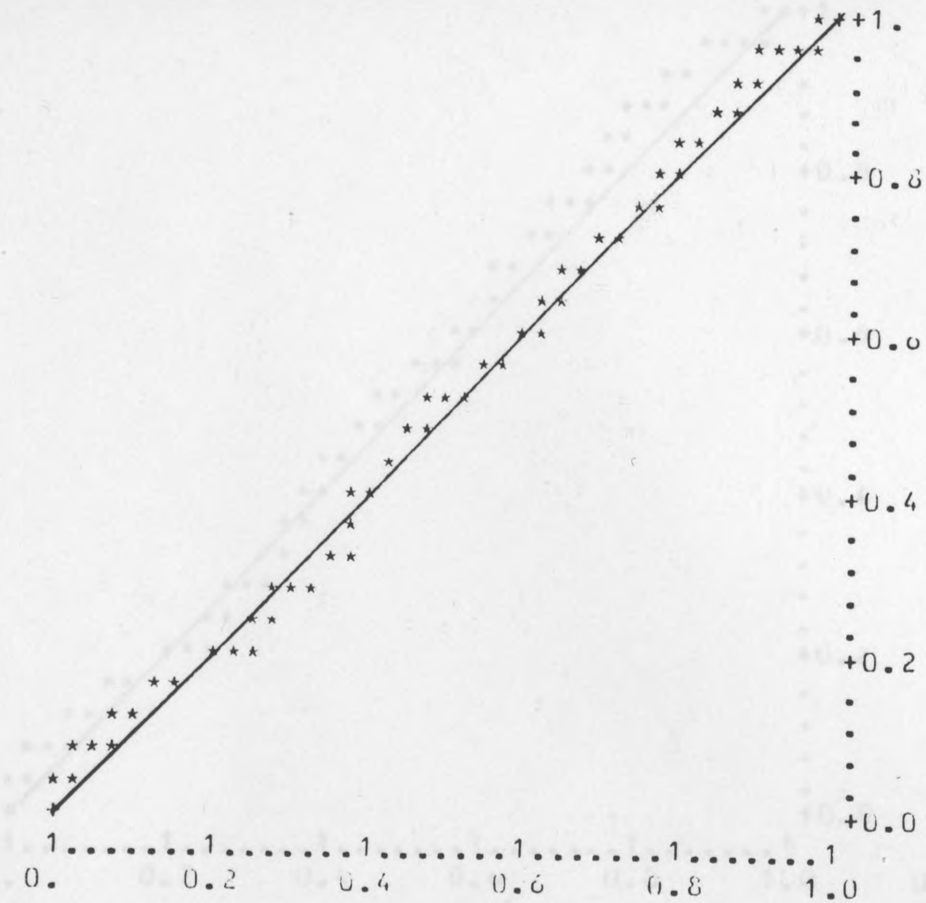


FIG. 5.49. BJM y-plots, data in Table 5.4, the plots based on the line printer output.

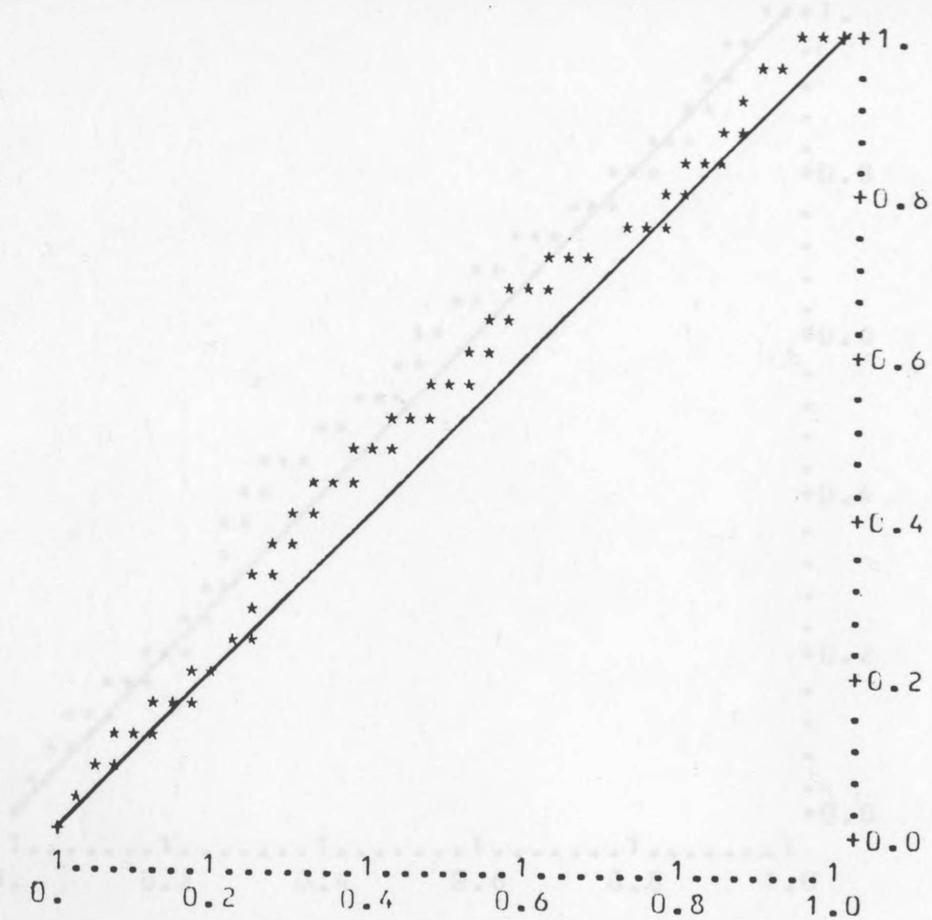


FIG. 5.48. BJM u-plots, data in Table 5.4, the plots based on the line printer output.

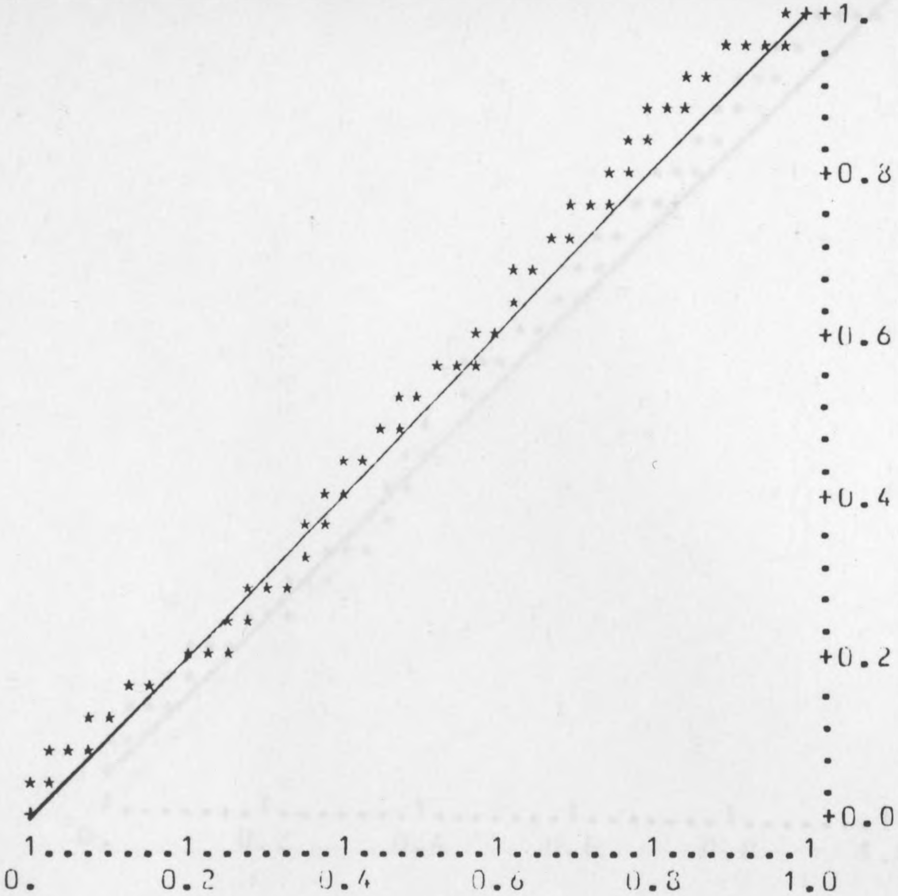


FIG.5.51. L y-plots, data in Table 5.4, the plots based on the line printer output.

FIG.5.50. L u-plots, data in Table 5.4, the plots based on the line printer output.

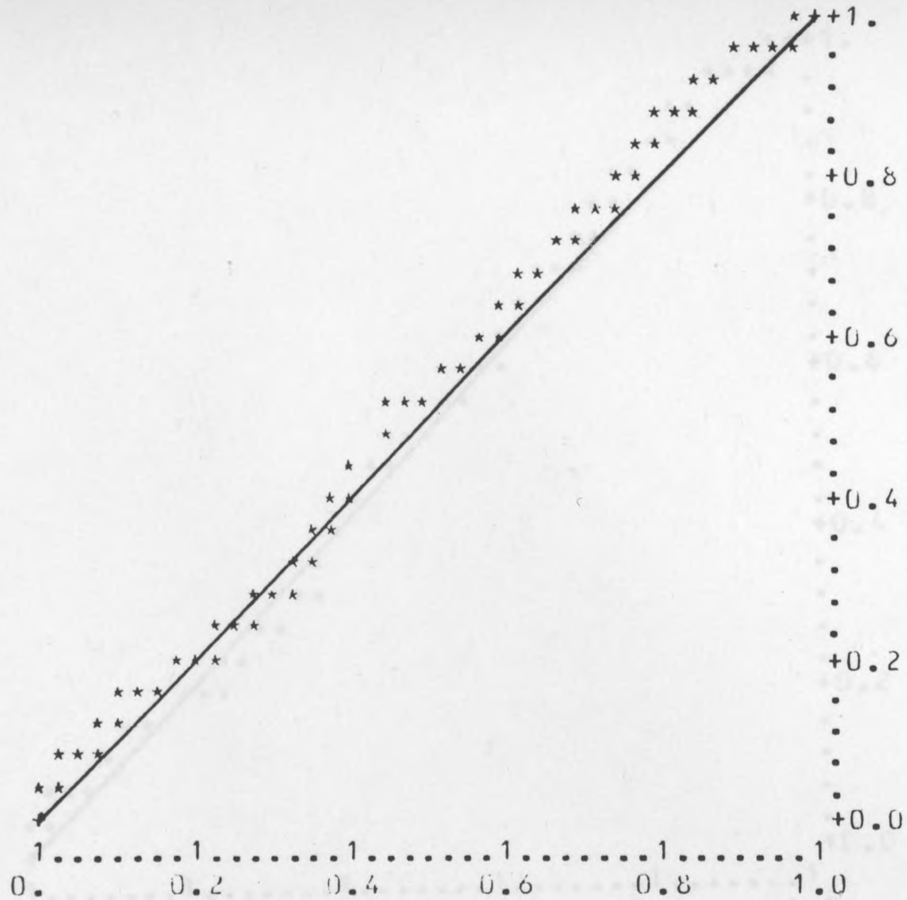


FIG.5.53. BL y-plots, data in Table 5.4., the plots based on the line printer output.

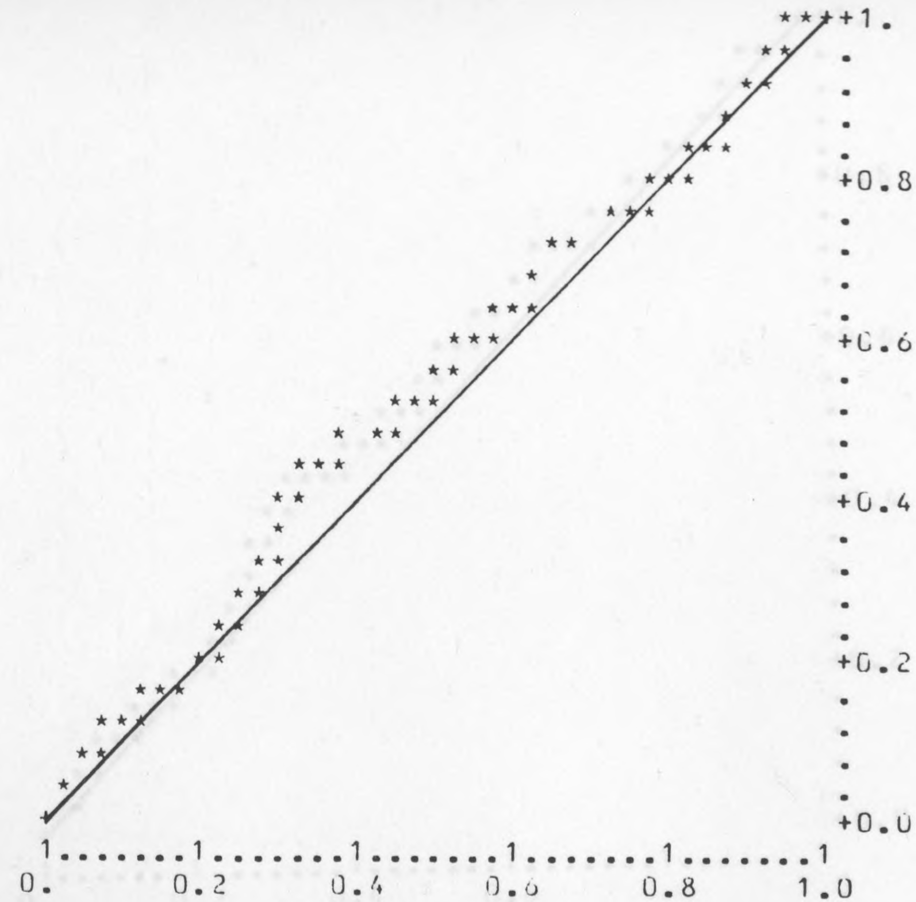


FIG. 5.52. BL u-plots, data in Table 5.4, the plots based on the line printer output.

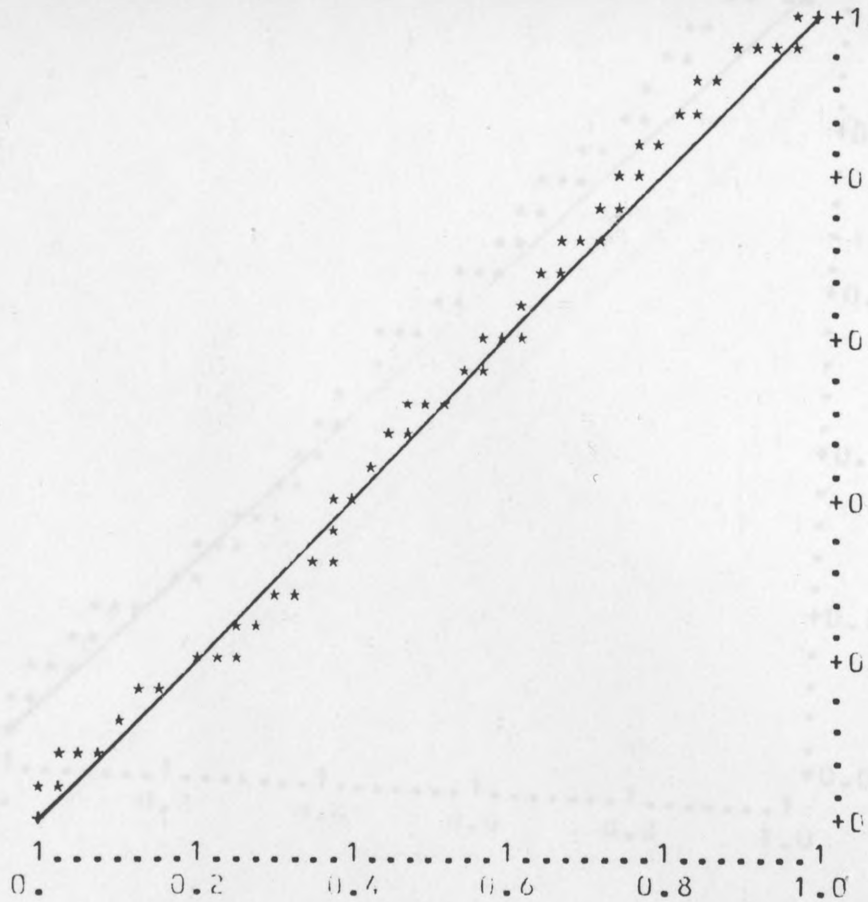


FIG. 5.55 JMNHPP y-plots, data in Table 5.4, the plots based on the line printer output.

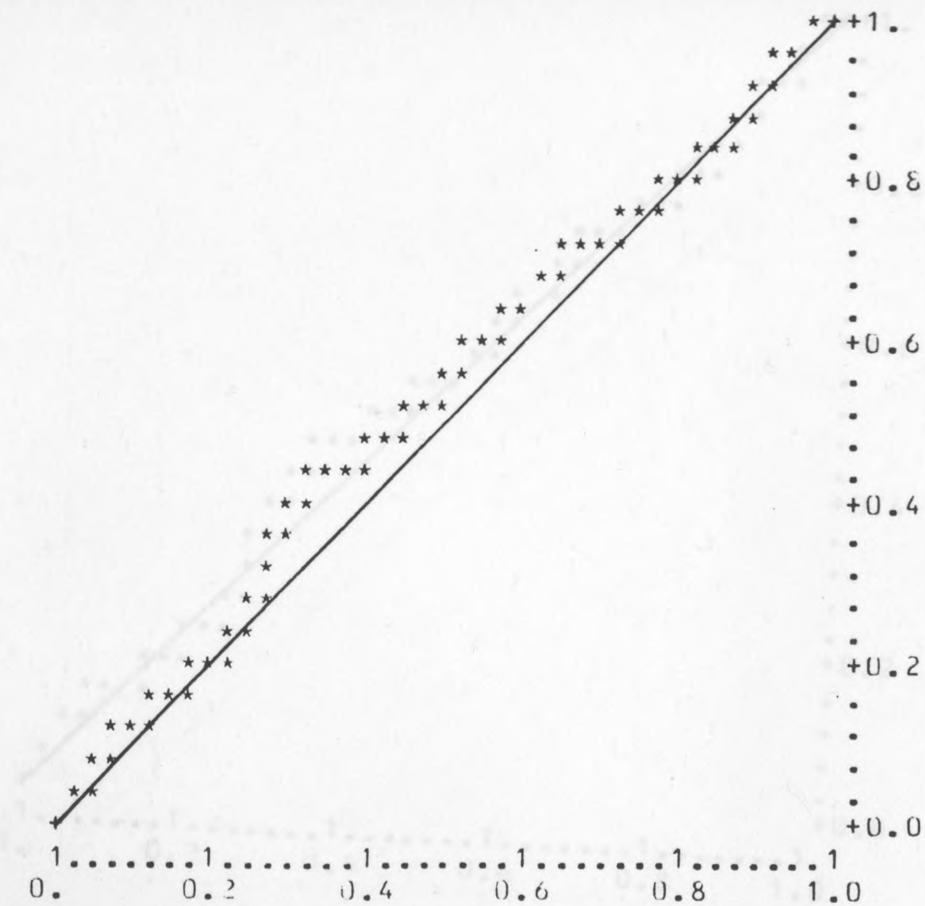


FIG. 5.54. JMNHPP u-plots, data in Table 5.4, the plots based on the line printer output.

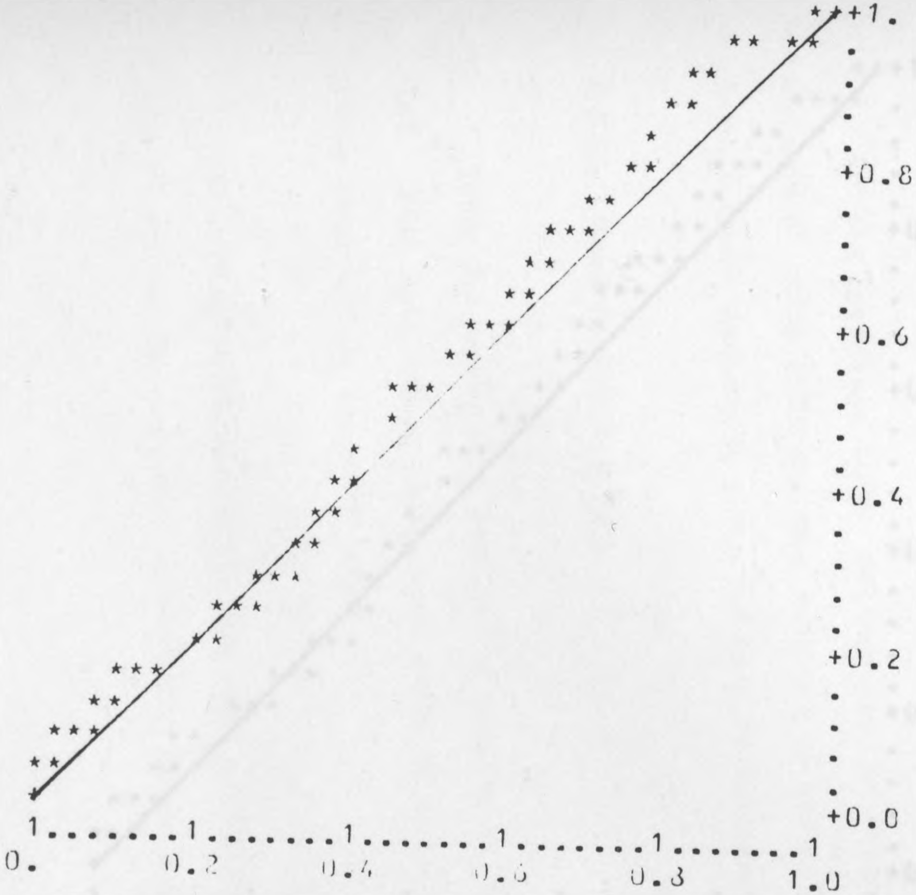


FIG.5.57. LNHPP y-plots, data in Table 5.4, the plots based on the line printer output.

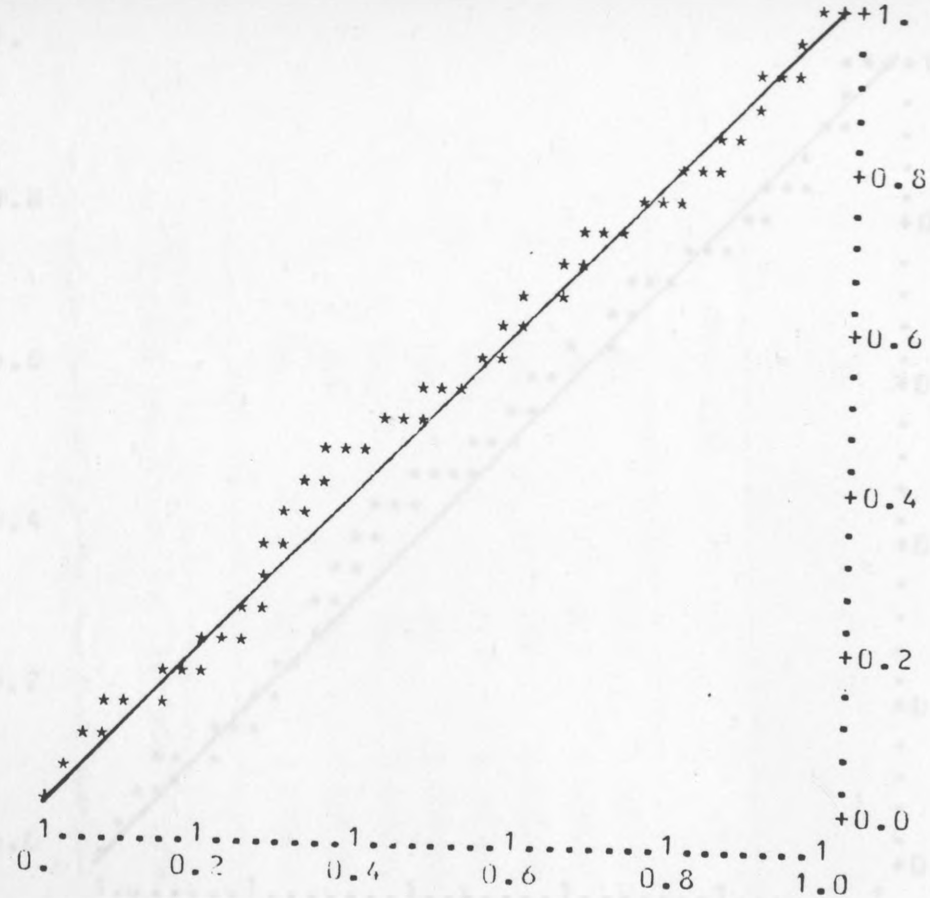


FIG. 5.56. LNHPP u-plots, data in Table 5.4, the plots based on the line printer output.

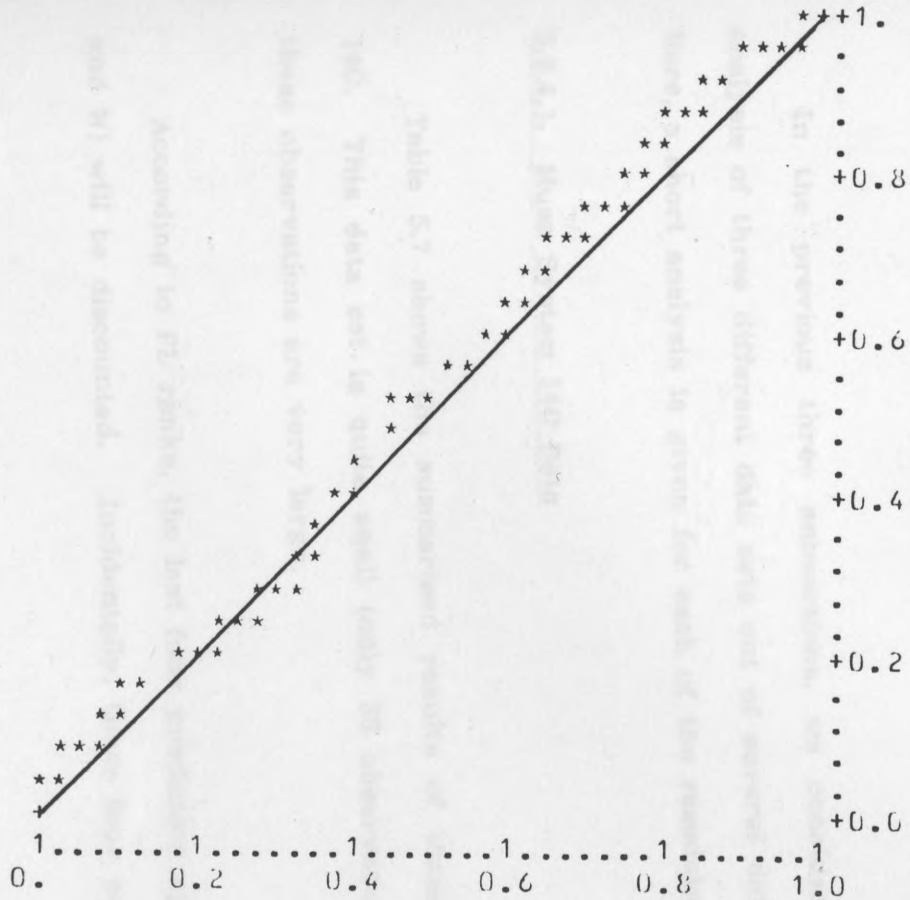


FIG.5.59 W y-plots, data in Table 5.4, the plots based on the line printer output.

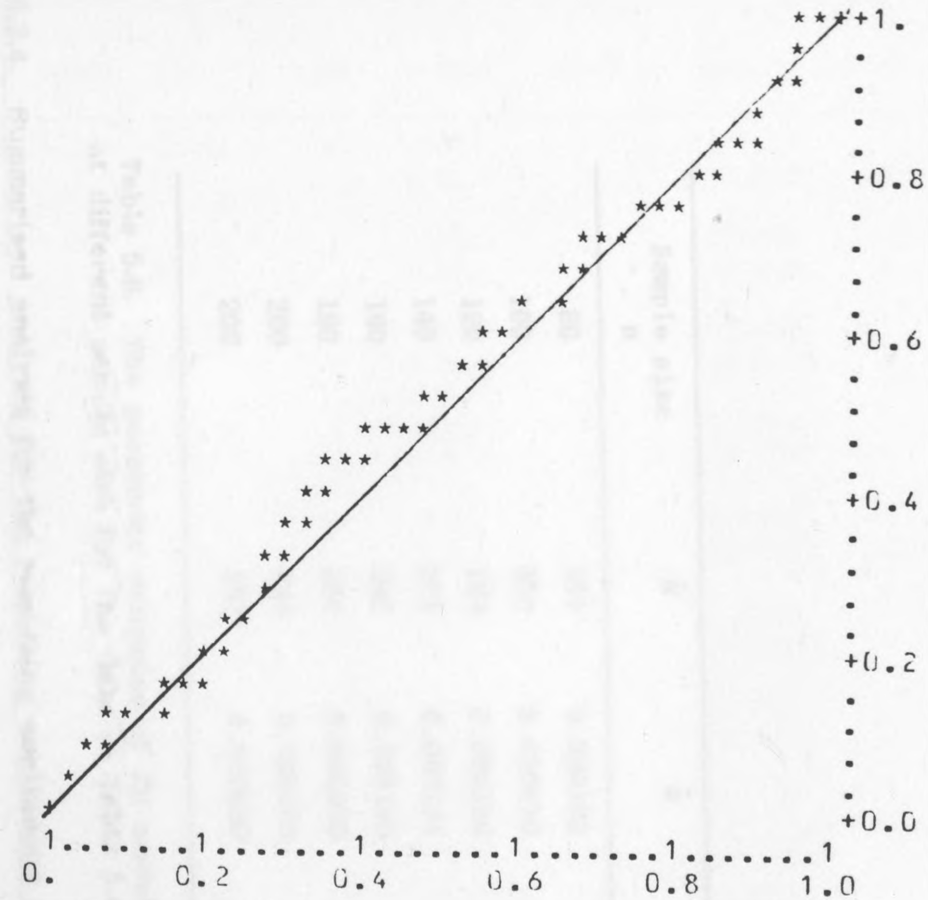


FIG. 5.58. W u-plots, data in Table 5.4, the plots based on the line printer output.

Sample size n	$\hat{N}$	$\hat{\Phi}$
80	257	0.000102
100	359	0.000070
120	183	0.000156
140	254	0.000101
160	245	0.000105
180	255	0.000099
200	244	0.000108
206	257	0.000099

Table 5.6. The parameter estimates of JM model at different sample size for the date in Table 5.4.

5.2.4. Summarised analysis for the remaining continuous data

In the previous three subsections, we considered the detailed analysis of three different data sets out of several data sets available. Here, a short analysis is given for each of the remaining data sets.

5.2.4.1. Musa System 14C Data

Table 5.7 shows the summarised results of Musa's (1979) system 14C. This data set is quite small (only 36 observations) and most of these observations are very large.

According to PL ranks, the last four prediction systems (LV, KL, D and W) will be discounted. Incidentally, these four predictions have

No.of Observations: = 36
Starting Sample Size : = 20

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	.9847	2	.8660	2	1.2219	3	.2682 20%	4	.1947 N.S.	6	12.7149 233.1722	3
BJM	1.9263	4	1.4431	4	1.1790	1	.2083 N.S.	1	.1519 N.S.	3	12.7766 230.9583	1
L	2.0003	6	1.7221	7	1.2542	4	.2644 20%	3	.2015 N.S.	7	12.7232 233.9658	4
BL	1.9263	4	1.4431	4	1.1790	1	.2083 N.S.	1	.1519 N.S.	3	12.7766 230.9583	1
LV(1)	4.8890	9	3.8391	9	***	8	.4413 1%	10	.1241 N.S.	2	12.8808 236.8394	7
KL(1)	3.4647	8	2.3503	8	***	8	.4115 1%	9	.1610 N.S.	5	12.2645 237.9225	8
D	2.0768	7	1.6939	6	1.3776	7	.3753 2%	8	.1225 N.S.	1	12.1569 241.8710	9
JMNHPP	.8183	1	.7556	1	1.2730	5	.2682 20%	4	.2135 N.S.	9	12.7150 234.3348	5
LNHPP	1.8826	3	1.2104	3	1.3343	6	.2682 20%	4	.2118 N.S.	8	12.7154 234.8436	6
W(1)	*	10	**	10	***	8	.3124 5%	7	.3123 5%	10	12.1569	10

TABLE 5.7. The summarised results of Musa's system 14C data.

(1) The model predicts; * infinite median, ** zero ROCOF and *** infinite or non-existent MTTF (or IMTBF). Same notations will be used in other Tables.

the highest ranks in all statistics (y-plot apart). The poor performance of these models seems mainly due to the large bias shown in their predictions (significant u-plot distances, Table 5.7). No value is assigned for Braun statistic in case of LV, KL and W because of the non-existence of MTTF. W predicts that the last fault has been removed for a range of values of n.

The remaining prediction systems (JM, BJM, L, BL, JMNHPP, and LNHPP) which are the best in PL ranks, are close in median plots (Figure 5.60). These plots are identical upto the last five predictions for JM, L, JMNHPP and LNHPP. JM performance is misleading on this data set because the data points did not satisfy Littlewood-Verrall condition (1981) except for the last five predictions.

BJM and BL are significantly better than the others on PL performance. These predictions are completely identical as a result of large (nearly infinite) $\hat{\beta}$ estimates in BL. These two models have the best u-plot distance, and are the only models for which this distance is non-significant. It is clear that BJM and BL should be preferred for future predictions.

5.2.4.2. Musa System 17 Data

Again, this data set is a short one and contains a very large observation at the end which affects the performance of most prediction systems (Figure 5.61).

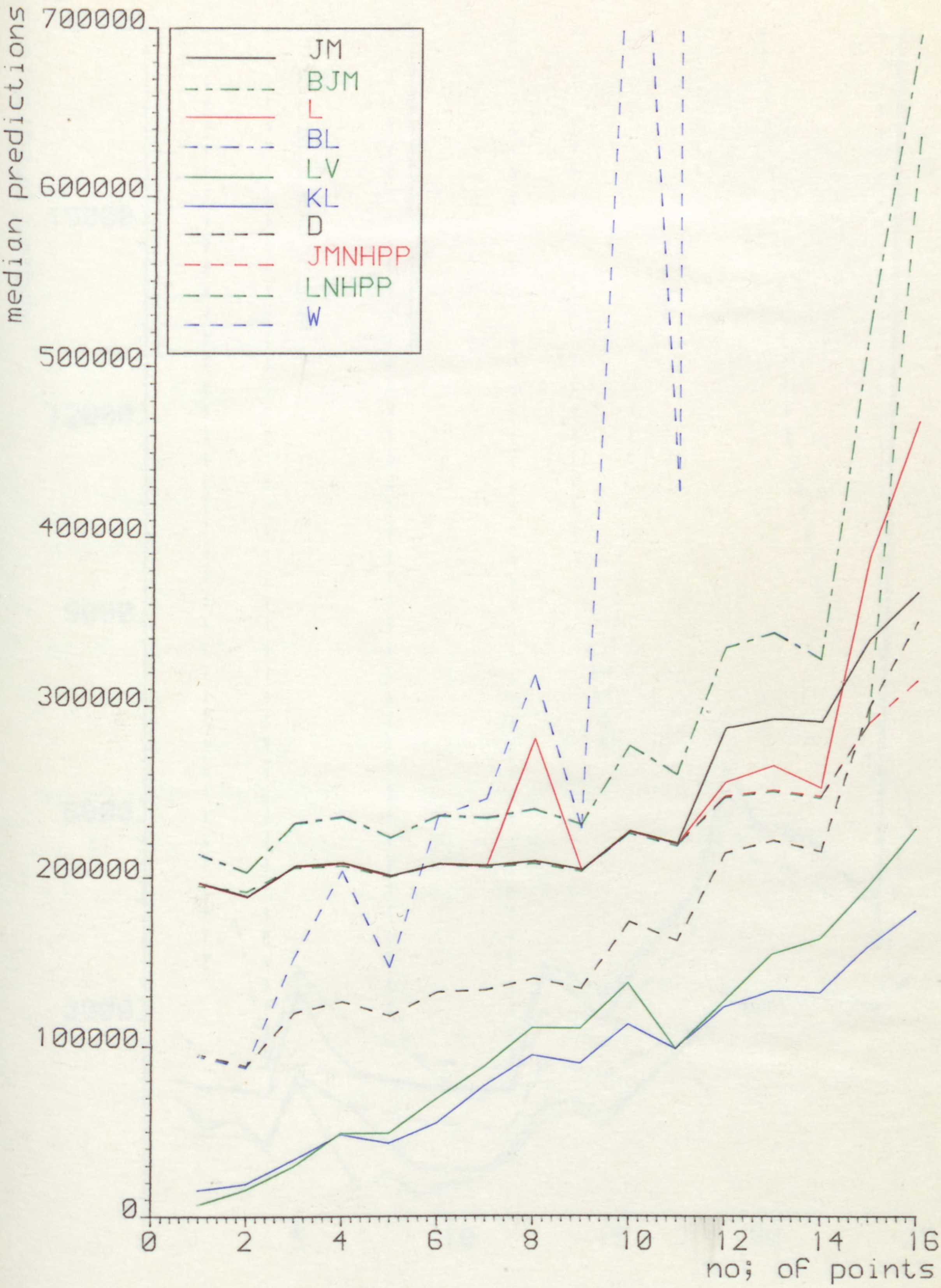


FIG.5.60. The plots of predictive medians for

Musa's system 14c data

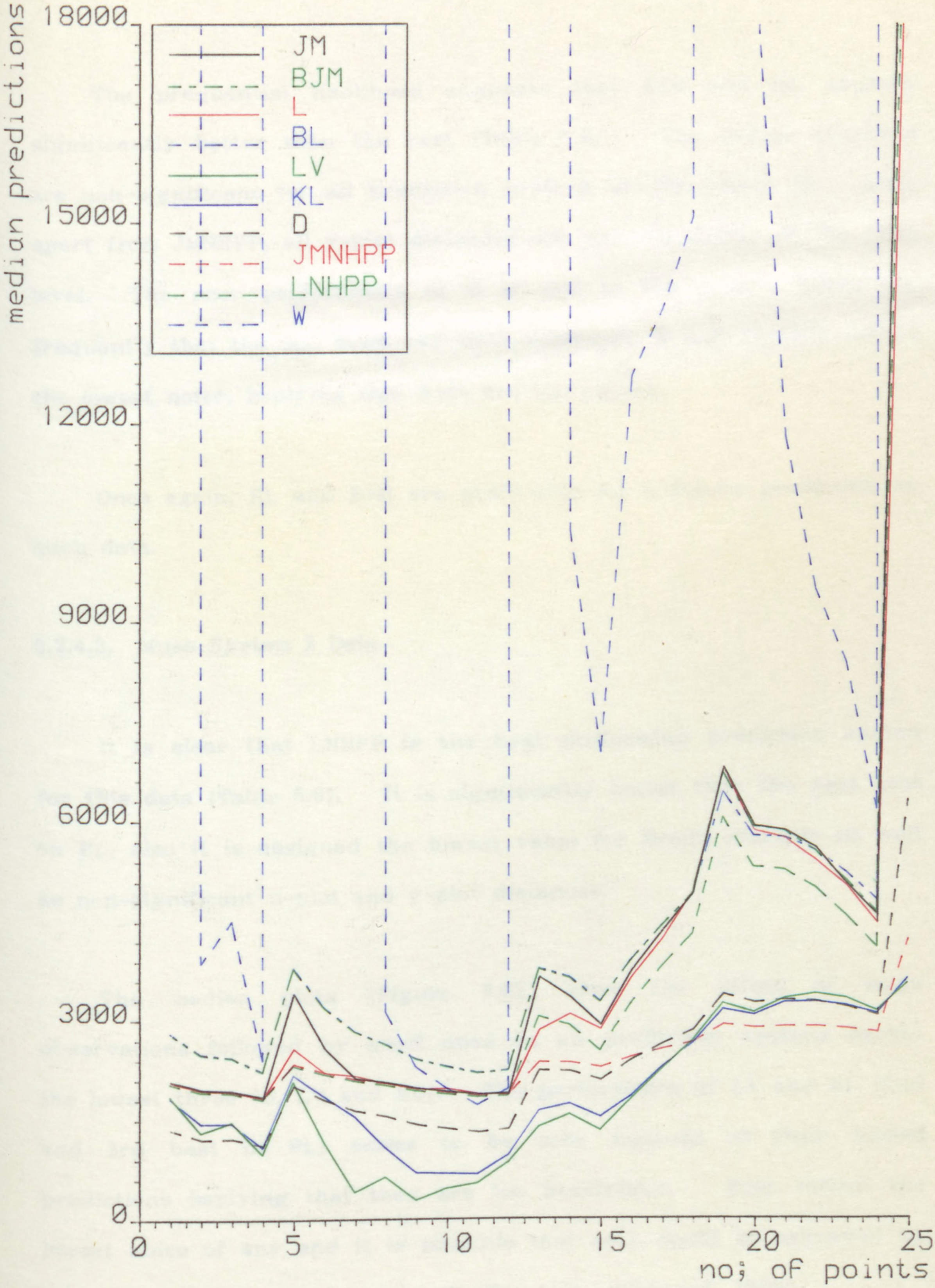


FIG.5.61. The plots of predictive medians for Musa's system 17 data

The prequential likelihood suggests that BJM and BL perform significantly better than the rest (Table 5.8). The u-plot distances are non-significant for all prediction systems at 10% except W. Also, apart from JMNHPP, all y-plot distances are non-significant at the same level. The poor performance of W is due to the system predicting frequently that the last fault has been removed. D and JMNHPP reflect the lowest noise, implying that both are too smooth.

Once again, BL and BJM are preferable for a future prediction on such data.

5.2.4.3. Musa System 2 Data

It is clear that LNHPP is the best performing prediction system for this data (Table 5.9). It is significantly better than the next best on PL, also it is assigned the lowest value for Braun statistic as well as non-significant u-plot and y-plot distances.

The median plots (Figure 5.62) show the effect of large observations followed by small ones on all prediction systems except the lowest three (D, LV and KL). The performance of LV and KL (2nd and 3rd best in PL) seems to be poor because of their biased predictions implying that they are too pessimistic. Both reflect the lowest noise of any and it is possible that each could be improved by applying an adaptive procedure [Keiller and Littlewood, 1984].

Test Continuous Data System Musa System 17

No.of Observations: = 38
Starting Sample Size : = 13

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	9.0333	9	3.9885	7	1.2158	9	.1884 N.S.	5	.2359 N.S.	7	8.2076 251.3175	7
BJM	7.9803	8	3.3716	6	1.1276	6	.2152 N.S.	6	.1826 N.S.	3	8.4309 245.3614	1
L	6.6834	4	3.2285	4	1.1443	8	.1828 N.S.	4	.2368 N.S.	8	8.2076 250.7667	6
BL	7.6366	7	3.2583	5	1.1199	5	.2152 N.S.	6	.1873 N.S.	4	8.4309 245.5081	2
LV	7.3109	5	8.3326	9	1.0769	2	.2735 20%	9	.1032 N.S.	1	8.1308 249.7701	4
KL	4.7902	3	4.0958	8	1.0709	1	.2236 N.S.	8	.1143 N.S.	2	8.1174 249.5152	3
D	3.8820	2	2.7078	2	1.1050	3	.1629 N.S.	1	.2258 N.S.	6	7.8825 261.3061	9
JMNHPP	1.7315	1	1.4530	1	1.1091	4	.1691 N.S.	2	.3106 10%	10	8.2076 257.7631	8
LNHPP	7.3532	6	33.0801	3	1.1339	7	.1682 N.S.	2	.2123 N.S.	5	8.2077 250.1423	5
W	*	10	**	10	***	10	.3975 2%	10	.2577 N.S.	9	48.5240 –	10

TABLE 5.8. The summarised results of Musa's system 17 data.

Test Continuous Data System Musa System 2

No.of Observations: = 54
Starting Sample Size : = 13

MODEL	$\sum_{i=2}^{n}\lvert 1-\frac{m_i}{m_{i-1}}\rvert$	RANK	$\sum_{i=2}^{n}\lvert 1-\frac{r_i}{r_{i-1}}\rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	12.2038	9	8.9883	8	2.6003	9	.1940 10%	9	.1627 N.S.	9	8.3129 363.3807	8
BJM	11.3677	7	6.8918	6	1.4564	7	.1557 N.S.	6	.1585 N.S.	7	8.5189 353.6922	9
L	11.9306	8	9.2025	9	1.6379	8	.1256 N.S.	2	.1335 N.S.	5	8.3385 359.9345	5
BL	4.9462	5	3.9656	5	.9579	5	.1364 N.S.	4	.1093 N.S.	3	8.3780 358.7478	4
LV	2.2742	1	2.0427	1	.9270	3	.1768 20%	7	.1277 N.S.	4	8.3923 358.2312	2
KL	2.3713	2	2.1393	2	.8990	2	.1786 20%	8	.1069 N.S.	2	8.4094 358.3392	3
D	3.6725	3	3.0947	3	.9425	4	.2063 10%	10	.1055 N.S.	1	8.8652 360.5915	6
JMNHPP	11.3411	6	7.2885	7	1.4301	6	.1332 N.S.	3	.1876 10%	10	8.6414 361.6410	7
LNHPP	4.7710	4	3.8646	4	.8752	1	.1212 N.S.	1	.1485 N.S.	6	8.6344 356.9698	1
W	*	10	**	10	***	10	.1540 N.S.	5	.1622 N.S.	8	8.1486 -	10

TABLE 5.9. The summarised results of Musa's system 2 data.
ML routine does not terminate normally for L, LV and KL.

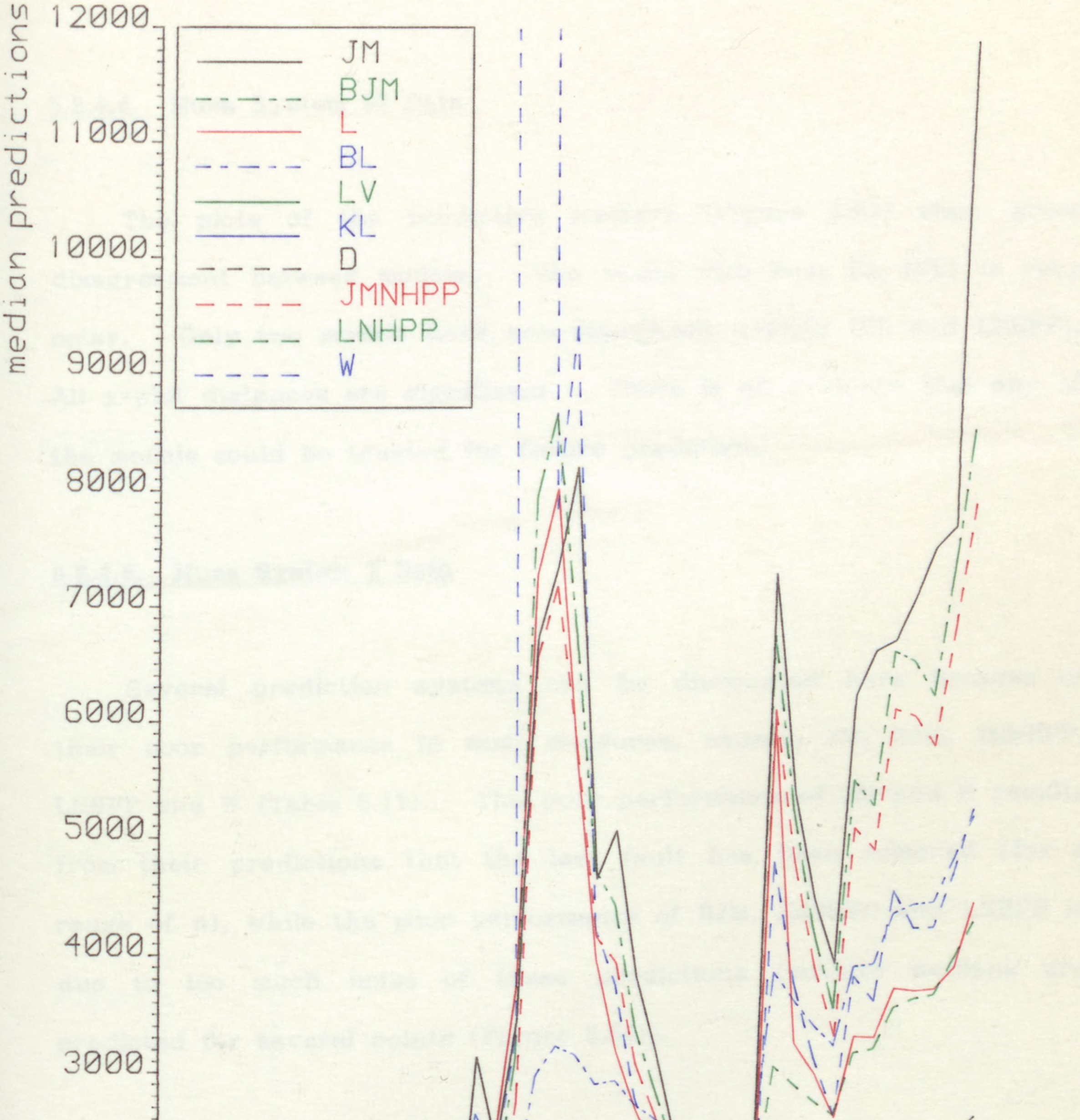


FIG.5.62. The plots of predictive medians for Musa's system 2 data

5.2.4.4. Musa System 27 Data

The plots of the predictive medians (Figure 5.63) show great disagreement between models. The model with best PL (BL) is very noisy. Only two models have non-significant u-plots (BL and LNHPP). All y-plot distances are significant. There is no evidence that any of the models could be trusted for future prediction.

5.2.4.5. Musa System 3 Data

Several prediction systems can be discounted here because of their poor performance in most measures, namely, JM, BJM, JMNHPP, LNHPP and W (Table 5.11). The poor performance of JM and W results from their predictions that the last fault has been removed (for a range of n), while the poor performance of BJM, JMNHPP and LNHPP is due to too much noise of these predictions (infinite medians are predicted for several points (Figure 5.64)).

LV, KL and D are reflecting the lowest noise but are too pessimistic as shown by u-plots which have highly significant distances.

Finally, L is the second best in PL performance but it reflects a very noisy prediction (Figure 5.64). So it is discredited in favour of BL. The later is significantly best on the prequential likelihood performance as well as u-plot and y-plot distances of the model being

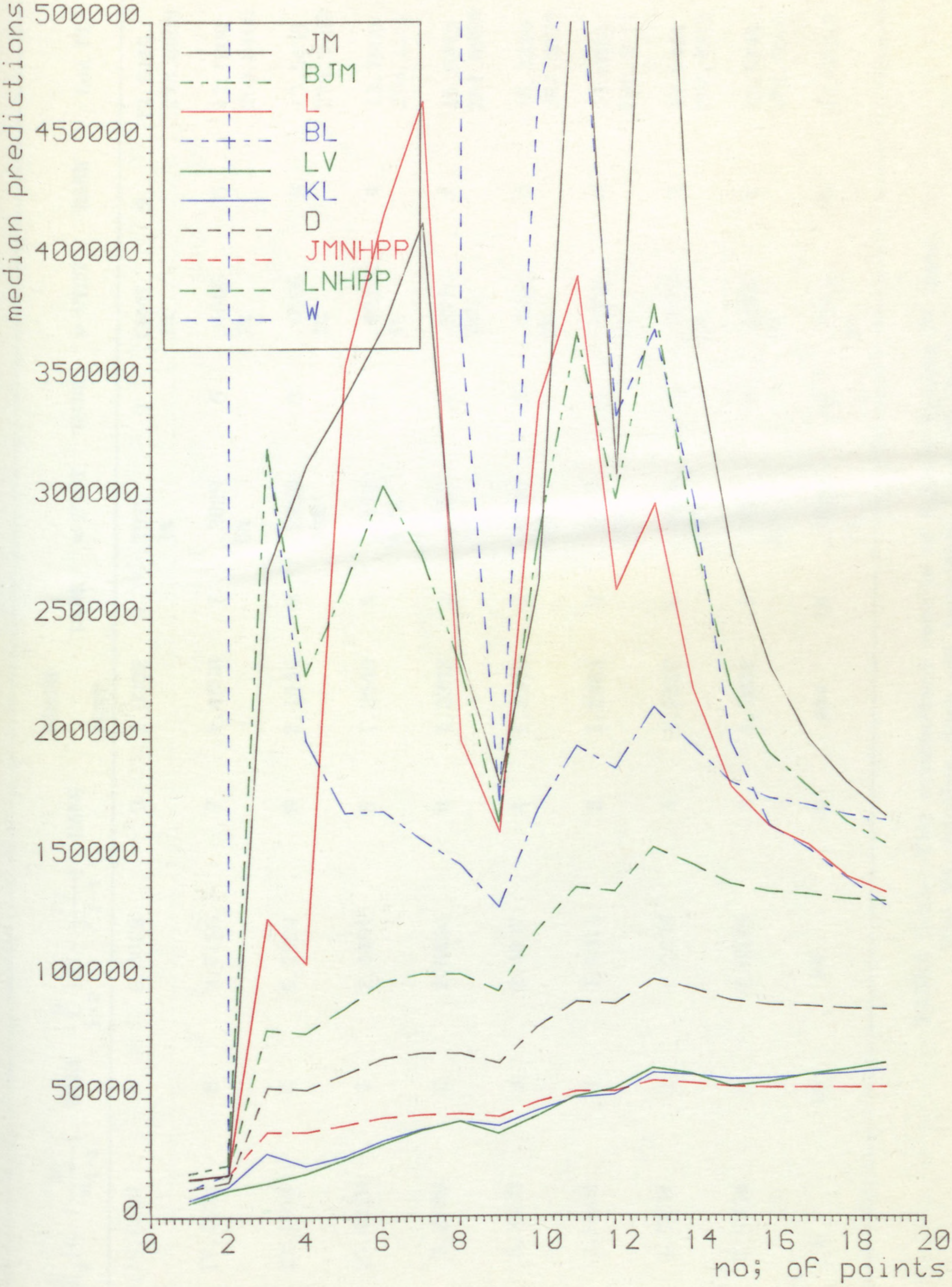


FIG.5.63. The plots of predictive medians for Musa's system 27 data

No.of Observations: = 41
Starting Sample Size : = 22

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	19.6627	9	6.3003	9	2.9302	9	.3922 1%	9	.6886 1%	9	13.5445 278.8008	7
BJM	17.1634	8	3.7156	7	1.4968	7	.3029 5%	5	.4866 1%	5	13.1800 260.8400	2
L	12.1591	6	5.7920	8	2.1446	8	.2759 10%	3	.6566 1%	8	13.5445 275.9653	6
BL	15.4028	7	2.4494	5	1.2549	5	.2112 N.S.	2	.4217 1%	4	13.1800 258.1462	1
LV	3.2780	2	2.6400	6	1.2252	2	.3707 1%	7	.2776 10%	1	15.0349 261.6959	3
KL	3.3695	3	2.4097	4	1.2545	4	.3589 2%	6	.3391 2%	3	15.3349 261.8679	4
D	3.9004	4	1.8311	2	1.2383	3	.2809 10%	4	.5264 1%	6	14.6021 281.4611	8
JMNHPP	1.7181	1	1.1636	1	1.3232	6	.3951 1%	10	.3379 2%	2	13.5444 282.3312	9
LNHPP	4.6358	5	1.8526	3	1.1830	1	.1553 N.S.	1	.5686 1%	7	13.5445 271.2016	5
W	*	10	**	10	***	10	.3815 1%	8	.7151 1%	10	14.60213	10

TABLE 5.10. The summarised results of Musa's system 27 data.
The ML routine does not terminate normally for L.

Test Continuous Data System Musa System 3

No.of Observations: = 38
Starting Sample Size : = 13

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	*	6	**	9	***	9	.4777 1%	10	.4569 1%	10	5.8706 –	9
BJM	*	6	8.6794	6	12.0615	7	.2314 20%	6	.2889 5%	7	5.9422 246.3822	8
L	16.6506	5	9.9025	8	15.3056	8	.1514 N.S.	1	.2360 20%	5	5.6678 239.1870	2
BL	7.8191	4	5.9750	4	.9366	2	.1526 N.S.	2	.1140 N.S.	2	6.5143 237.1442	1
LV	6.3929	2	5.2275	2	.9825	3	.3320 1%	9	.1358 N.S.	4	5.2509 243.7027	4
KL	6.8478	3	5.8724	3	1.2929	4	.2914 5%	7	.0991 N.S.	1	5.4883 242.6179	5
D	5.1718	1	3.9850	1	.8737	1	.3001 2%	8	.1260 N.S.	3	5.6461 241.2496	3
JMNHPP	*	6	7.7733	5	9.9270	6	.1832 N.S.	3	.3597 1%	9	5.5487 246.1144	7
LNHPP	*	6	8.7511	7	9.9257	5	.2064 N.S.	4	.3394 1%	8	5.6670 245.7035	6
W	*	6	**	9	***	9	.2194 20%	5	.2771 5%	6	5.8161 –	9

TABLE 5.11. The summarised results of Musa's system 3 data

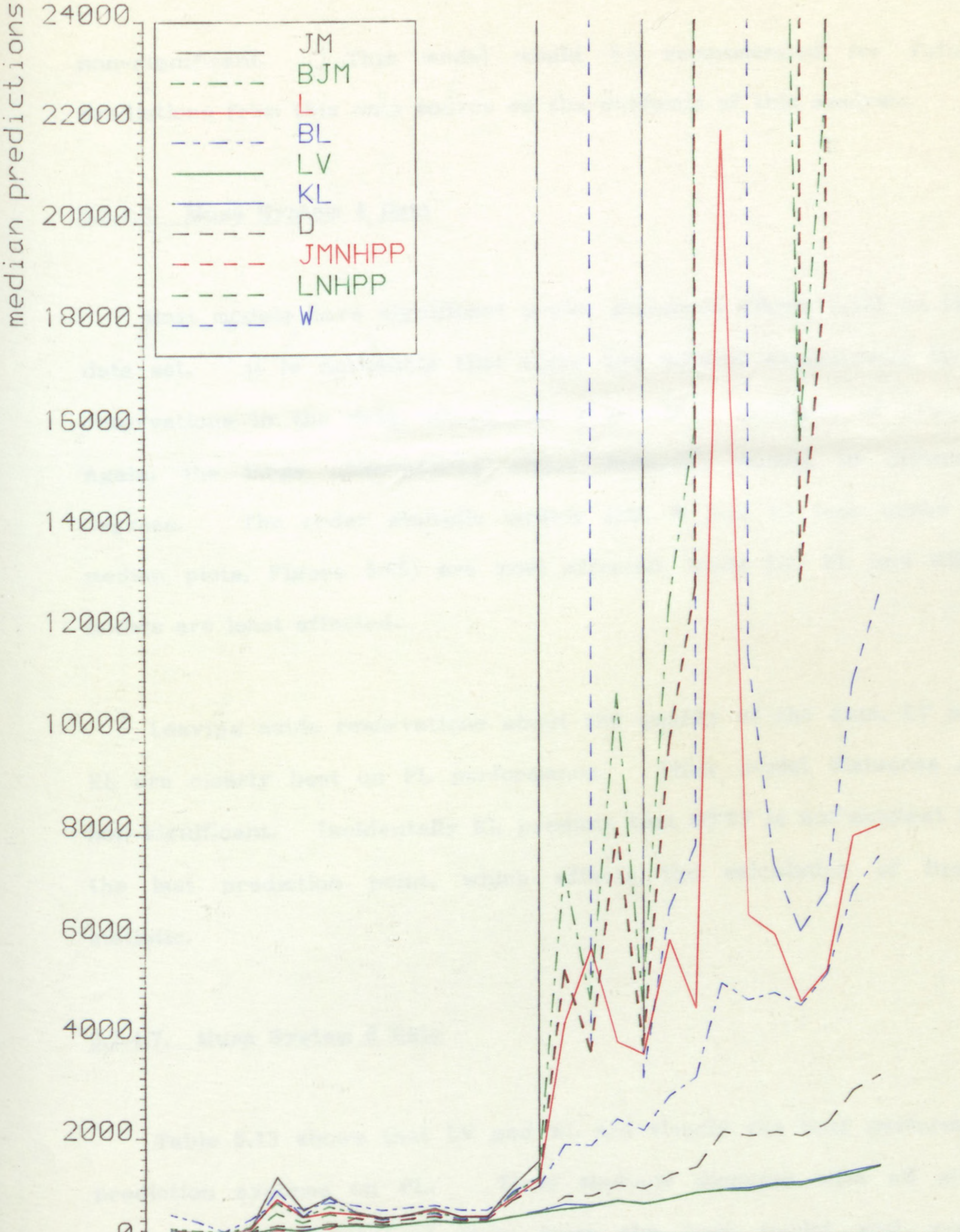


FIG.5.64. The plots of predictive medians for Musa's system 3 data

non-significant. This model would be recommended for future predictions from this data source on the evidence of this analysis.

5.2.4.6. Musa System 4 Data

Most models have significant y-plot distances (Table 5.12) on this data set. it is noticeable that there are several suspiciously large observations in the data, which also shows little evidence of growth. Again, the large observations affect different models to different degrees. The order statistic models (JM, W and L) (see peaks in median plots, Figure 5.65) are most affected, while LV, KL and NHPP models are least affected.

Leaving aside reservations about the quality of the data, LV and KL are clearly best on PL performance. Their u-plot distances are non-significant. Incidentally KL predicts that MTTF is not existent for the last prediction point, which affects the calculation of Braun statistic.

5.2.4.7. Musa System 6 Data

Table 5.13 shows that LV and KL are clearly the best performing prediction systems on PL. Their medians disagree with all other models (Figure 5.66), yet they have the best u-plot and y-plot distances. Indeed all other models have highly significant u- and

No.of Observations: = 53
Starting Sample Size : = 13

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	*	6	**	9	***	8	.1688 20%	7	.2459 2%	5	5.6696 –	9
BJM	*	6	6.9797	7	15.3460	7	.1851 20%	9	.1290 N.S.	1	5.7769 302.6742	3
L	14.2267	5	8.8457	8	.9738	2	.1197 N.S.	2	.3072 1%	8	5.6765 313.4945	6
BL	*	5	6.4017	6	4.7661	5	.1722 20%	8	.1768 2%	3	5.9939 304.1979	4
LV	5.2574	1	5.3402	3	1.0248	4	.1289 N.S.	3	.2135 5%	4	5.9705 301.0855	2
KL	5.6576	3	5.8844	5	***	8	.1431 N.S.	6	.1671 20%	2	5.7362 300.0959	1
D	5.5204	2	3.6919	2	1.0126	3	.1188 N.S.	1	.3787 1%	9	5.6205 321.8510	7
JMNHPP	*	6	3.0736	1	11.5527	6	.2085 10%	10	.4537 1%	10	5.4662 362.6669	8
LNHPP	8.4656	4	5.3878	4	.9695	1	.1350 N.S.	4	.2856 1%	7	5.4910 311.5787	5
W	*	6	**	9	***	8	.1418 N.S.	5	.2678 1%	6	5.6205 –	9

TABLE 5.12. The summarised results of Musa's system 4 data.
The ML routine does not terminate normally for L.

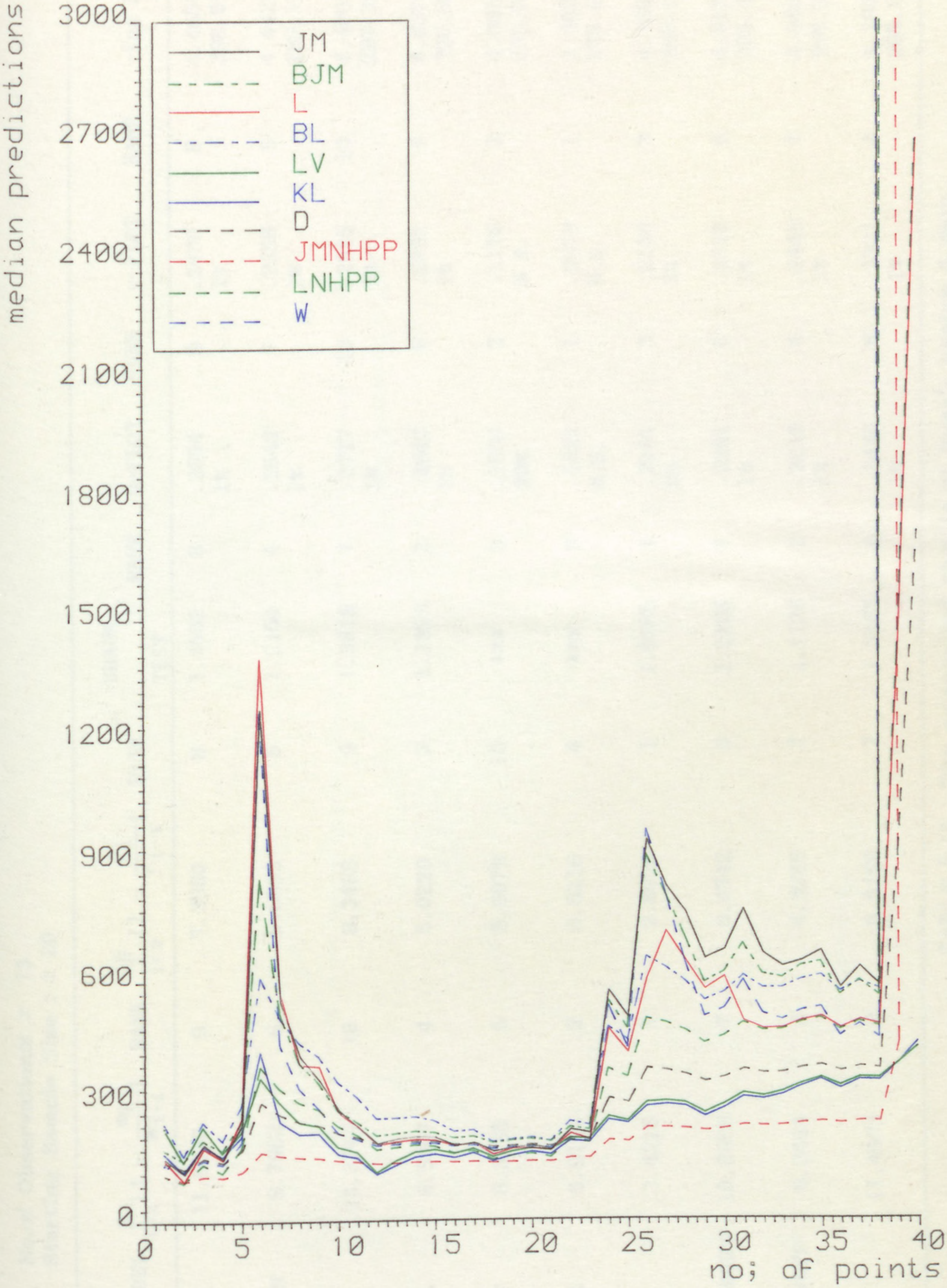


FIG.5.65. The plots of predictive medians for Musa's system 4 data

Test Continuous Data System Musa System 6

No.of Observations: = 73
Starting Sample Size : = 20

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	11.9530	9	7.3860	8	1.4552	8	.2654 1%	9	.2489 1%	8	4.4650 298.0149	10
BJM	9.7962	6	5.6393	5	1.2154	4	.2543 1%	7	.2356 1%	5	4.4422 289.1597	3
L	13.3744	10	8.3488	9	1.3816	7	.2717 1%	10	.2515 1%	10	4.4606 297.3525	9
BL	6.5967	4	5.0229	3	1.1827	3	.2643 1%	8	.2492 1%	9	4.6371 290.6683	4
LV	8.3735	5	8.6079	10	***	9	.1554 20%	2	.1176 N.S.	2	4.0814 276.5956	2
KL	5.9156	2	5.6216	4	***	9	.1421 N.S.	1	.0979 N.S.	1	3.8825 275.6645	1
D	3.4316	1	2.8070	1	1.0585	1	.2191 2%	3	.2130 2%	3	4.1634 290.9713	5
JMNHPP	10.6868	7	6.4742	6	1.2166	5	.2291 1%	5	.2379 1%	6	4.3124 295.4205	7
LNHPP	6.3464	3	4.9263	2	1.1176	2	.2210 1%	4	.2443 1%	7	4.3459 293.3122	6
W	11.4503	8	6.5156	7	1.3167	6	.2436 1%	6	.2321 1%	4	4.3962 295.6072	8

TABLE 5.13. The summarised results of Musa's system 6 data. The ML routine does not terminate normally for L.

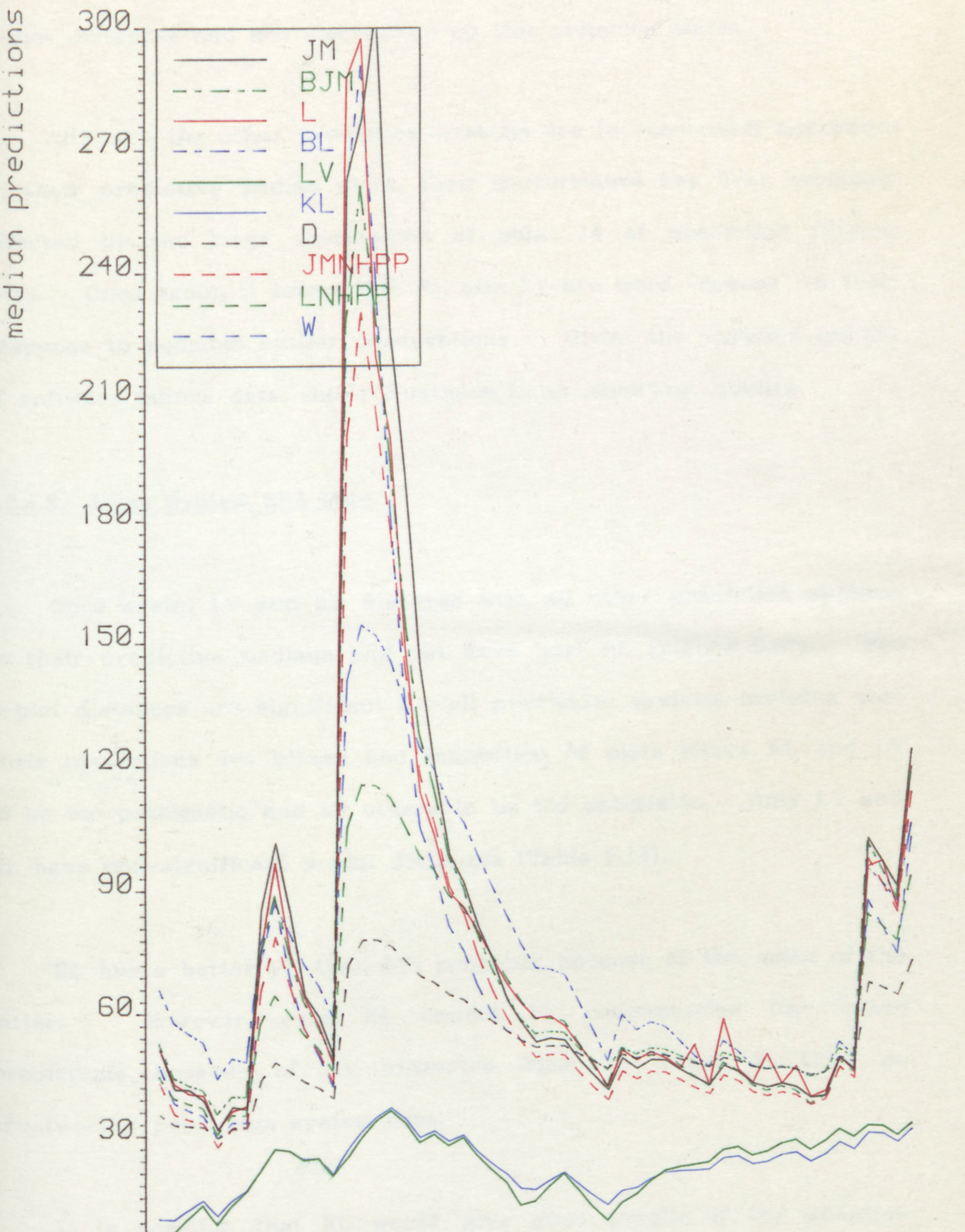


FIG.5.66. The plots of predictive medians for Musa's system 6 data

y-plot distances and are discredited on this criterion alone.

Although the other prediction systems are in reasonable agreement on their predictive median plots, their performance has been seriously affected by the large observation at point 14 of prediction (Figure 5.66). Once again, it seems that KL and LV are more "robust" in their response to possible outlier observations. Given the variable quality of software failure data, this robustness is an important quality.

5.2.4.8. Musa System SS4 Data

Once again, LV and KL disagree with all other prediction systems in their predictive medians and yet have best PL (Figure 5.67). The u-plot distances are significant for all prediction systems implying that their predictions are biased and inspection of plots shows KL and LV to be too pessimistic and all others to be too optimistic. Only LV and KL have non-significant y-plot distances (Table 5.14).

KL has a better PL than LV, probably because of the noise of the latter. However, even KL cannot be recommended for future predictions because of its excessive bias. There is thus no trustworthy prediction system here.

It is possible that KL would give good results if the adaptive procedure [Keiller and Littlewood, 1984] were used.

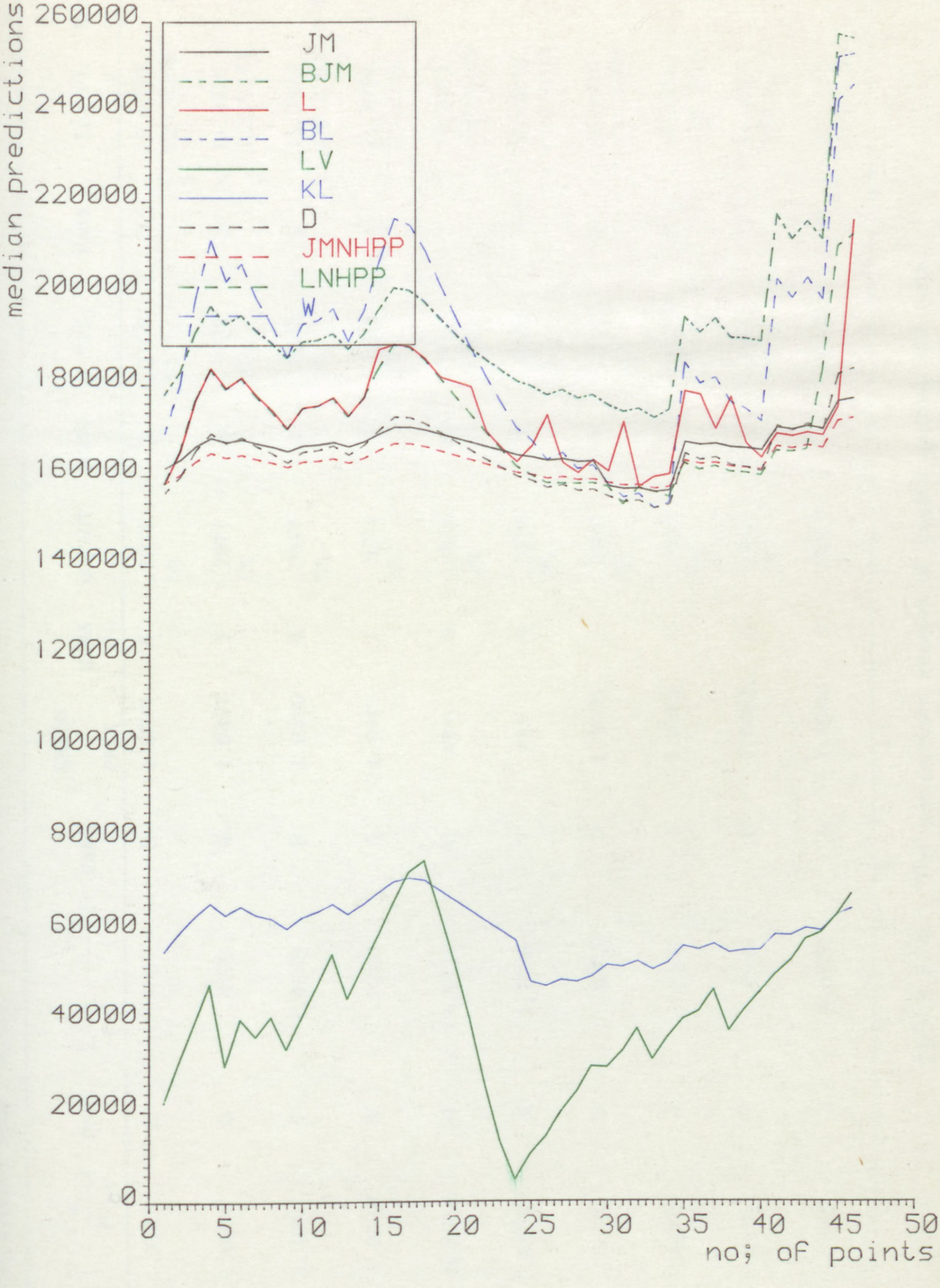


FIG.5.67. The plots of predictive medians for Musa's system ss4 data

No.of Observations: = 196
Starting Sample Size : = 150

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	.3713	2	.3656	2	1.0537	5	.2031 5%	3	.2165 5%	7	15.0250 629.8999	7
BJM	1.0898	6	.8170	4	1.0407	1	.2071 5%	6	.1637 20%	3	14.8776 628.5343	3
L	1.4670	7	1.4085	8	1.0590	8	.2047 5%	4	.2242 2%	9	15.0592 630.2209	10
BL	1.0648	5	.8226	5	1.0407	1	.2071 5%	6	.1641 20%	4	14.8776 628.5369	4
LV	9.2754	10	11.1531	10	***	9	.2912 1%	10	.0821 N.S.	1	16.4371 628.1619	2
KL	1.5015	8	1.2558	7	***	9	.2275 2%	9	.0994 N.S.	2	15.8784 624.7500	1
D	.6326	3	.6172	3	1.0570	7	.2027 5%	2	.2159 5%	6	15.0835 630.1699	9
JMNHPP	.3186	1	.3147	1	1.0562	6	.2023 5%	1	.2156 5%	5	15.0591 630.0878	8
LNHPP	1.0162	4	.9096	6	1.0497	3	.2047 5%	4	.2194 5%	8	15.0585 629.6334	6
W	1.7808	9	1.6748	9	1.0502	4	.2079 5%	8	.2249 2%	10	14.9507 629.6286	5

TABLE 5.14. The summarised results of Musa's system SS4 data.
The ML routine does not terminate normally for L.
The calculations of BJM involve overflo and underflo in some predictions.

5.2.4.9 System B7 Data

Apart from D and JMNHPP, all prediction systems give similar PL performance (Table 5.15). The poor performance of D and JMNHPP is due to their biased predictions as confirmed by their significant u-plot distances. Thus both models are too pessimistic. Since both have good noise statistics, they could perhaps be improved by adapting. W performs quite poorly on PL because it is the most noisy prediction system, despite having lowest u-plot and y-plot distances (Figure 5.68).

There is little to choose between the remaining prediction systems and their detailed predictions seem very close.

5.2.4.10 System SYSEN Data

This data contains too many observations with zero-value which affect the u-plot distances of all prediction systems. Apart from JMNHPP, all prediction systems have the same u-plot distance which is highly significant as is that of JMNHPP (Table 5.16). All predictors are capturing the trend in the data as evident by their non-significant y-plot distance at 10% level.

Evidence of reliability growth is obtained from the predictive median plots (Figure 5.69). It is noticeable that JMNMHPP, D, LV and KL are giving more pessimistic predictions. However, the nature of

No.of Observations: = 45
Starting Sample Size : = 15

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	4.8212	8	3.6699	8	1.0581	7	.1220 N.S.	3	.1402 N.S.	3	7.6806 260.8545	2
BJM	3.8262	7	2.6846	7	1.0166	5	.0920 N.S.	1	.1467 N.S.	4	7.7924 261.1539	4
L	5.1404	9	4.1264	9	1.0629	8	.1320 N.S.	5	.1318 N.S.	2	7.6698 260.9169	3
BL	3.3569	6	2.5772	5	.9557	1	.1235 N.S.	4	.1665 N.S.	6	7.8289 260.8446	1
LV	2.3303	3	2.1061	3	.9889	3	.1738 N.S.	6	.2098 20%	9	7.6480 261.6604	6
KL	2.6128	4	2.3393	4	.9906	4	.1744 N.S.	7	.2046 20%	8	7.6436 261.7054	7
D	1.8958	2	1.6500	2	1.0472	6	.2432 5%	9	.2025 20%	7	7.5008 263.4670	9
JMNHPP	1.1297	1	1.0669	1	1.1397	9	.2859 2%	10	.2460 5%	10	7.5516 267.5985	10
LNHPP	3.1029	5	2.5898	6	.9735	2	.2024 20%	8	.1602 N.S.	5	7.5524 261.2601	5
W	9.3565	10	6.2689	10	2.2163	10	.1010 N.S.	2	.1215 N.S.	1	7.8794 262.1153	8

TABLE 5.15. The summarised results of system B7 HH60D 2FNS. The ML routine does not terminate normally for L, LV and KL. The calculations of BL involve overflo in some predictions.

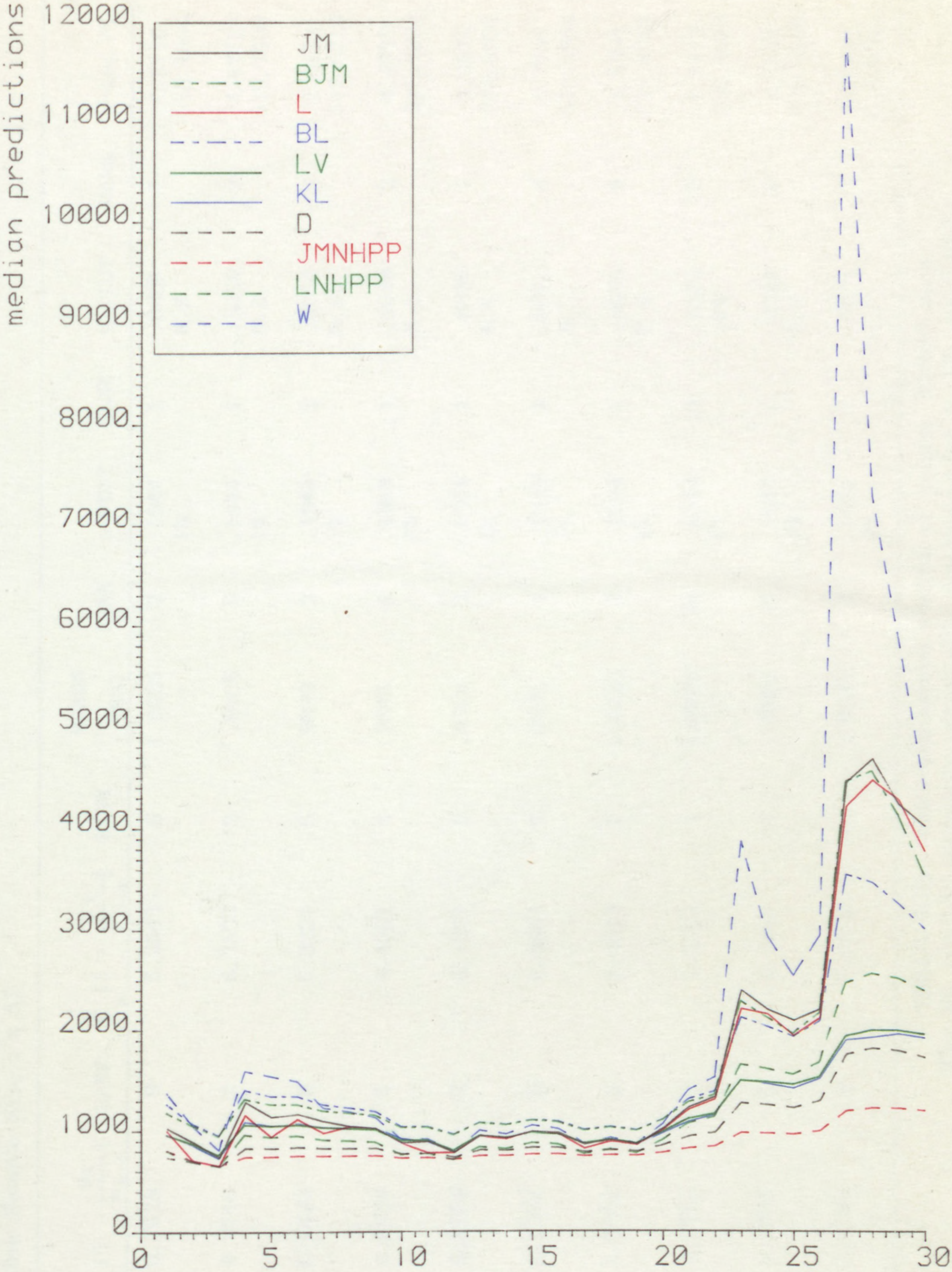


FIG.5.68. The plots of predictive medians for data-B7 HH60D 2FNS

Test Continuous Data System D-SYSEN

No.of Observations: = 159
Starting Sample Size : = 51

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	5.7628	9	5.3549	6	1.0025	7	.1944 1%	1	.0805 N.S.	7	4.5114 518.1664	4
BJM	4.7746	3	6.3633	9	.9905	5	.1944 1%	1	.0669 N.S.	3	4.4772 515.7593	2
L	5.6446	8	5.3227	5	.9838	3	.1944 1%	1	.0754 N.S.	6	4.5113 517.6213	3
BL	4.8670	4	4.4567	3	.9845	4	.1944 1%	1	.0640 N.S.	2	4.4772 515.7023	1
LV	5.5386	7	6.1959	8	.9709	2	.1944 1%	1	.0626 N.S.	1	5.3871 518.5931	6
KL	5.2951	6	6.0661	7	.9694	1	.1944 1%	1	.0672 N.S.	4	5.2798 519.2958	7
D	2.5982	2	2.4913	2	1.0228	8	.1944 1%	1	.0689 N.S.	5	4.5594 521.8581	9
JMNHPP	1.3725	1	1.3420	1	1.0892	10	.2149 1%	10	.1109 20%	10	4.5113 530.4926	10
LNHPP	5.1361	5	4.7365	4	.9961	6	.1944 1%	1	.0840 N.S.	8	4.5114 518.3128	5
W	7.7381	10	7.0685	10	1.0347	9	.1944 1%	1	.0871 N.S.	9	4.5594 519.6421	8

TABLE 5.16. The summarised results of system SYSEN data.
The ML routine terminate abnormally sometimes for L, BL and LNHPP.

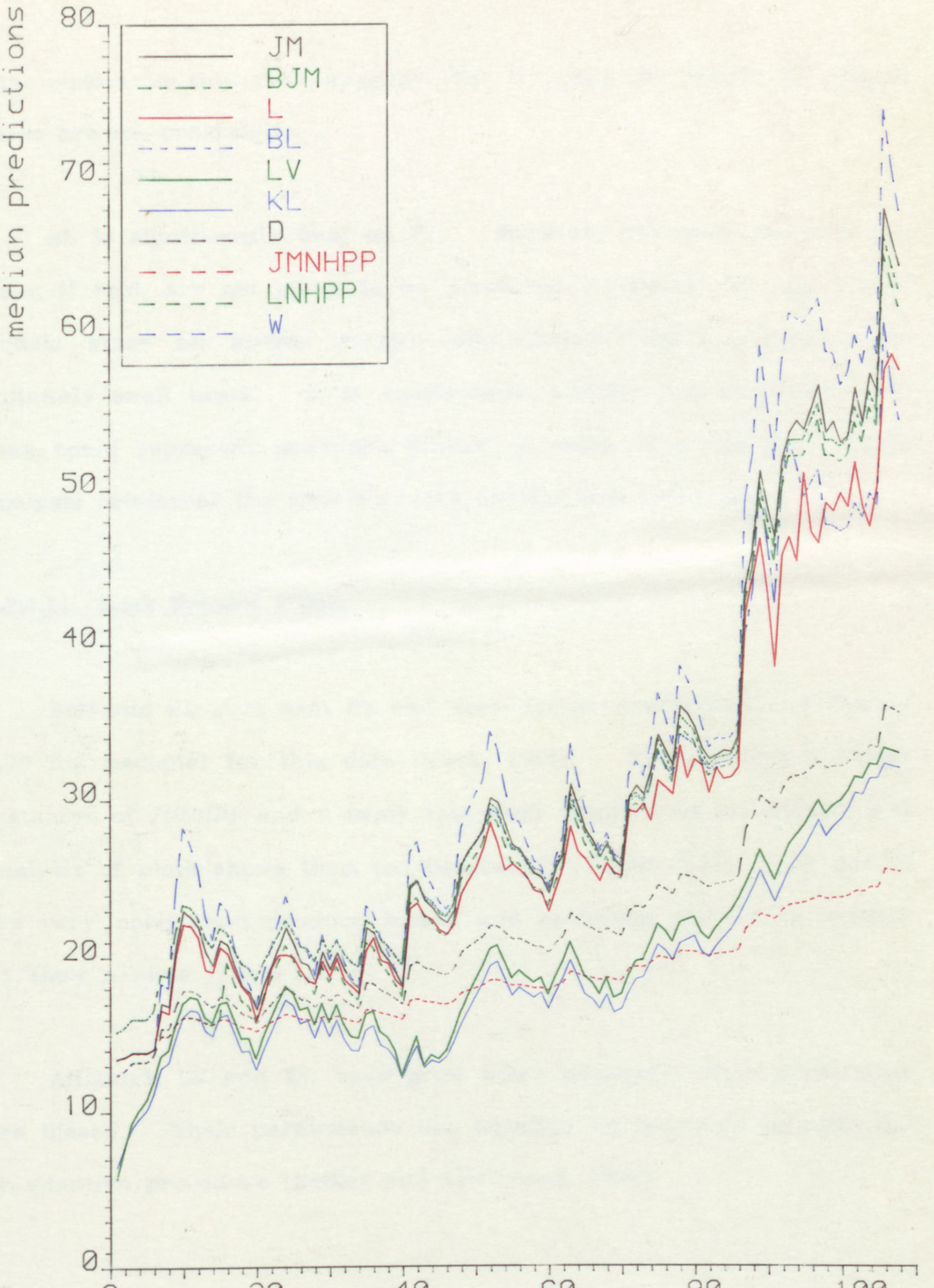


FIG.5.69. The plots of predictive medians for system SYSEN data

the u-plots in this case suggest that it would be unsafe to assume these are too pessimistic.

BL is significantly best on PL. However, the many zeros in the data, if real, are not going to be predicted accurately by <u>any</u> of the models since all models assign infinitesimally small probability to infinitely small times. It is questionable whether the zeros are real: they could represent immediate failure on retry of a bad fix. This analysis reinforces the need for very careful data collection.

5.2.4.11 Moek Project 1 Data

BJM and BL give best PL and very similar predictions (see Figure 5.70 for medians) for this data [Moek, 1984]. The significant u-plot distances of JMNHPP and D imply that their predictions are biased, and analysis of plots shows them too pessimistic (Figure 5.70). JM and W are very noisy, and produce biased and optimistic results as evident by their u-plots (Table 5.17).

Although LV and KL have good noise measures, their predictions are biased. Their performance can possibly be improved by applying an adaptive procedure [Keiller and Littlewood, 1984].

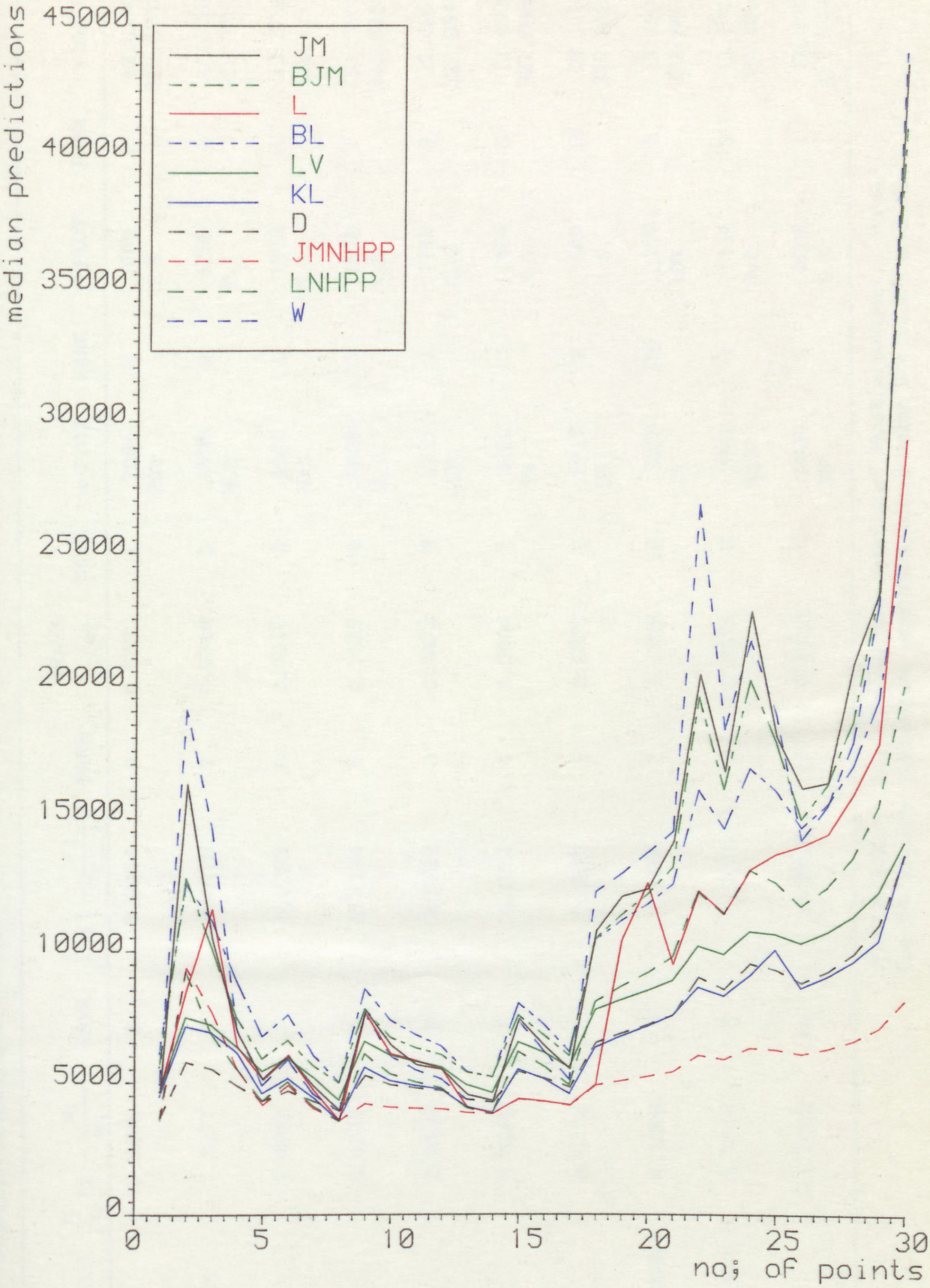


FIG.5.70. The plots of predictive medians for Moek's project 1 data

Test Continuous Data System D-DA1

No.of Observations: = 43
Starting Sample Size : = 13

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	10.3235	9	7.3464	9	0.6233	1	.2085 20%	6	.1059 N.S.	2	12.4190 320.5702	3
BJM	7.3377	7	5.0778	7	0.6348	2	.1596 N.S.	2	.1280 N.S.	4	12.0288 319.3761	1
L	7.6662	8	6.3033	8	0.7013	3	.1528 N.S.	1	.1376 N.S.	5	12.3789 322.1217	8
BL	6.2217	5	4.8124	5	0.7527	4	.1605 N.S.	3	.1575 N.S.	7	11.9003 319.5159	2
LV	3.8446	2	3.2812	3	0.9473	8	.2131 20%	7	.1747 N.S.	9	12.4202 321.8549	4
KL	4.5363	4	4.1251	4	0.9004	7	.2460 5%	8	.1494 N.S.	6	12.4900 321.8883	5
D	3.4484	1	2.9287	1	0.9883	9	.2717 2%	9	.1665 N.S.	8	13.1158 323.1427	9
JMNHPP	4.1204	3	3.1251	2	1.1928	10	.3329 1%	10	.2327 10%	10	13.1907 332.8487	10
LNHPP	6.7113	6	4.9473	6	0.8379	6	.1881 N.S.	4	.1134 N.S.	3	13.1616 321.8308	7
W	11.7462	10	8.3883	10	0.7125	5	.2069 20%	5	.1000 N.S.	1	12.2861 321.3446	6

TABLE 5.17. The summarised results of Moek's project 1 data.
The ML routine sometimes terminate abnormally for L, LV and KL.

5.2.4.12. Moek Project 2 Data

All models have significant u-plot distances; all but four have significant y-plot distances. It is these latter four which give the best PL.

These poor performances seem to be caused by a discontinuity in the data, occurring at about the 20th prediction. The smallest observation following that point is three times larger than the previous greatest observation.

None of the predictions here can be trusted. If the sudden change in the data can be explained, it might be the case that only data following the change should be used. No such information was available.

5.2.4.13. Braun-Paine GE1 Data

This data set is presented by Braun and Paine (1977). The analysis of this data set shows little reliability growth as evidenced by the predictive median plots (Figure 5.72). All u-plot and y-plot distances are non-significant at 10% level (Table 5.19).

PL performance as well as the median plots suggest that all predictions are very close to each other. It is not obvious to choose

Test Continuous Data System D-DA2

No.of Observations: = 159
Starting Sample Size : = 51

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	*	8	**	10	***	10	.3636 1%	6	.3939 1%	10	2.1863 –	10
BJM	*	8	5.0335	6	34.7584	9	.3069 1%	4	.2582 2%	8	2.3983 154.3278	9
L	7.6807	6	5.8542	8	2.3310	7	.2419 5%	1	.1140 N.S.	3	2.1846 138.6310	2
BL	4.0060	4	3.3139	4	1.0139	6	.3105 1%	5	.1003 N.S.	2	2.3859 138.7237	3
LV	2.5002	2	2.2542	2	0.9829	5	.5112 1%	8	.2054 20%	7	2.1592 146.1738	6
KL	2.4777	1	2.2355	1	0.9357	4	.5130 1%	9	.1870 20%	5	2.1617 146.2094	7
D	3.7535	3	3.1079	3	0.6141	2	.5615 1%	10	.0983 N.S.	1	2.1960 141.4016	4
JMNHPP	*	8	5.1079	7	29.5559	8	.2830 1%	3	.2765 2%	9	2.2008 151.1262	8
LNHPP	4.3329	5	3.5339	5	0.5815	1	.3976 1%	7	.1179 N.S.	4	2.1891 136.6748	1
W	11.3707	7	8.1966	9	0.8003	3	.2745 2%	2	.1895 20%	6	2.1959 141.8954	5

TABLE 5.18. The summarised results of Moek's project 2 data.
The ML routine sometimes terminate abnormally for L, LV and KL.

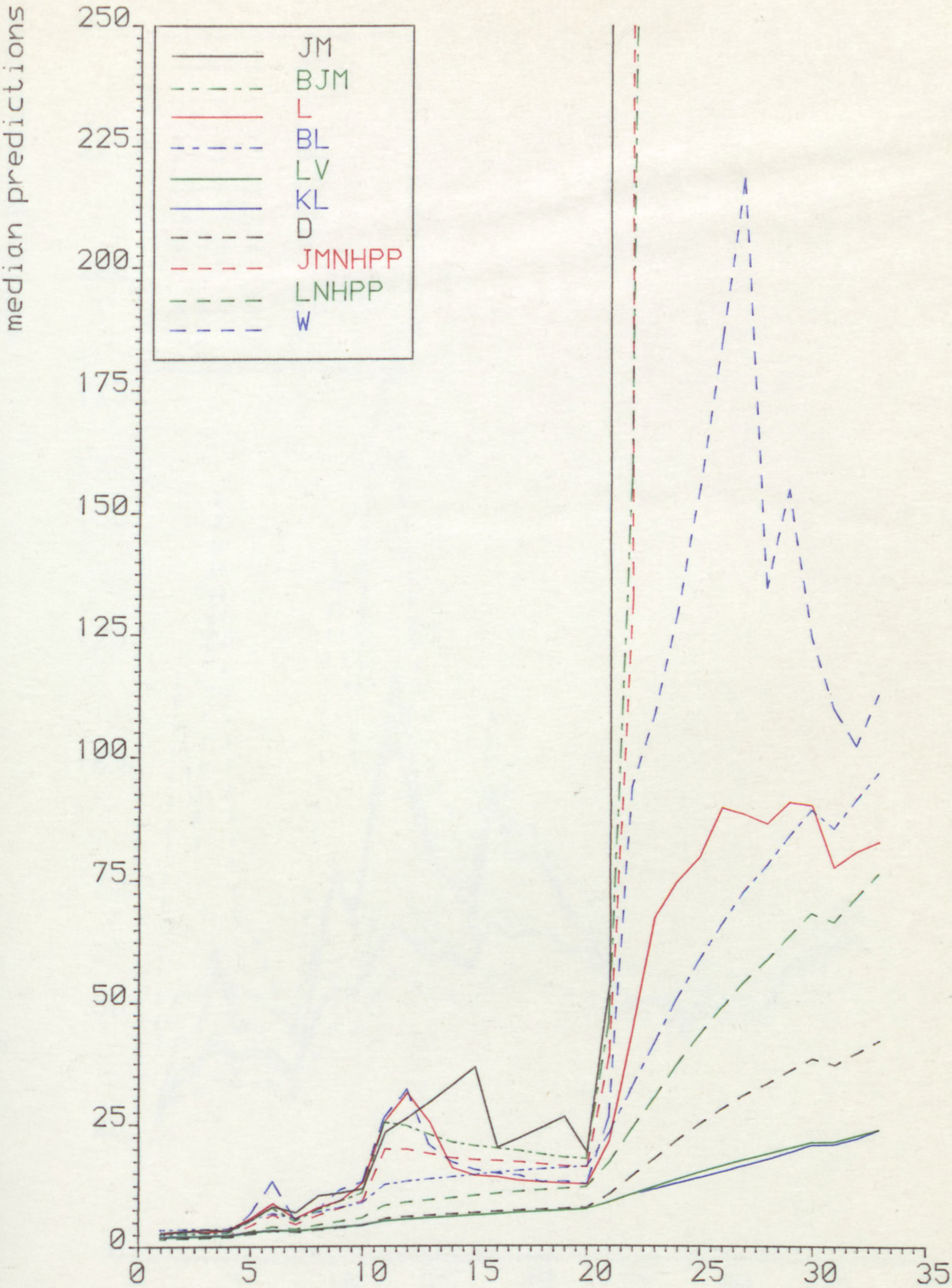


FIG.5.71. The plots of predictive medians for

Moek's project 2 data

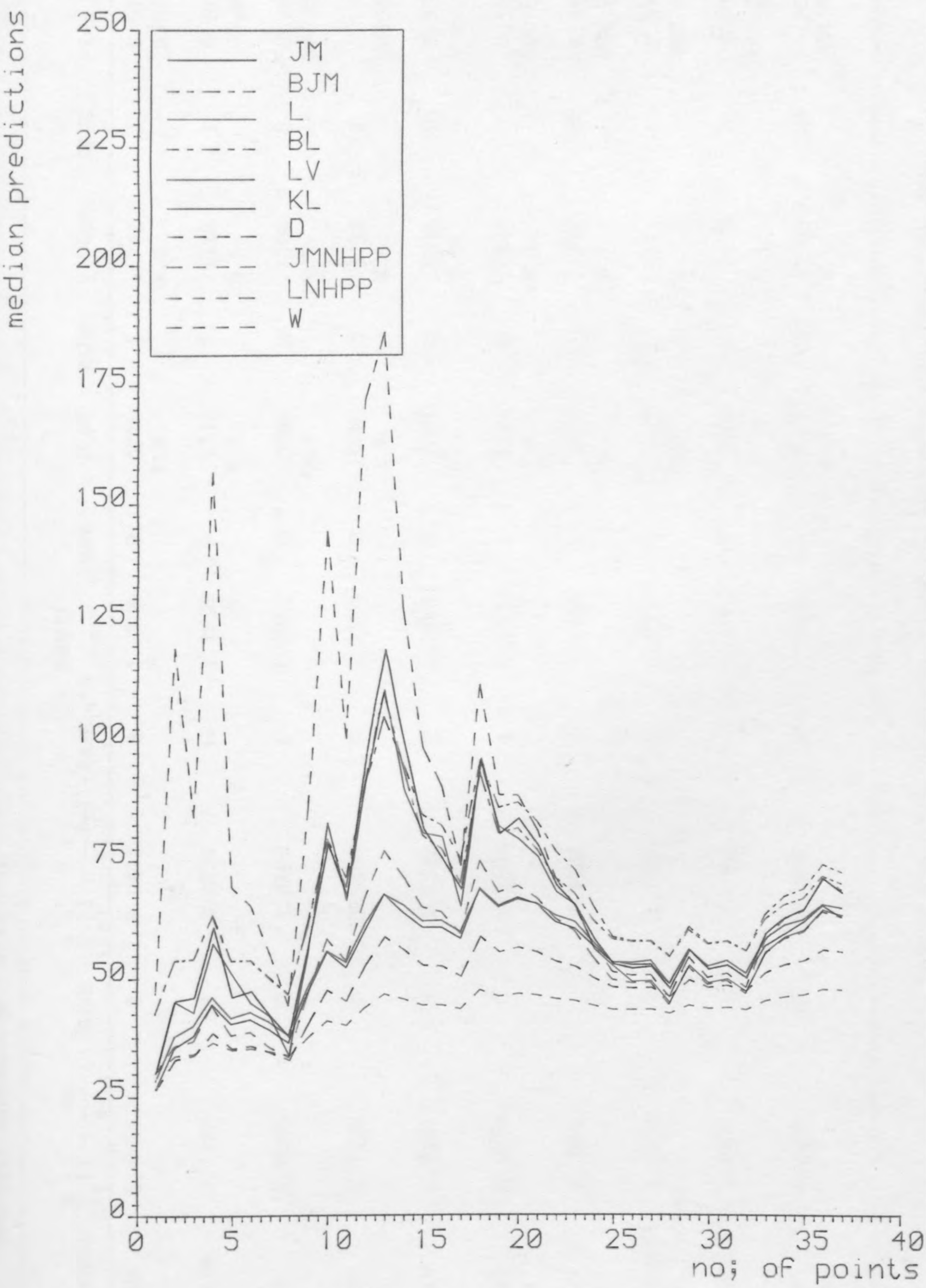


FIG.5.72. The plots of predictive medians for Braun&Paine GE-1 data

Test Continuous Data System D-GE1

No.of Observations: = 52
Starting Sample Size : = 15

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	5.4155	8	4.8472	8	1.3735	8	.0778 N.S.	1	.1347 N.S.	4	6.5115 203.1776	6
BJM	3.7177	7	3.1521	6	1.2326	6	.1046 N.S.	3	.1116 N.S.	1	6.1790 203.3556	7
L	5.6235	9	4.9198	9	1.3751	9	.0899 N.S.	2	.1304 N.S.	3	6.5115 203.6417	9
BL	3.5408	5	3.1062	5	1.2726	7	.1046 N.S.	3	.1119 N.S.	2	6.1790 203.4518	8
LV	2.7201	3	2.6438	3	1.1501	3	.1053 N.S.	5	.1380 N.S.	6	6.5378 202.3496	1
KL	2.9279	4	2.9285	4	1.1382	1	.1071 N.S.	6	.1480 N.S.	7	6.6211 202.6167	2
D	2.2671	2	2.0766	2	1.1499	2	.1651 N.S.	9	.1615 N.S.	9	6.7823 202.7873	4
JMNHPP	1.1534	1	1.1025	1	1.1595	4	.1894 20%	10	.1354 N.S.	5	6.5115 202.9216	5
LNHPP	3.5453	6	3.1946	7	1.1874	5	.1164 N.S.	7	.1533 N.S.	8	6.5115 202.6596	3
W	9.2897	10	8.2933	10	2.2530	10	.1215 N.S.	8	.1662 N.S.	10	5.9601 204.8924	10

TABLE 5.19. The summarised resudlts of Braun and Paine GE1 data.
The ML routine sometimes terminates abnormally for L and KL.
The calculation of BL involve overflo in some predictions

among the prediction systems in such situations, however, LV is slightly better in PL than all others. LV and KL predictions are very close.

5.2.4.14. System MDSIM Data

Table 5.20a and Figure 5.73a show the summarised results and predictive median plots of this data. In fact this data set contains a very large observation which affects the performance of all prediction systems. A comparison of Tables 5.20a and 5.20b and Figures 5.73a and 5.73b, which show the summarised results and the predictive median plots of this data with and without this particular observation, show the great effect it has on some models. These results show, again, the importance of careful data collection.

It is noticeable, however, that the effect varies greatly from one model to another. There is no clear effect in LV and KL performances while there is slight effect in D and LNHPP models. The performance of the remaining prediction systems was improved greatly by omitting this observation.

According to PL performance, LNHPP and BL are best performing models in both situations. But their y-plot distances are the best when the large observation is omitted. The poor performance of LV and KL is due to their biased and pessimistic predictions.

Test Continuous Data System D-MDSIM

No.of Observations: = 110
Starting Sample Size : = 45

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	BRAUN -Log PL	RANK
JM	7.1756	8	5.3797	8	1.9691	9	.1503 10%	5	.2399 1%	10	9.3332 595.5835	9
BJM	6.0766	6	4.3367	6	1.5290	8	.1525 10%	6	.2217 1%	8	9.3741 593.8734	5
L	7.4257	9	6.3917	9	1.3711	6	.1285 N.S.	4	.2024 1%	6	9.4489 591.6977	4
BL	5.8597	5	4.0167	5	1.1722	5	.0977 N.S.	1	.1637 10%	5	9.3723 589.6879	2
LV	1.6368	2	1.5684	2	1.0121	3	.1955 2%	8	.1297 N.S.	3	9.6304 594.4594	7
KL	1.5550	1	1.5277	1	1.0008	2	.2003 1%	9	.1000 N.S.	1	9.6424 589.7729	3
D	2.1267	3	1.9034	3	1.0215	4	.2054 1%	10	.1127 N.S.	2	9.7737 596.5922	10
JMNHPP	6.1374	7	4.4904	7	1.5267	7	.1272 N.S.	3	.2310 1%	9	9.4228 593.2627	6
LNHPP	2.8897	4	2.5514	4	0.9916	1	.1256 N.S.	2	.1457 20%	4	9.4242 589.5902	1
W	10.7693	10	7.3094	10	2.2872	10	.1618 10%	7	.2045 1%	7	9.1428 594.5641	8

TABLE 5.20a. The summarised results of system MDSIM data.
The ML routine does not terminate normally for L, LV, KL and LNHPP.

Test Continuous Data System D-MDSIM

No.of Observations: = 109
Starting Sample Size : = 45

MODEL	$\sum_{i=2}^{n} \left|1 - \frac{m_i}{m_{i-1}}\right|$	RANK	$\sum_{i=2}^{n} \left|1 - \frac{r_i}{r_{i-1}}\right|$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	4.9083	8	4.4961	8	1.1827	10	.0909 N.S.	4	.1149 N.S.	9	9.3332 571.9626	5
BJM	4.3409	6	3.8254	6	1.0698	8	.1002 N.S.	6	.1115 N.S.	7	9.3741 572.9803	7
L	5.2572	9	4.9475	9	1.0353	6	.0743 N.S.	1	.0952 N.S.	5	9.4489 571.0885	3
BL	3.7165	5	3.3272	5	.9282	2	.0904 N.S.	3	.0591 N.S.	1	9.3723 570.9162	2
LV	1.5302	1	1.4961	1	.9805	4	.1888 2%	8	.1139 N.S.	8	9.6303 574.1616	9
KL	1.5701	2	1.7617	3	.9638	3	.1936 2%	9	.0933 N.S.	4	9.6424 573.0126	8
D	1.8118	3	1.7246	2	1.0038	5	.2208 1%	10	.1003 N.S.	6	9.7737 576.3963	10
JMNHPP	4.4392	7	3.9760	7	1.0674	7	.0764 N.S.	2	.1160 N.S.	10	9.4228 571.6527	4
LNHPP	2.5529	4	2.3957	4	.9161	1	.1345 20%	7	.0726 N.S.	2	9.4242 570.5298	1
W	6.0733	10	5.6574	10	1.1409	9	.0954 N.S.	5	.0914 N.S.	3	9.1428 572.2261	6

TABLE 5.20b. The summarised results of system MDSIM data.
The ML routine does not terminate normally for L, LV, KL and LNHPP.
Omits one large observation = 33340.

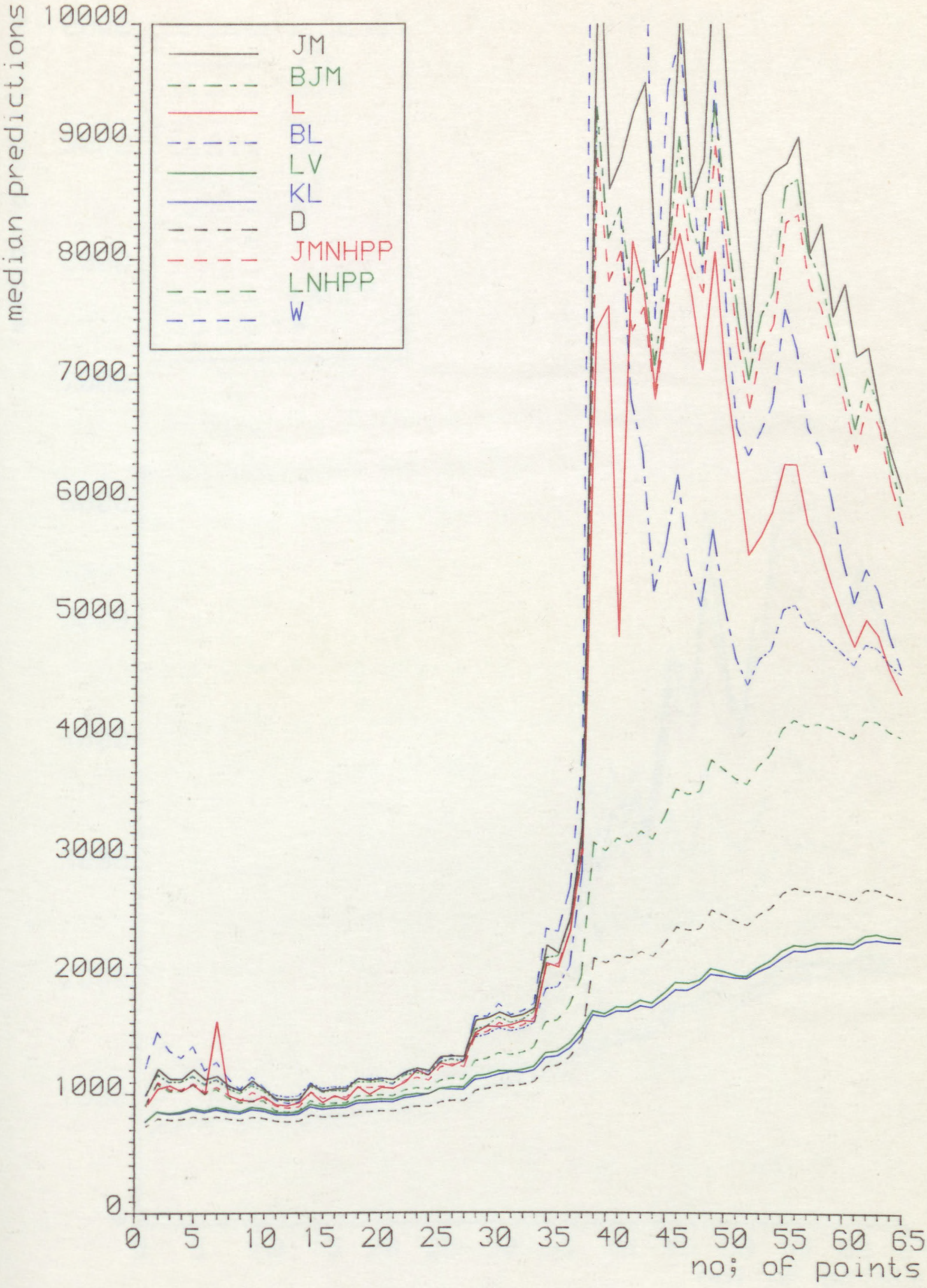


FIG.5.73a. The plots of predictive medians for system MDSIM data

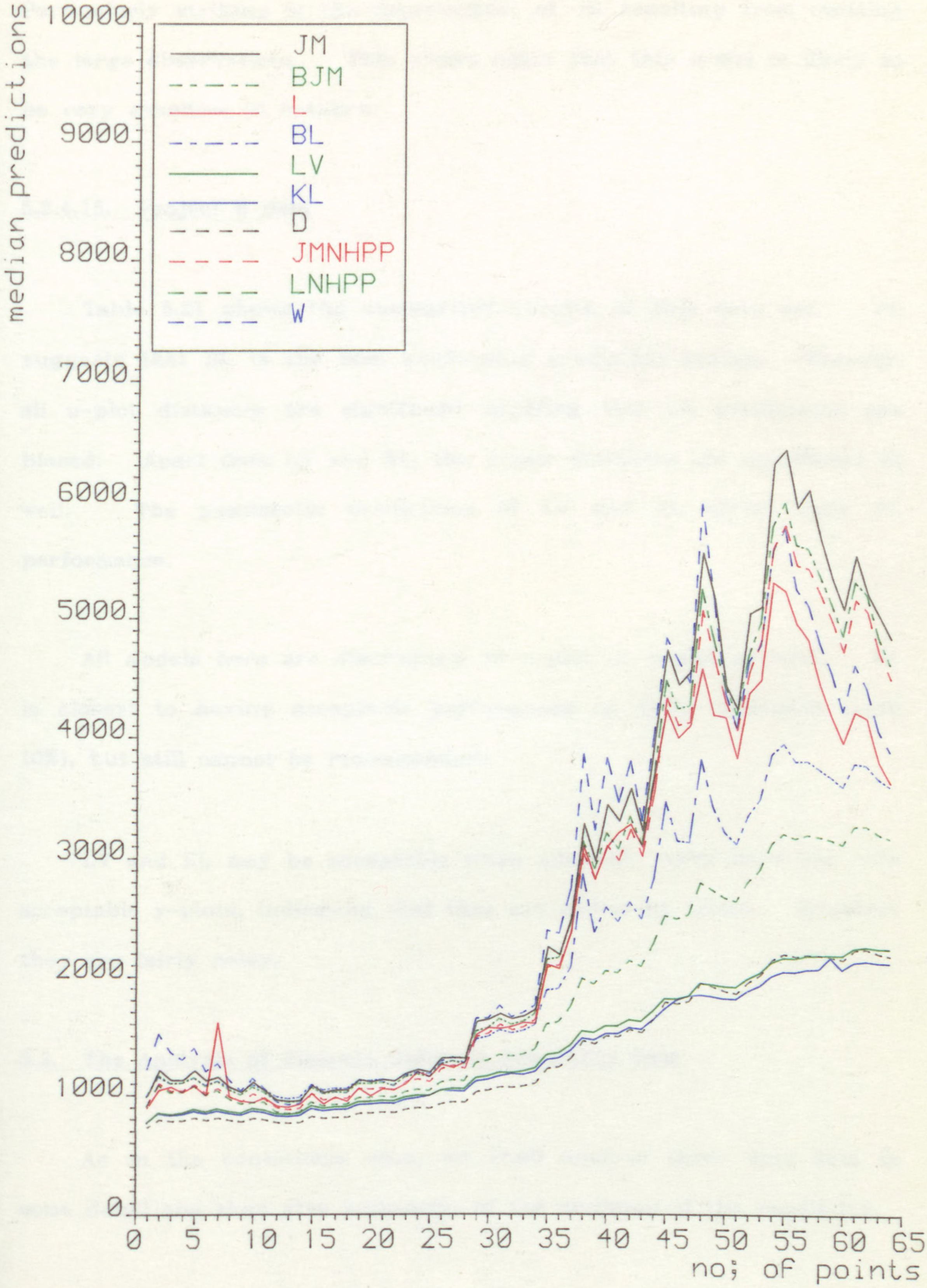


FIG.5.73b. The plots of predictive medians for system MDSIM data without the large obs; which equal to 33340.

Particularly striking is the improvement of JM resulting from omitting the large observations. This shows again that this model is likely to be very sensitive to outliers.

5.2.4.15. Project B Data

Table 5.21 shows the summarised results of this data set. PL suggests that BL is the best performing prediction system. However, all u-plot distances are significant implying that all predictions are biased. Apart from LV and KL, the y-plot distances are significant as well. The pessimistic predictions of LV and KL affect their PL performance.

All models here are discredited on u-plot or y-plot or both. BL is closest to having acceptable performance on these measures (both 10%), but still cannot be recommended.

LV and KL may be acceptable when adapted: they have the only acceptable y-plots, indicating that they are capturing trend. However, they are fairly noisy.

5.3. The Analysis of Discrete Software Reliability Data

As in the continuous case, we shall analyse three date sets in some detail and then give summaries of the analyses of the remainder.

Test Continuous Data System D-PB

No.of Observations: = 168
Starting Sample Size : = 50

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	u-PLOT	RANK	y-PLOT	RANK	-Log PL	RANK
JM	6.5528	5	5.0640	5	1.2061	9	.1373 5%	4	.1836 1%	8	3.2673 624.4777	4
BJM	6.3755	4	4.8656	4	1.0969	7	.1410 2%	5	.1471 2%	4	3.3554 615.4024	2
L	25.9014	10	24.0092	10	1.0715	5	.1239 5%	2	.1554 1%	5	3.2573 624.9961	5
BL	6.8022	6	5.2499	6	0.9368	2	.1185 10%	1	.1274 10%	3	3.3554 613.4539	1
LV	10.1568	8	9.7484	8	0.9941	4	.2817 1%	10	.0925 N.S.	1	2.3351 639.7976	8
KL	8.7539	7	7.7941	7	0.9346	1	.2538 1%	9	.0953 N.S.	2	2.5742 639.1233	7
D	3.4886	2	3.1237	2	0.9673	3	.1674 1%	6	.1610 1%	6	3.0997 642.8100	9
JMNHPP	1.8398	1	1.7375	1	1.0765	6	.2394 1%	8	.2430 1%	9	3.2673 669.8266	10
LNHPP	6.2536	3	4.7274	3	1.0971	8	.1292 5%	3	.1752 1%	7	3.2673 623.2801	3
W	18.8375	9	10.4845	9	2.0592	10	.1734 1%	7	.2491 1%	10	3.0997 633.4722	6

TABLE 5.21. The summarised results of Project B data.
The ML routine does not terminate normally for L.

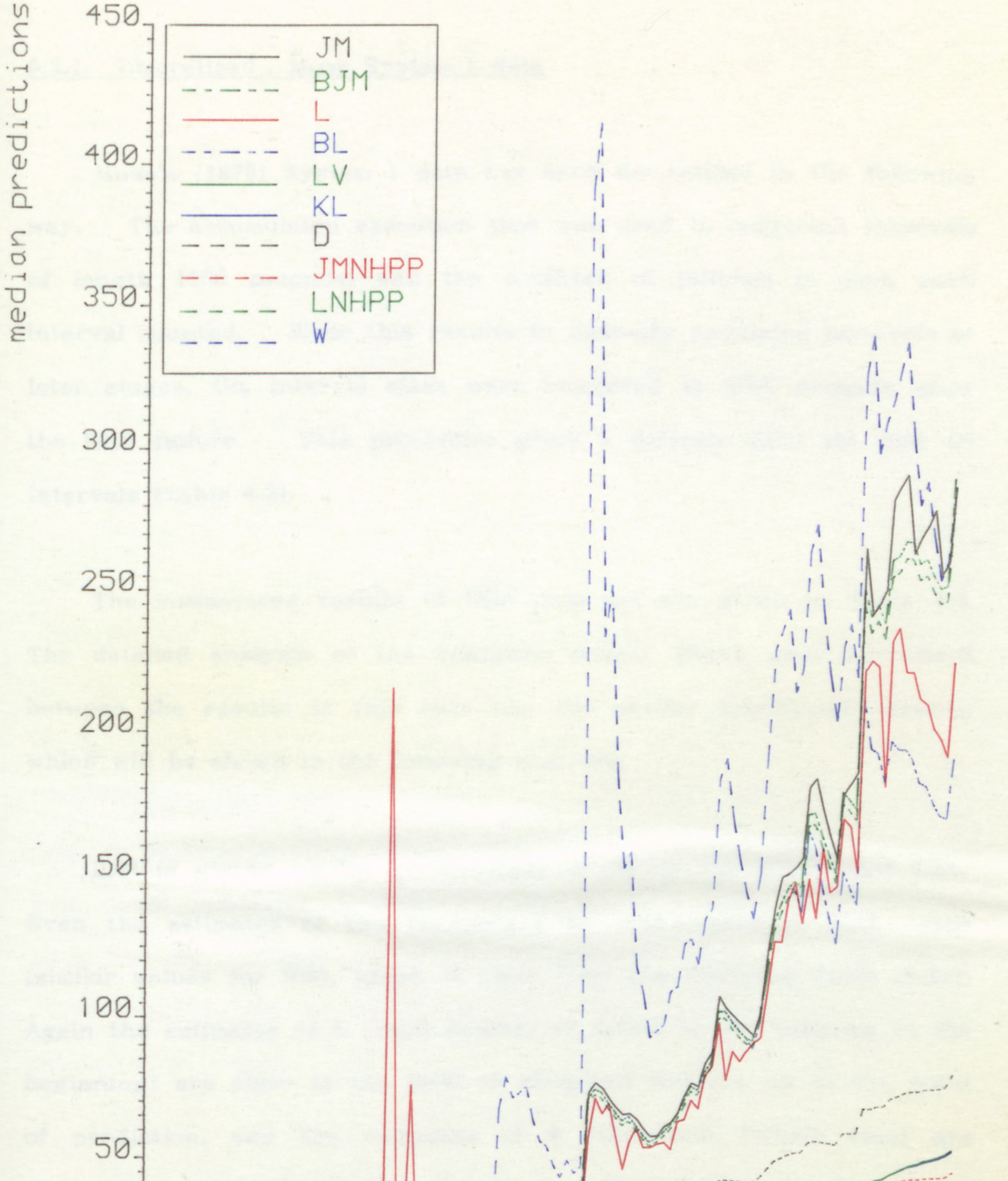


FIG.5.74. The plots of predictive medians for project B data

5.3.1. Discretized Musa System 1 data

Musa's (1979) System 1 data has been discretized in the following way. The accumulated execution time was used to construct intervals of length 1000 seconds, and the numbers of failures in each such interval counted. Since this results in sparsely populated intervals at later stages, the interval sizes were increased to 2000 seconds after the 80th failure. This procedure gives a discrete data set with 55 intervals (Table 4.3).

The summarized results of this data set are given in Table 4.4. The detailed analysis of the computer output shows good agreement between the results in this case and the earlier continuous version, which will be shown in the following analysis.

DJM is giving similar performance to JM on the continuous data. Even the estimates of the parameters are very close in both cases (similar values for $\hat{N}, \hat{\Phi}$), which is clear from the following Table (5.22). Again the estimates of N (total number of faults in the program at the beginning) are close to the total of observed failures up to the point of prediction, and the estimates of Φ (the fault failure rate) are decreasing as the sample size increases. Thus these estimates cannot be trusted.

	DJM			JM	
Sample size	$\hat{N}$	$\hat{\Phi}$	Sample size	$\hat{N}$	$\hat{\Phi}$
13	80	0.000105	62	84	0.000102
20	99	0.000076	79	102	0.000076
30	102	0.000069	97	104	0.000070
40	134	0.000039	122	138	0.000038
50	139	0.000036	132	141	0.000036
54	141	0.000035	135	142	0.000035

Table 5.22. The estimates of DJM parameters and those of JM at (approximately) the same points.

The u-plot distance is significant at 1% level, and a detailed study shows that the plots are everywhere below the line of unit slope (Figures 5.76),i.e. the model is overestimating the cdf. This means it is overestimating the reliability. (Note: large cdf of inter-event time implies small reliability in continuous case, large cdf of failure count in discrete case implies large reliability). The noise statistics based on medians and rates show great noiseness in DJM predictions compared with other prediction systems. Also the Chi-square value is very large and significant at 1% level. All this results in the model being discredited in PL, where it has the worst performance among the prediction systems.

The u-plot distance of DJMNHPP is also highly significant and the u-plot is everywhere below the line of the unit slope (Figure 5.78) implying that the predictions of DJMNHPP are also too optimistic. The predictions of both JM and DJM models are close as evidenced by

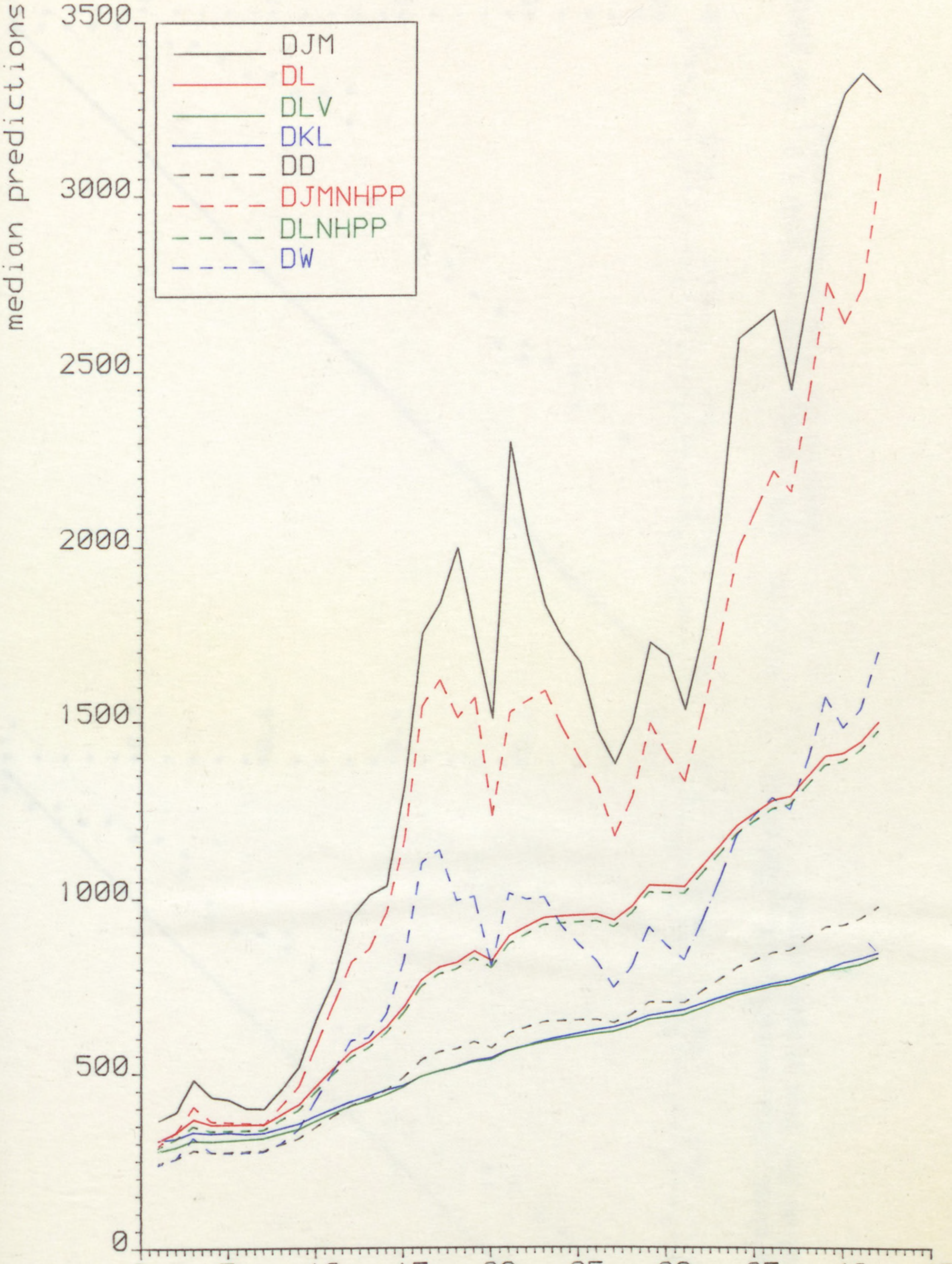


FIG.5.75. The plots of predictive median time to 1st failure in each prediction interval for the data in Table 4.3

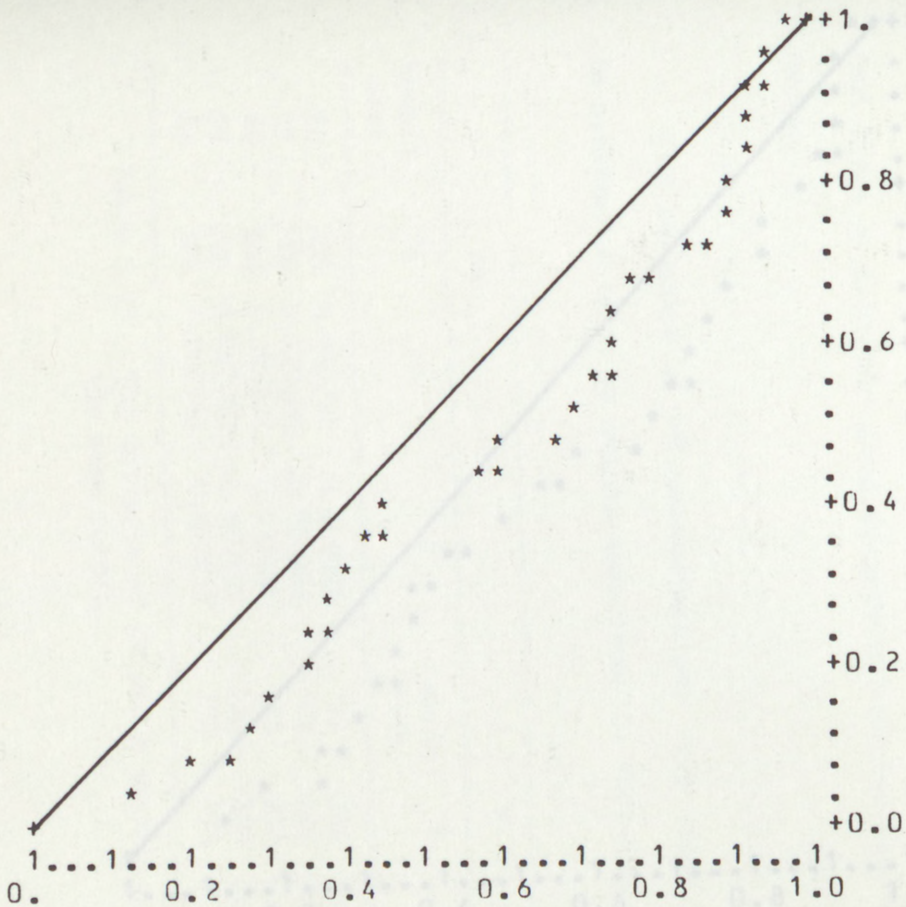


FIG.5.77. DL u-plots, data in Table 4.3., the plots based on the line printer output.

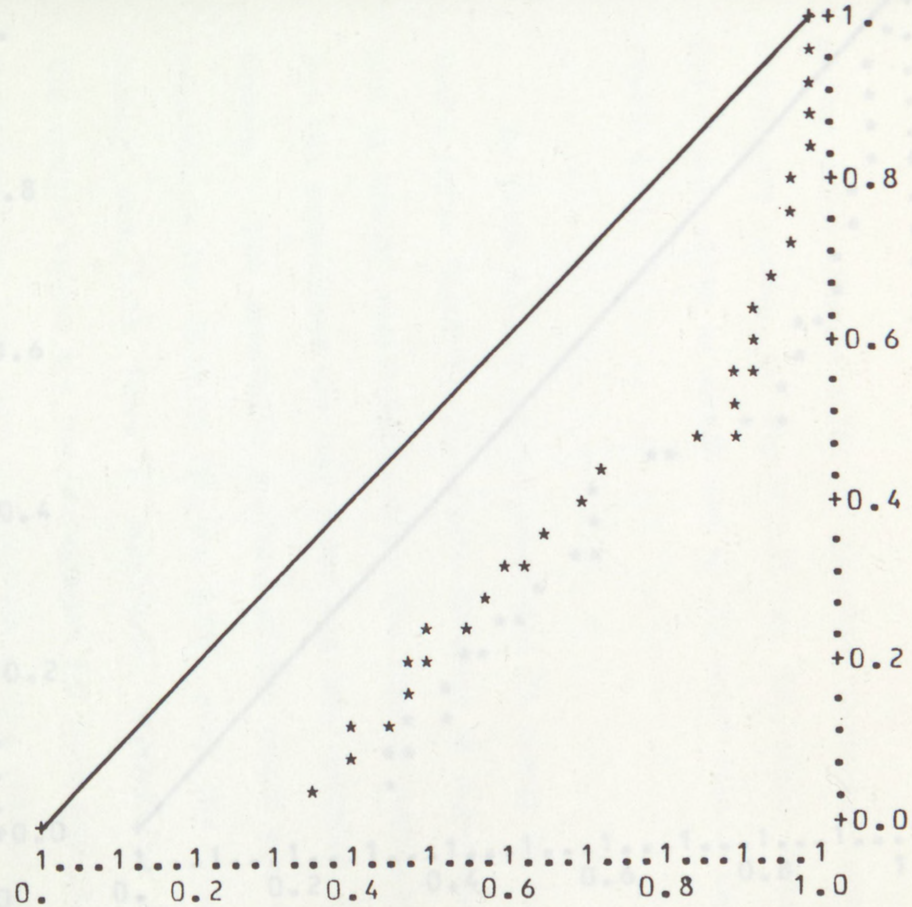


FIG.5.76. DJM u-plots, data in Table 4.3., the plots based on the line printer output.

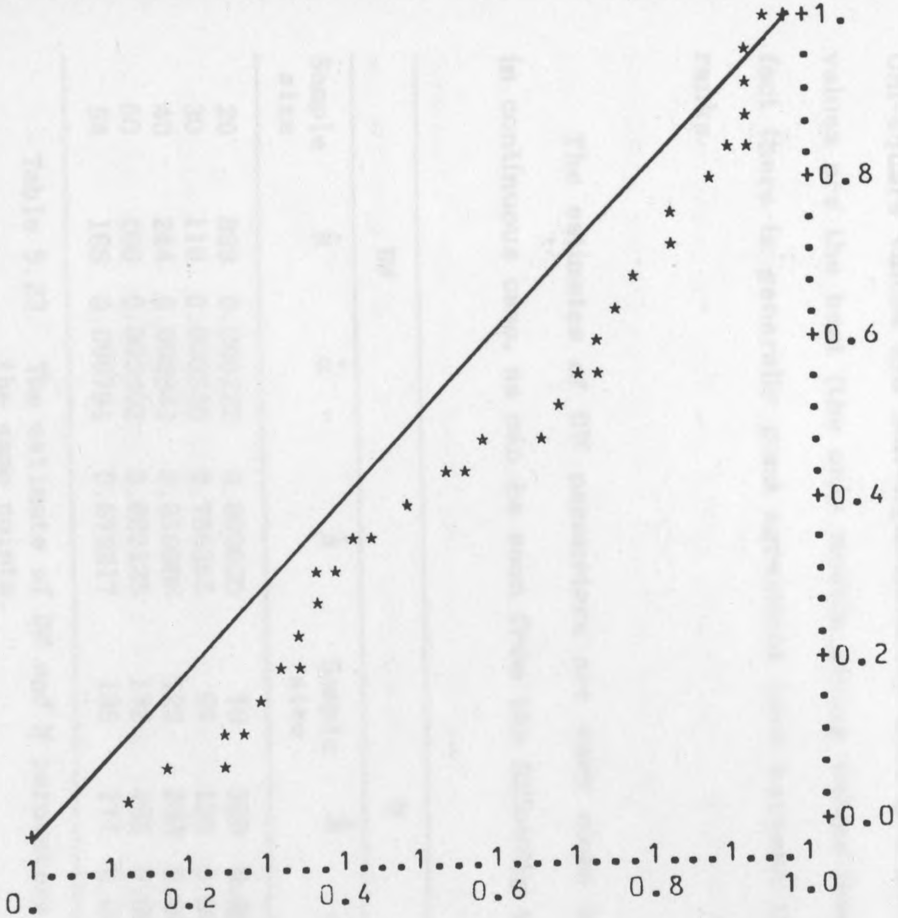


FIG.5.79. DLNHPP u-plots data in Table 4.3., the plots based on the line printer output.

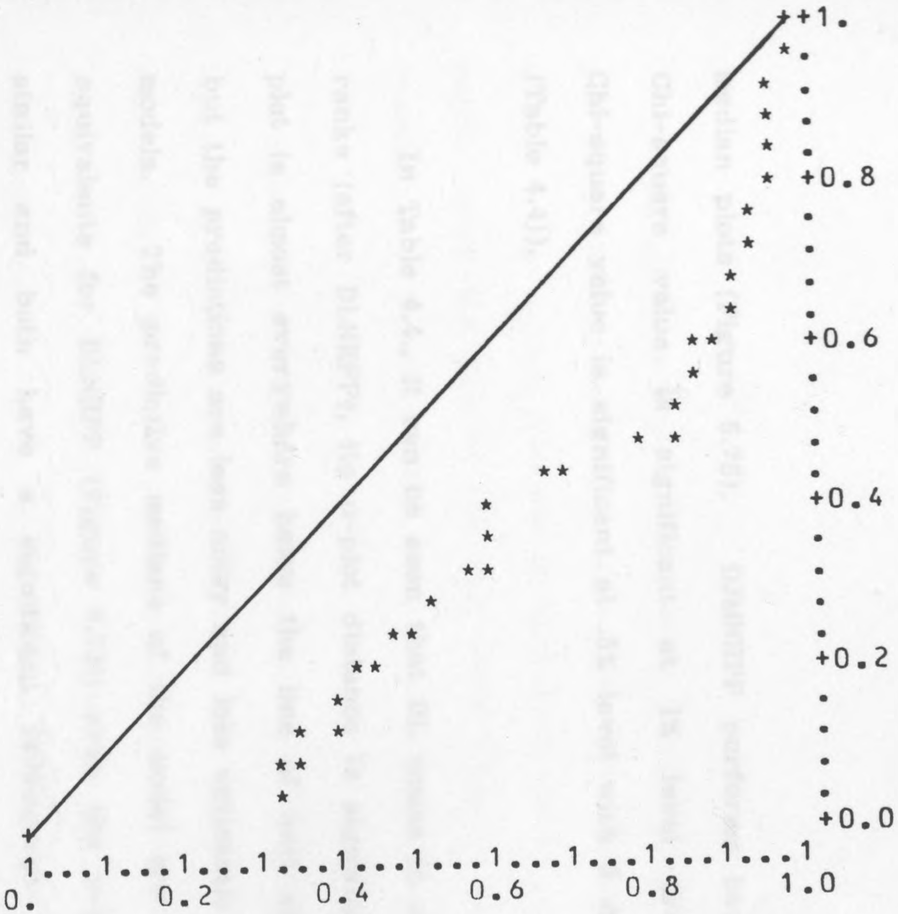


FIG.5.78. DJMNHPP u-plots, data in Table 4.3, the plots based on the line printer output.

median plots (Figure 5.75). DJMNHPP performs badly in PL and its Chi-square value is significant at 1% level (even the grouped Chi-square value is significant at .5% level with 9 degrees of freedom (Table 4.4)).

In Table 4.4., it can be seen that DL comes in second place in PL ranks (after DLNHPP), its u-plot distance is significant at 5% and the plot is almost everywhere below the line of unit slope (Figure 5.77), but the predictions are less noisy and less optimistic than the previous models. The predictive medians of the model are very close to the equivalents for DLNHPP (Figure 5.75); even the u-plots of both look similar and both have a significant Kolmogorov distance. The Chi-square values are non-significant for both and their Braun statistic values are the best (the only models giving values less than one). In fact there is generally good agreement here between the Braun and PL ranks.

The estimates of DW parameters are very close to their estimates in continuous case, as can be seen from the following table.

DW				W			
Sample size	$\hat{N}$	$\hat{\alpha}$	$\hat{\beta}$	Sample size	$\hat{N}$	$\hat{\alpha}$	$\hat{\beta}$
20	893	0.000222	0.609025	79	320	0.000357	0.675454
30	119	0.000549	0.756255	97	120	0.000548	0.758828
40	244	0.000841	0.610086	122	238	0.000671	0.636723
50	166	0.000693	0.682125	132	182	0.000684	0.670819
54	168	0.000701	0.679217	135	171	0.000662	0.683607

Table 5.23. The estimate of DW and W parameters at approximately the same points.

DW predictions show a lot of noise (more than DL and DLNHPP but less than DJM and DJMNHPP), as is clear from the median plot (Figure 5.75). The u-plot distance is significant at 5% and the plots are mainly below the line of the unit slope (Figure 5.81), i.e. optimistic predictions. Its Chi-square value is non-significant.

The remaining prediction systems - DD, DLV and DKL - are least noisy. Their u-plot distances and Chi-square values are non-significant but their u-plots are mainly above the line of the unit slope (Figures 5.80, 5.82 and 5.83) indicating that their predictions are too pessimistic.

The picture so far, then, is that we can eliminate DJM, DJMNHPP, and DW because these models exhibit significant "bias" as evidenced by their u-plots, as well as being too noisy. Comparison between the remaining models shows some apparent contradictions. DL and DLNHPP are best in PL and Braun statistic but they exhibit "biased" predictions. On the other hand DLV, DKL and DD seem to be unbiased, and are best on the Chi-square criterion, but are significantly inferior in their PL performance.

The observed numbers of failures per unit time, as well as the predictions of the expected numbers of failures per unit time for DL, DLV, DKL and DLNHPP are plotted in Figures 5.84.

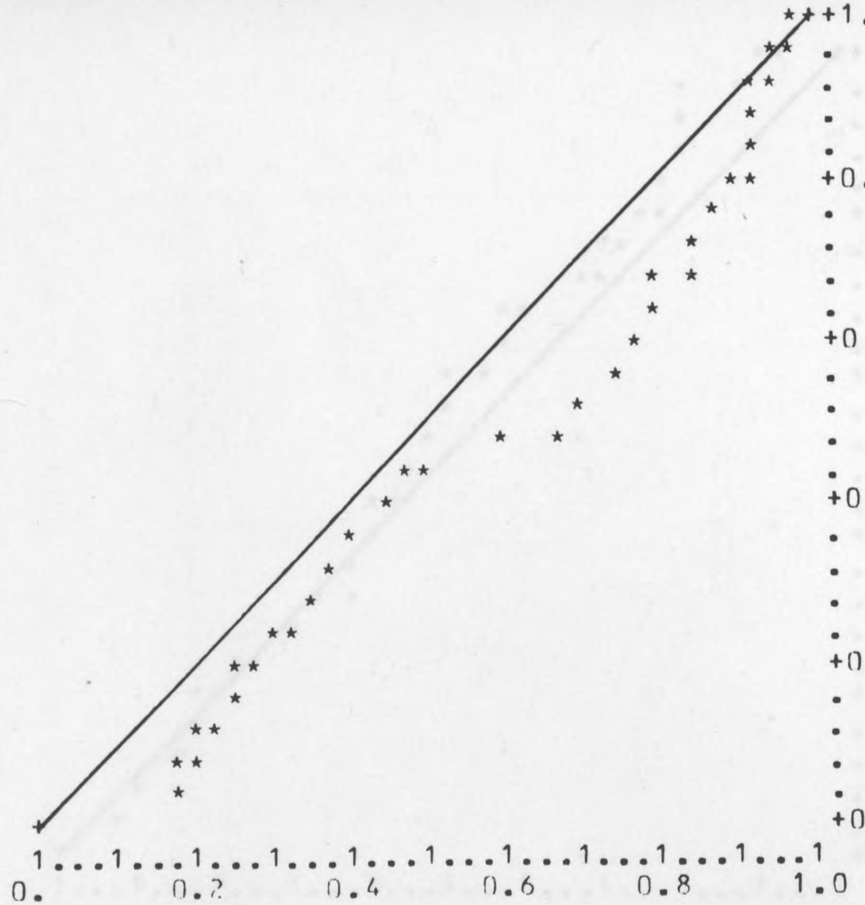


FIG.5.81. DW u-plots, data in Table 4.3., the plots based on the line printer output.

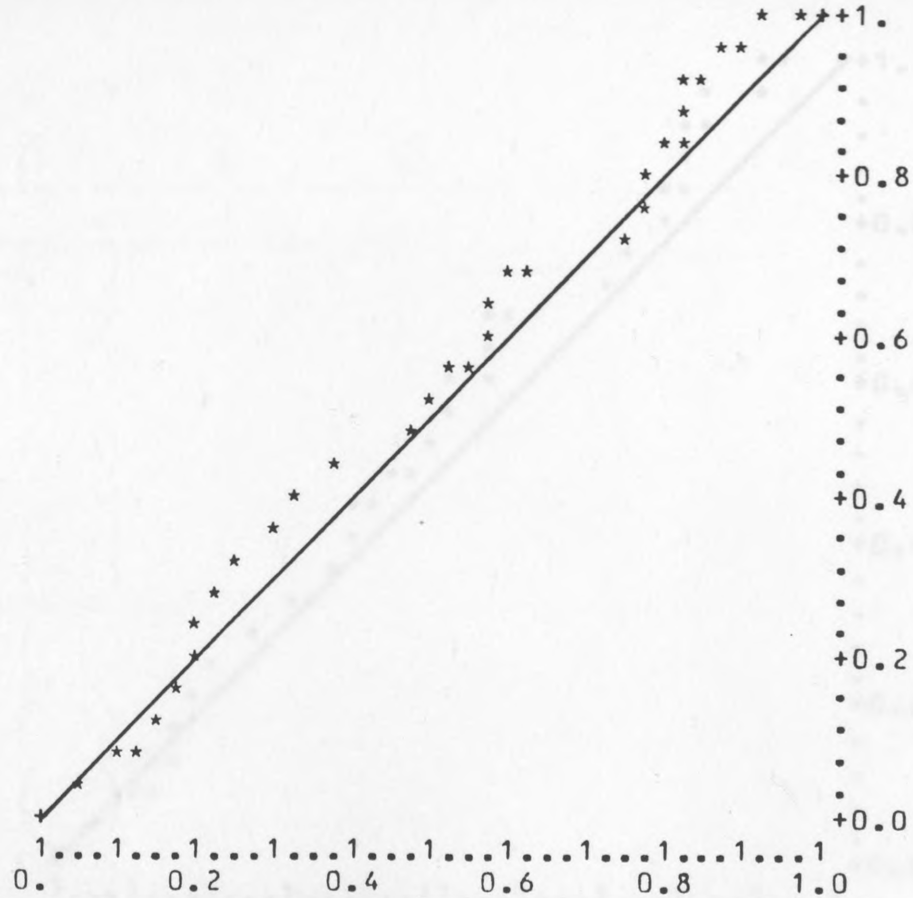


FIG. 5.80. DD u-plots, data in Table 4.3, the plots based on the line printer output.

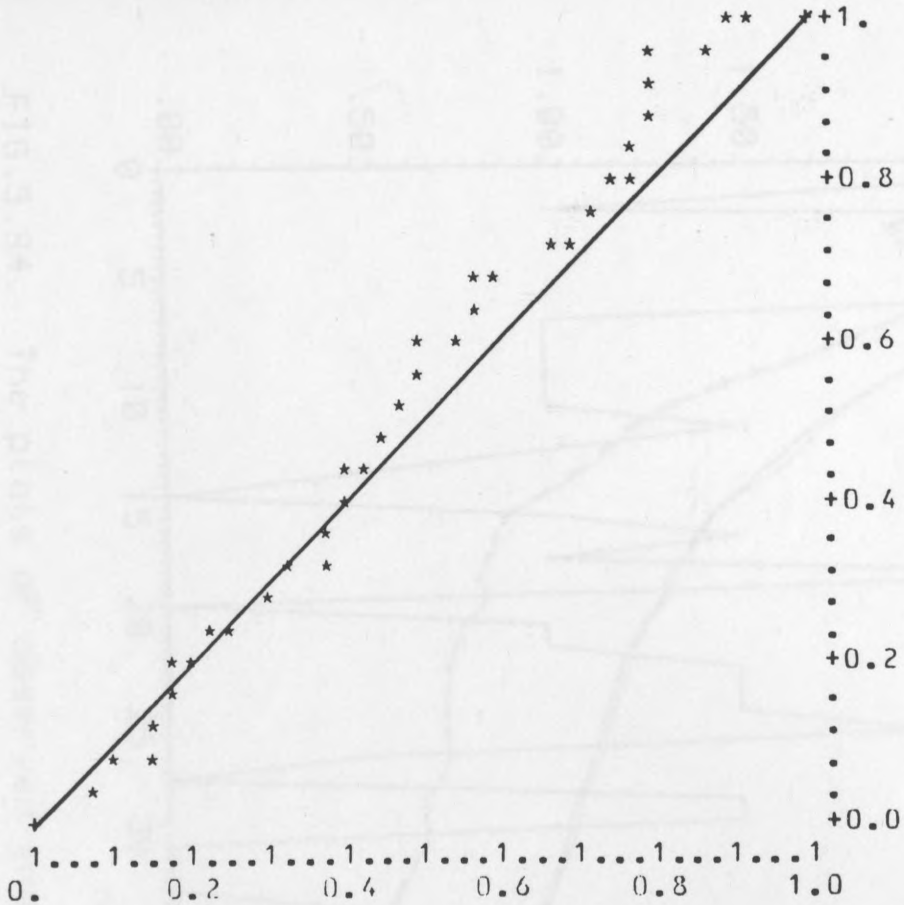
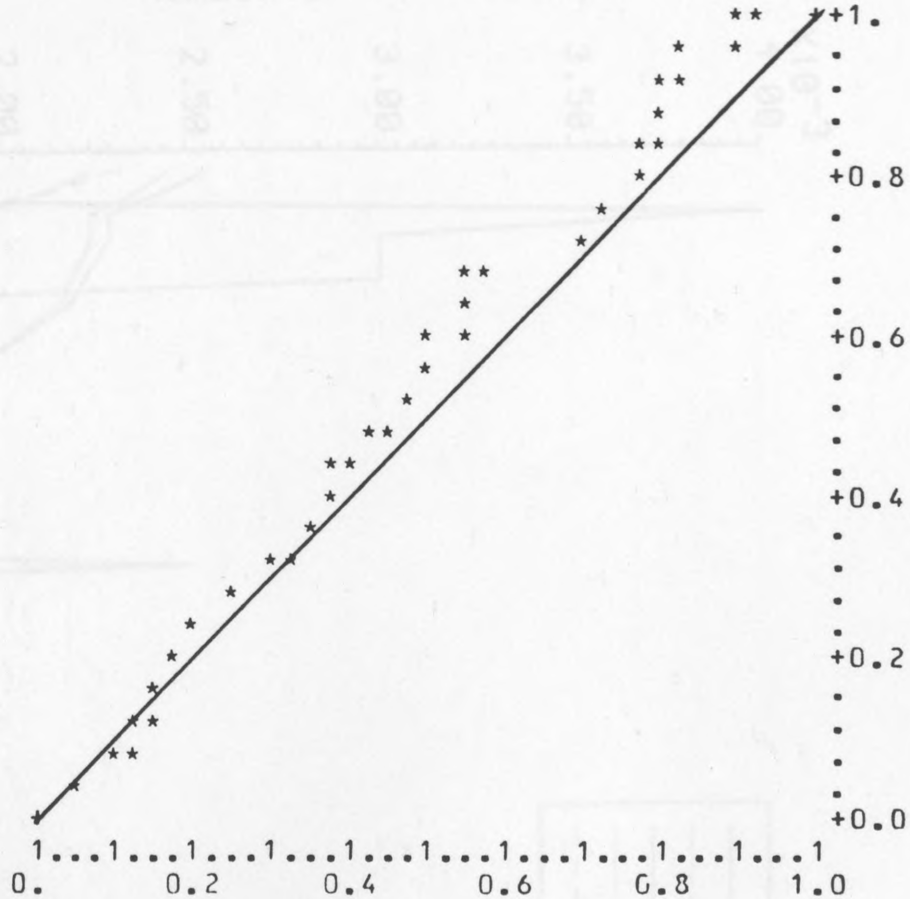


FIG.5.83. DKL u-plots, data in Table 4,3., the plots based on the line printer output.



FIG,5,82. DLV u-plots, data in Table 4.3, the plots based on the line printer output.

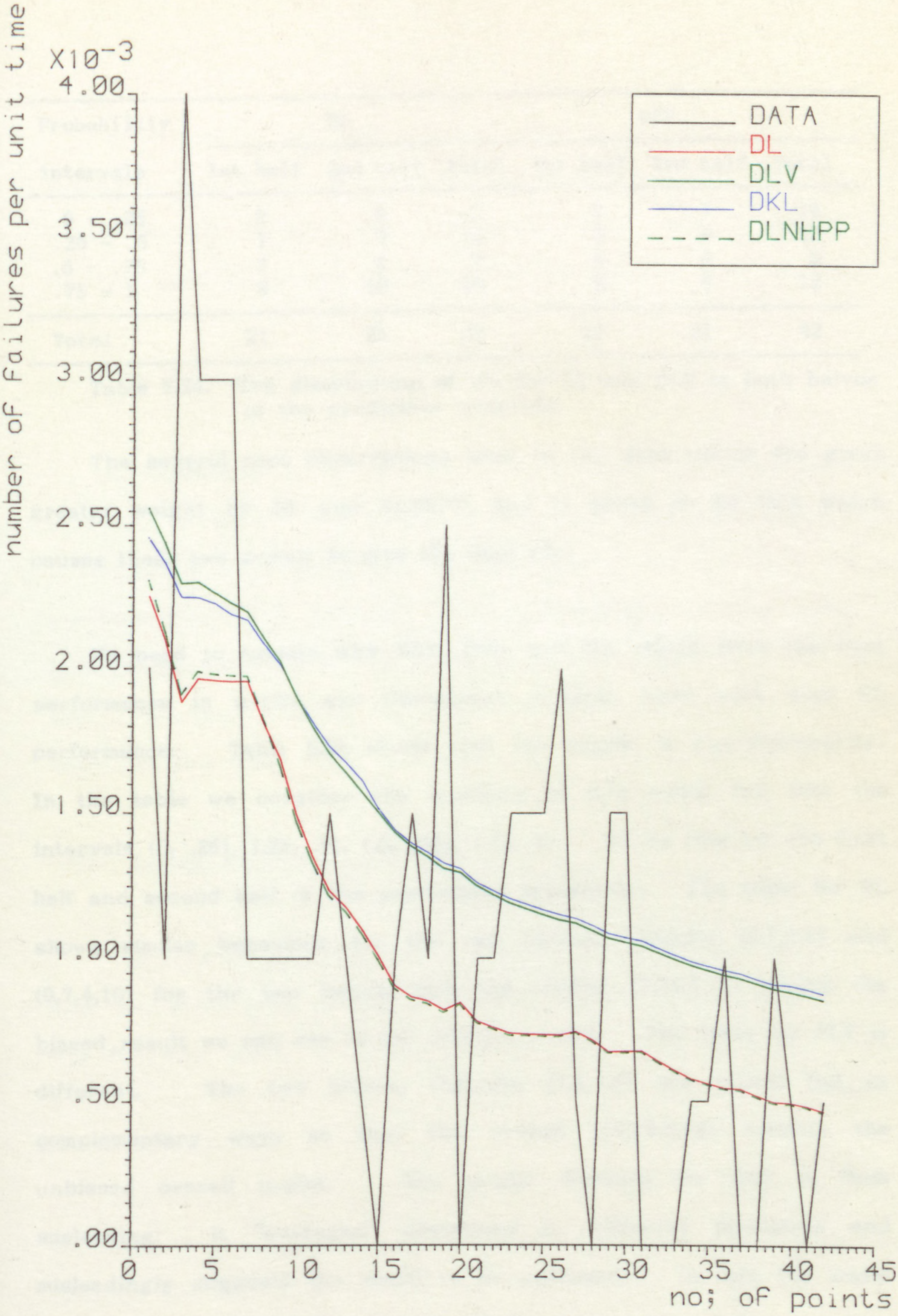


FIG.5.84. The plots of observed and predictive expected numbers of failures per time unit for the data in Table 4.3

Probability	DL			DLV		
intervals	1st half	2nd half	Total	1st half	2nd half	Total
0 - .25	2	0	2	3	7	10
.25 - .5	7	7	14	8	4	12
.5 - .75	3	4	7	4	4	8
.75 - 1	9	10	19	6	6	12
Total	21	21	42	21	21	42

Table 5.24. The distribution of u's for DL and DLV in both halves of the prediction intervals.

The several zero observations later in the data vector are given greater weight by DL and DLNHPP, and it seems to be this which causes these two models to give the best PL.

We need to explain why DLV, DKL, and DD, which have the best performance in u-plot and Chi-square criteria, have such poor PL performance. Table 5.24 shows that the reason is non-stationarity. In the table we consider the numbers of u_i's which fall into the intervals (0, .25), (.25, .5), (.5, .75), (.75, 1). We do this for the first half and second half of the predictions separately. The table for DL shows similar behaviour for the two halves: counts (2,7,3,9) and (0,7,4,10) for the two halves with the overall (2,14,7,19) giving the biased result we can see in the detailed u-plot. The table for DLV is different. The two halves, (3,8,4,6), (7,4,4,6) are biased but in complementary ways so that the overall (10,12,8,12) reflects the unbiased overall u-plot. The u-plot distance for DLV is thus misleading: it "averages" deviations in different directions and misleadingly suggests the model to be unbiased. In fact the model has a varying bias; it is not capturing the trend accurately.

It is notable that this changing bias is operating only for the left-hand portion of the predictive distributions. The median predictions, for example, seem consistently good. This is reflected in the very good Chi-square performance which is based solely on means.

The above comments also apply to DKL and DD.

To summarise, then, we again have a trade-off between different criteria. If we are only interested in predicting the expected number of failures in a future interval, we would prefer DD, DLV, DKL. If we are interested in a prediction of the distribution of the number of failures we might prefer DL, DLNHPP even though these show some bias. The possibility of developing a discrete adaptive procedure for cases like this could be an interesting topic for investigation.

5.3.2. System P8751 Data (AD-P758)

The original data set was in the form of continuous data. The same procedure used to discretize Musa's System 1 data (section 5.3.1.) was followed, giving a discrete data set containing 50 points (Table 5.25).

The estimated values of N obtained by DJM increased, and the estimates of Φ decreased, as the sample size increased. DJM and DJMNHPP are very close in their parameter estimates (Table 5.26).

I	NR(I)	TN(I)	TP(I)
1	12	0.	1000.00
2	11	12.00	1000.00
3	12	23.00	1000.00
4	9	35.00	1000.00
5	11	44.00	1000.00
6	7	55.00	1000.00
7	5	62.00	1000.00
8	14	67.00	1000.00
9	10	81.00	1000.00
10	4	91.00	1000.00
11	2	95.00	1000.00
12	5	97.00	1000.00
13	4	102.00	1000.00
14	5	106.00	1000.00
15	2	111.00	1000.00
16	2	113.00	1000.00
17	3	115.00	1000.00
18	4	118.00	1000.00
19	2	122.00	1000.00
20	3	124.00	1000.00
21	2	127.00	1000.00
22	5	129.00	1000.00
23	3	134.00	1000.00
24	1	137.00	1000.00
25	5	138.00	1000.00
26	8	143.00	2000.00
27	3	151.00	2000.00
28	7	154.00	2000.00
29	7	161.00	2000.00
30	4	168.00	2000.00
31	2	172.00	2000.00
32	4	174.00	2000.00
33	5	178.00	2000.00
34	4	183.00	2000.00
35	11	187.00	2000.00
36	6	198.00	2000.00
37	6	204.00	2000.00
38	3	210.00	2000.00
39	0	213.00	2000.00
40	7	213.00	2000.00
41	3	220.00	2000.00
42	5	223.00	2000.00
43	6	228.00	2000.00
44	2	234.00	2000.00
45	1	236.00	2000.00
46	3	237.00	2000.00
47	3	240.00	2000.00
48	1	243.00	2000.00
49	8	244.00	2000.00
50	8	252.00	2000.00

TABLE 5.25. System P8751 data (AD-P758), read left to right test interval, numbers of failures inside the interval, total observed numbers of failures at the beginning of the interval, and the interval length in time unit.

This closeness is reflected in their predictions (Figure 5.85). The u-plot distance is significant at 1% level for both. The plots themselves are faraway below the line of the unit slope (Figures 5.86 and 5.88) implying that their predictions are too optimistic. This optimism affects the calculation of noise, Chi-square and Braun statistic and their PL performance as well (Table 5.27).

Sample	DJM		DJMNHPP	
size	$\hat{N}$	$\hat{\Phi}$	$\hat{\mu}$	$\hat{\Phi}$
20	150	0.000089	151.241	0.000092
25	168	0.000073	169.578	0.000074
30	200	0.000054	203.689	0.000053
35	232	0.000041	238.471	0.000039
40	256	0.000034	261.690	0.000033
45	271	0.000031	274.453	0.000031
49	288	0.000028	292.517	0.000027

Table 5.26. MLE's for DJM and DJMNHPP obtained by using the data shown in Table 5.25.

DL and DLNHPP exhibit "bias" in their predictions as manifested by their significant u-plot distances. The u-plots are everywhere below the line of the unit slope (Figures 5.87 and 5.89) indicating that their predictions are optimistic.

DW is the noisiest prediction system on this data set (Table 5.27). The model exhibits a significant Chi-square value and u-plot distance and its predictions are too optimistic (Figure 5.90).

No.of Observations: = 50
Starting Sample Size: = 20

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	2.0782	6	1.9829	6	1.6149	8	135.363	94.772	.4709	8	1.3075	
							.1%	.1%(12)	1%		91.1080	8
DL	1.8972	5	1.8121	5	1.2180	5	76.966	44.389	.3453	5	1.3196	
							.1%	.1%(14)	1%		77.1474	5
DLV	.7391	2	.7035	2	.9866	1	41.049		.1221	3	1.8821	
							N.S.		N.S.		70.0116	2
DKL	.7328	1	.6945	1	.9885	2	40.582		.1069	2	1.8741	
							N.S.		N.S.		69.1755	1
DD	.9077	3	.8727	3	1.0509	3	41.690		.1005	1	2.0229	
							N.S.		N.S.		71.5934	3
DJMNHPP	2.2255	7	2.1115	7	1.5268	7	120.674	62.644	.4475	7	1.3104	
							.1%	.1%(13)	1%		87.7770	7
DLNHPP	1.6351	4	1.5859	4	1.1758	4	73.021	42.860	.3253	4	1.3107	
							.1%	.1%(14)	1%		76.4439	4
DW	2.6027	8	2.5720	8	1.3433	6	92.549	54.234	.3821	6	1.3191	
							.1%	.1%(14)	1%		81.8563	6

TABLE 5.27. The summarized results ofthe data shwon in Table 5.25.
The ML routine does not always terminate normally for DL.

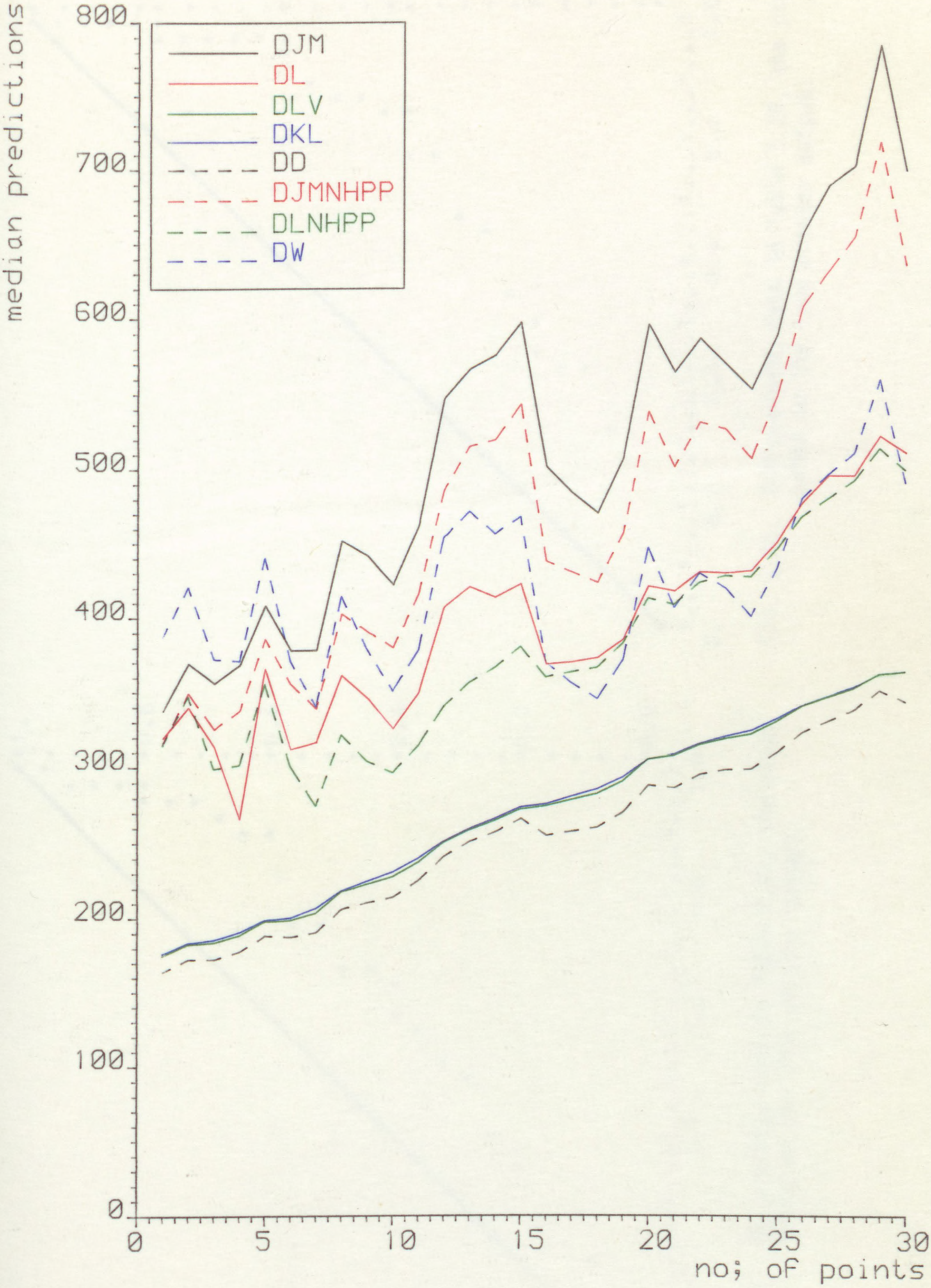


FIG.5.85. The plots of predictive median time to 1st failure in each prediction interval for the data in Table 4.25

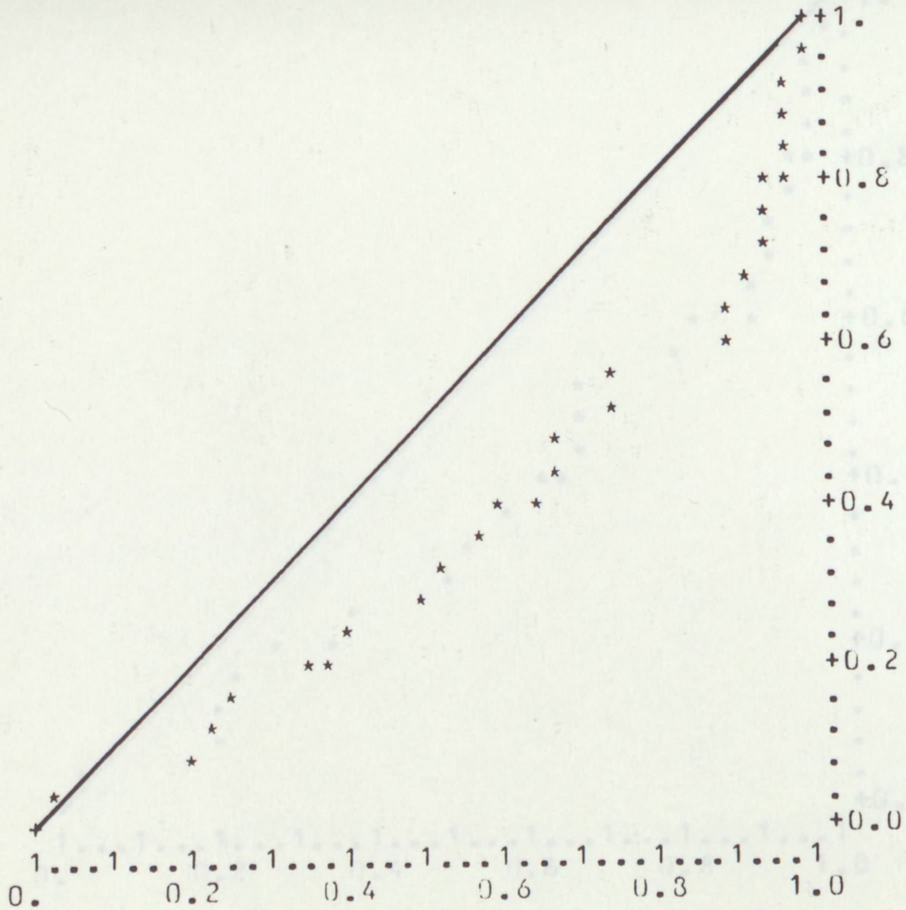


FIG.5.87. DL u-plots, data in Table 5.25, the plots based on the line printer output.

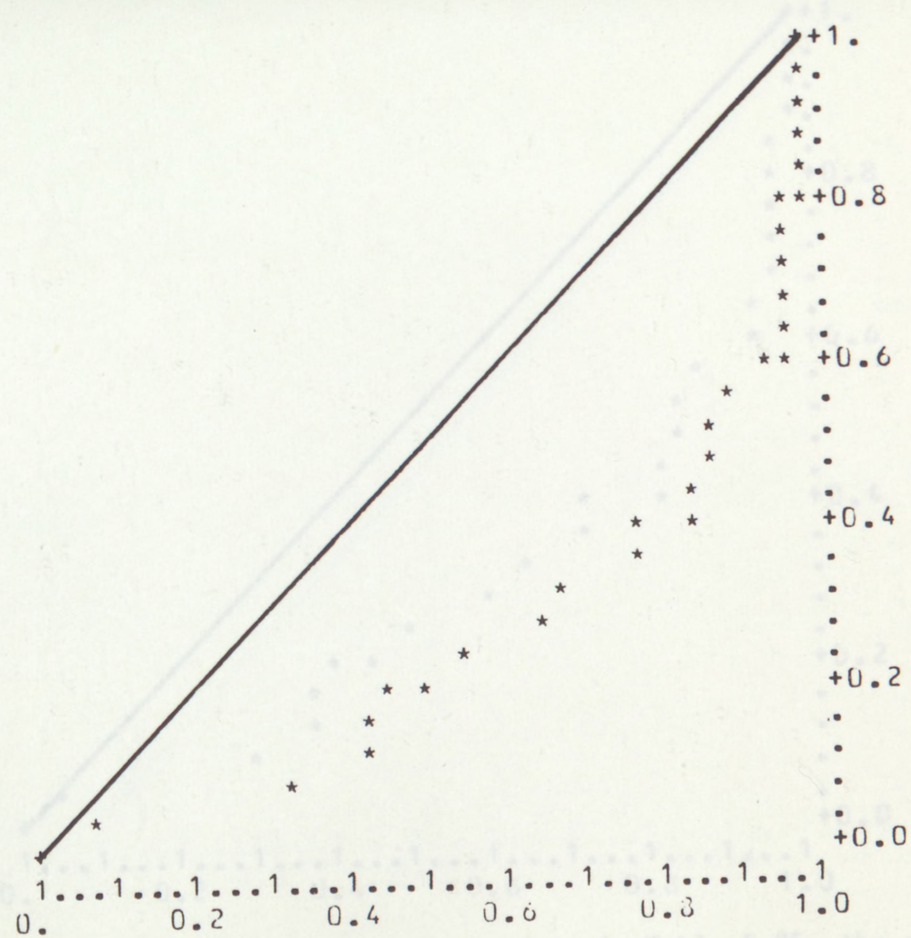


FIG. 5.86. DJM u-plots, data in Table 5.25, the plots based on the line printer output.

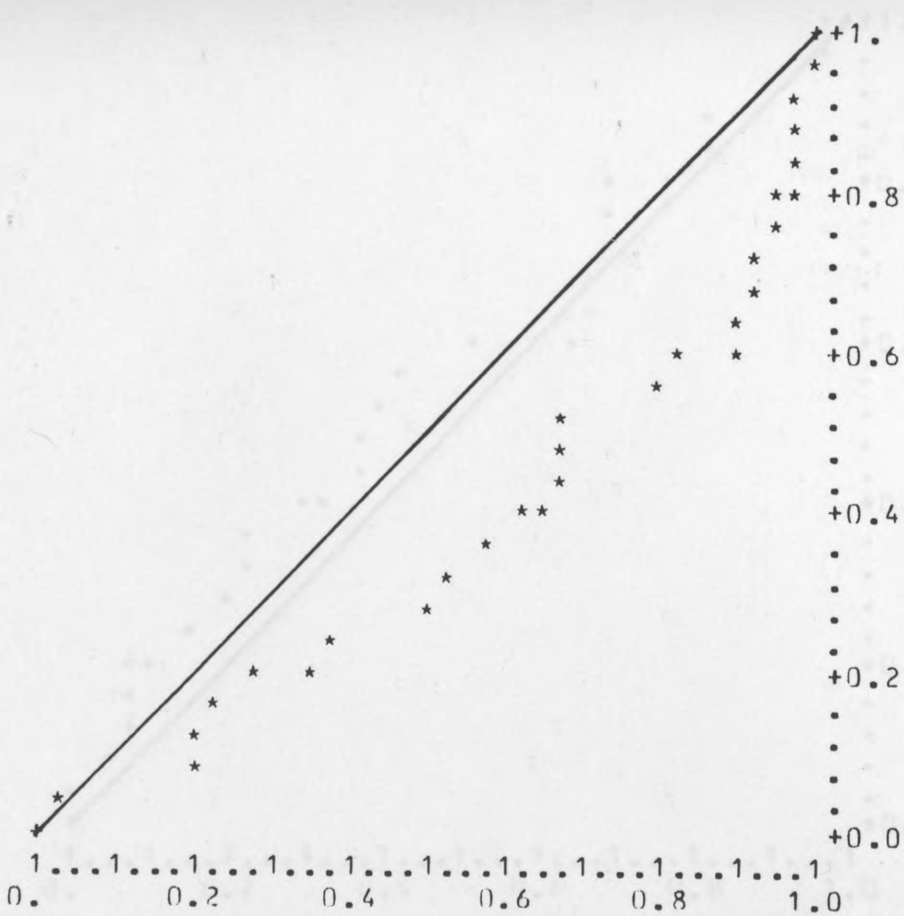


FIG.5.89. DLNHPP u-plots, data in Table 5,25, the plots based on the line printer output.

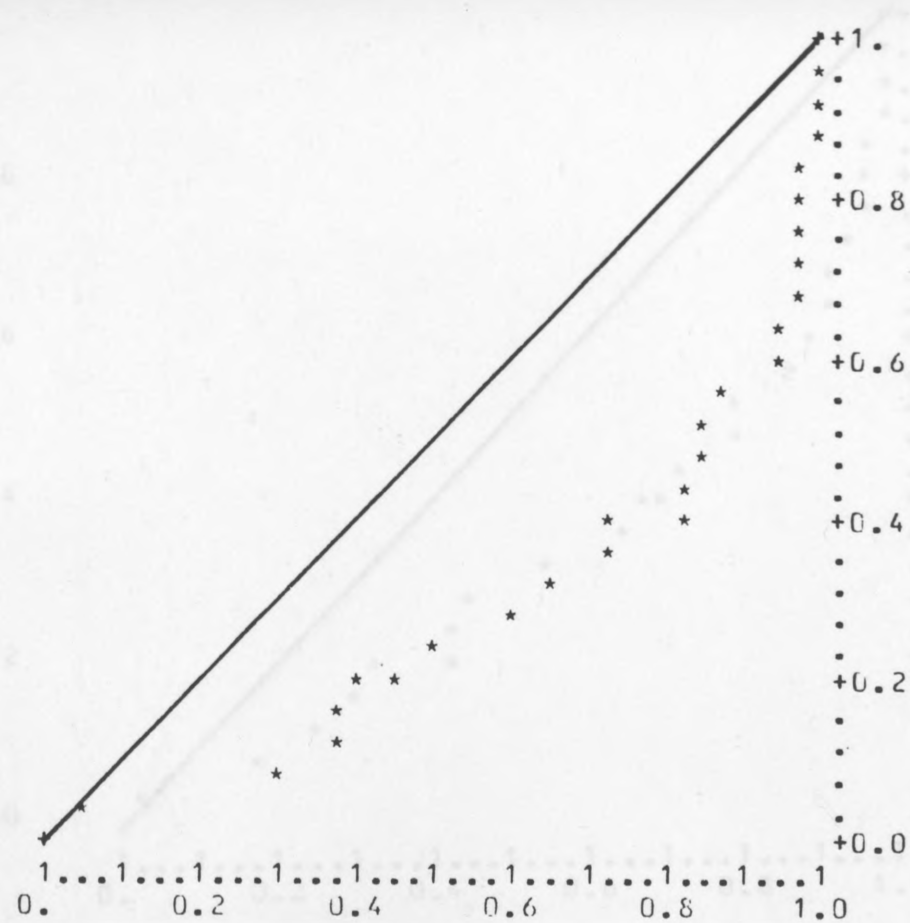


FIG. 5.88. DJMNHPP u-plots, data in Table 5,25, the plots based on the line printer output.

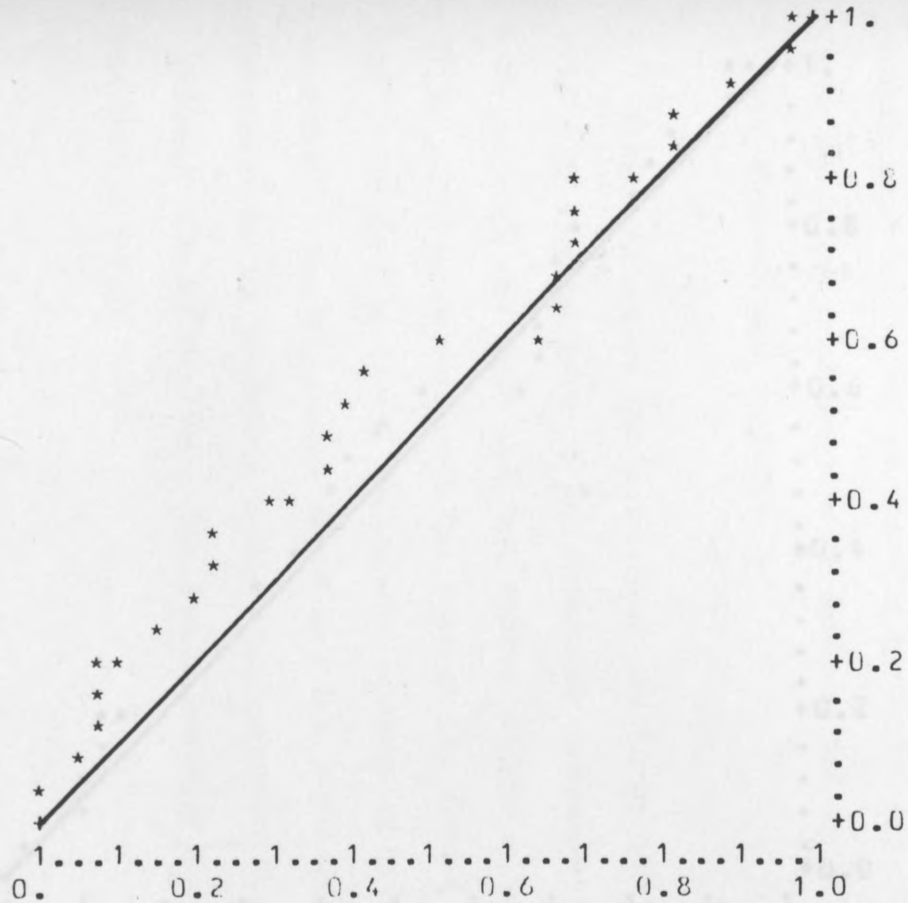


FIG.5.91. DD u-plots, data in Table 5,25, the plots based on the line printer output.

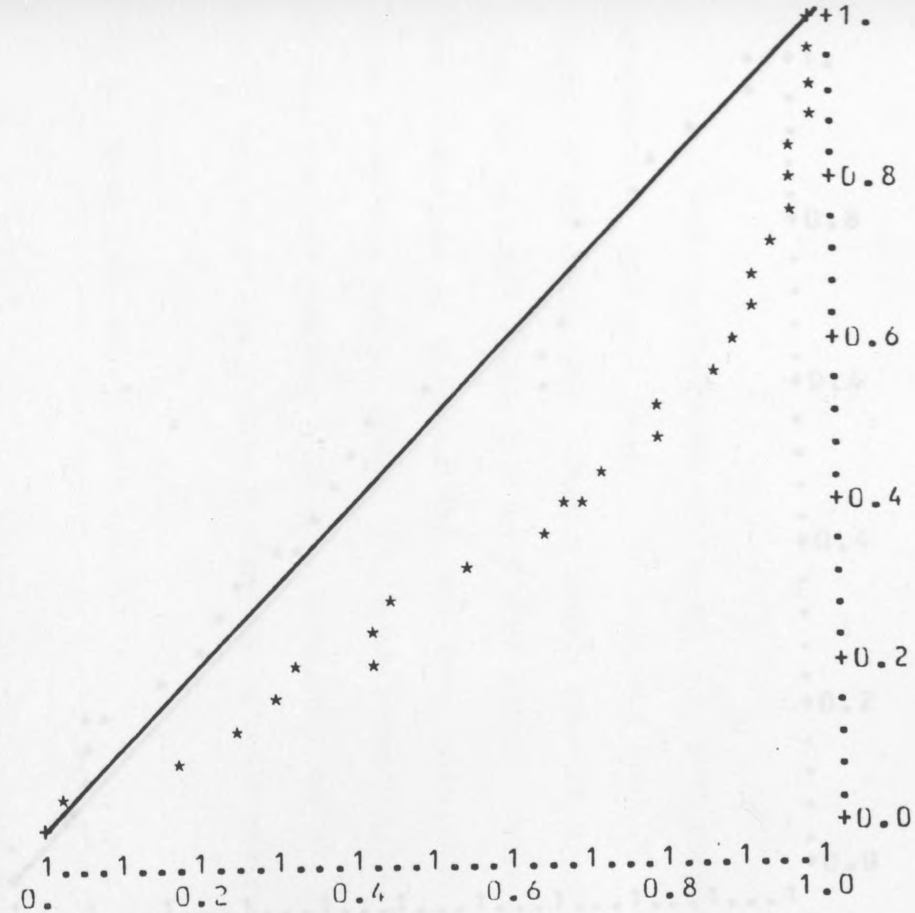


FIG.5.90. DW u-plots, data in Table 5.25, the plots based on the line printer output.

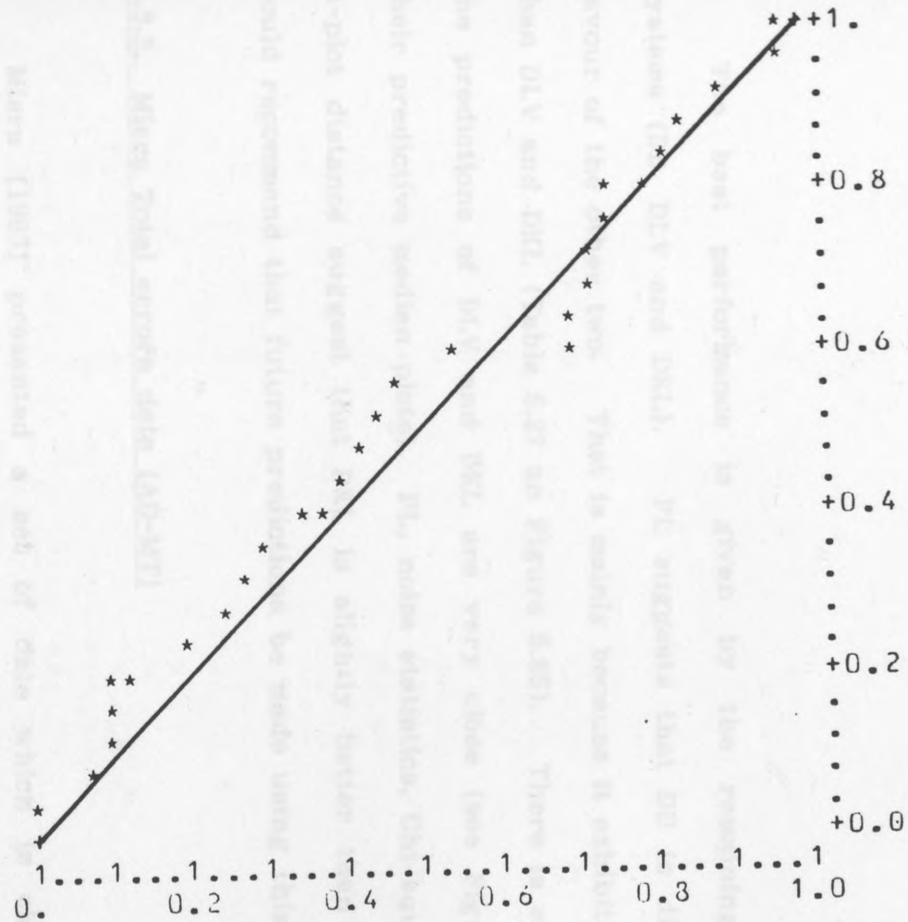
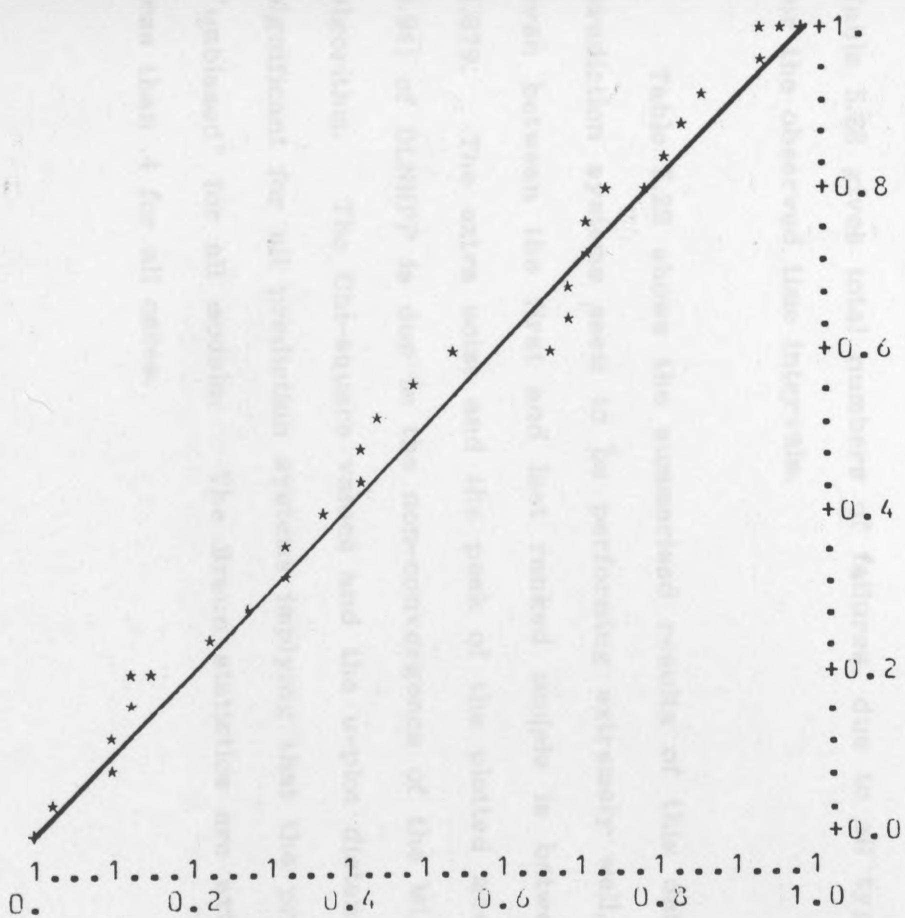


FIG. 5.92. DLV u-plots, data in Table 5.25, the plots based on the line printer output.



FIG,5.93. DKL u-plots, data in Table 5.25, the plots based on line printer output.

The best performance is given by the remaining prediction systems (DD, DLV and DKL). PL suggests that DD is discredited in favour of the other two. That is mainly because it exhibits more noise than DLV and DKL (Table 5.27 an Figure 5.85). There is evidence that the predictions of DLV and DKL are very close (see Figure 5.85 for their predictive median plots). PL, noise statistics, Chi-square and the u-plot distance suggest that DKL is slightly better than DLV, so we could recommend that future predictions be made using this system.

5.3.3. Misra Total errors data (AD-MT)

Misra (1983) presented a set of data which is classified into critical, major and minor errors of a particular system. The data of Table 5.28 gives total numbers of failures, due to all types of error, for the observed time intervals.

Table 5.29 shows the summarised results of this data set. All prediction systems seem to be performing extremely well. The PLR even between the first and last ranked models is between .638 and 1.879. The extra noise and the peak of the plotted medians (Figure 5.94) of DLNHPP is due to the non-convergence of the ML optimisation algorithm. The Chi-square values and the u-plot distances are non-significant for all prediction systems implying that the predictions are "unbiased" for all models. The Braun statistics are extremely good: less than .4 for all cases.

I	NR(I)	TN(I)	TP(I)
1	15	0.	62.50
2	6	15.00	44.00
3	8	21.00	40.00
4	8	29.00	68.00
5	8	37.00	62.00
6	4	45.00	66.00
7	4	49.00	73.00
8	8	53.00	73.50
9	6	61.00	92.00
10	2	67.00	71.40
11	7	69.00	64.50
12	8	76.00	64.70
13	3	84.00	36.00
14	5	87.00	54.00
15	5	92.00	39.50
16	8	97.00	68.00
17	8	105.00	61.00
18	6	113.00	62.60
19	12	119.00	98.70
20	5	131.00	25.00
21	2	136.00	12.00
22	5	138.00	55.00
23	6	143.00	49.00
24	9	149.00	64.00
25	1	158.00	26.00
26	4	159.00	66.00
27	2	163.00	49.00
28	4	165.00	52.00
29	4	169.00	70.00
30	9	173.00	84.50
31	6	182.00	83.00
32	1	188.00	60.00
33	3	189.00	72.50
34	6	192.00	90.00
35	6	198.00	58.00
36	3	204.00	60.00
37	14	207.00	168.00
38	10	221.00	111.50

TABLE 5.28. Misra total errors data, read left to right, test interval, numbers of failures inside the interval, observed numbers of failures at the beginning of the interval, and the interval length in time unit.

Test Discrete Data System: AD-MT

No.of Observations: = 38
Starting Sample Size: = 20

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	.8270	7	.7923	7	.3695	8	15.280 N.S.		.2060 N.S.	7	1.5707 38.3281	7
DL	.7578	5	.7272	5	.3495	6	14.885 N.S.		.2035 N.S.	5	1.5841 38.1320	4
DLV	.6712	3	.6716	3	.3065	3	14.590 N.S.		.2252 N.S.	8	1.6788 38.0529	3
DKL	.6936	4	.6823	4	.3066	4	14.629 N.S.		.2046 N.S.	6	1.6795 38.1847	5
DD	.3772	1	.3698	1	.2585	1	13.704 N.S.		.1406 N.S.	1	1.6253 37.7755	1
DJMNHPP	.8144	6	.7777	6	.3692	7	15.177 N.S.		.1986 N.S.	4	1.5460 38.4062	8
DLNHPP	2.5163	8	.8079	8	.3354	5	14.843 N.S.		.1962 N.S.	3	1.5629 38.2865	6
DW	.3772	1	.3698	1	.2585	1	13.704 N.S.		.1406 N.S.	1	1.6253 37.7755	1

TABLE 5.29. The summarised results of Misra's total errors data (Table 5.28).
The ML routine does not terminate normally for DL,DKL and DLNHPP.

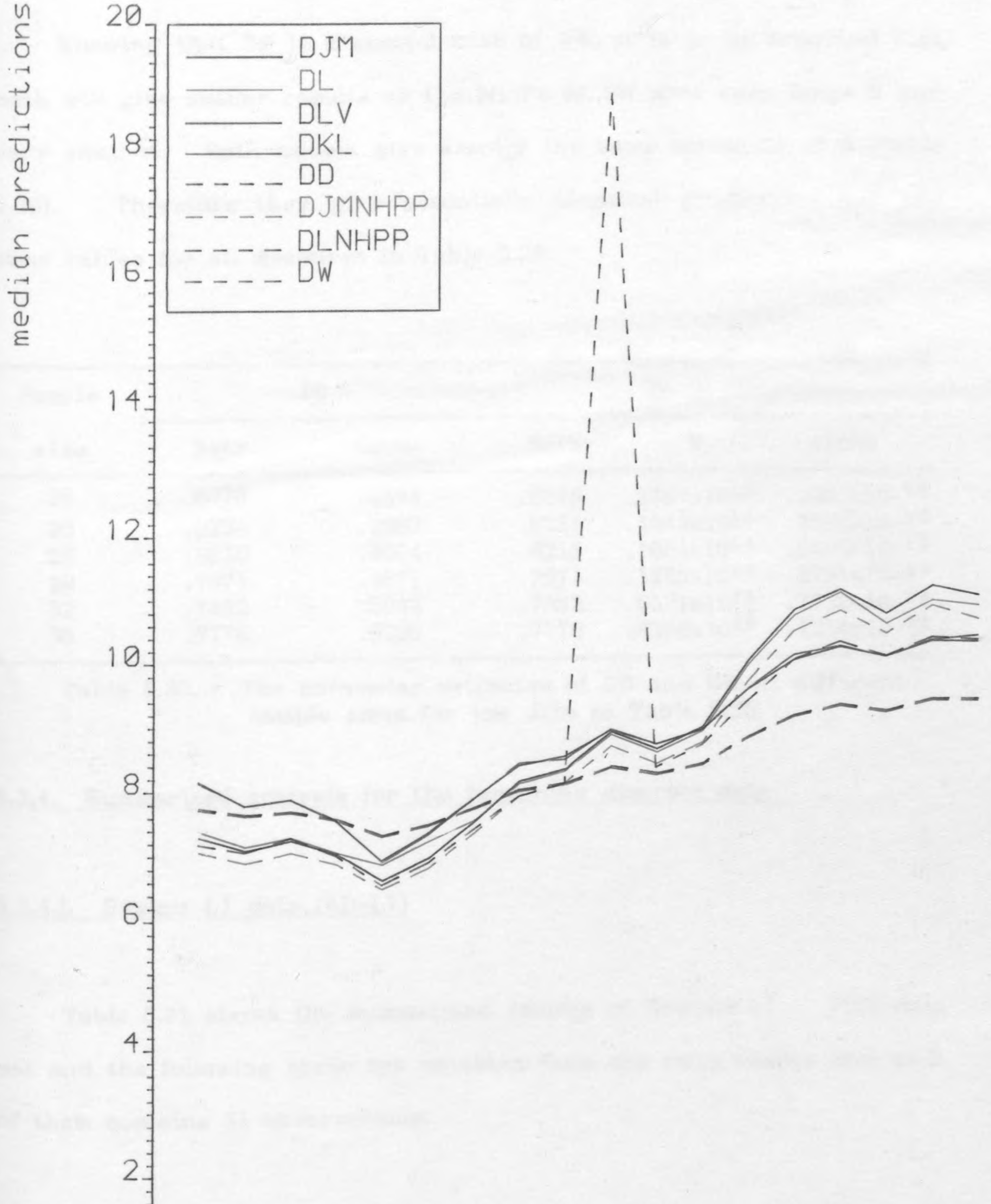


FIG.5.94. The plots of predictive median time to 1st failure in each prediction interval for the data in Table 4.28

Knowing that DD is a special case of DW, it is to be expected that both will give similar results as the MLE's of DW give very large $\hat{N}$ and very small $\hat{\alpha}$. Both models give exactly the same estimates of β (Table 5.30). Therefore they give essentially identical predictions and the same values for all measures in Table 5.29.

Sample	DD		DW		
size	Beta	Gamma	Beta	N	Alpha
20	.8070	.4374	.8070	$.2364 \times 10^{13}$	$.1851 \times 10^{-12}$
23	.8231	.3967	.8231	$.1549 \times 10^{14}$	$.2561 \times 10^{-13}$
26	.8218	.4004	.8218	$.1664 \times 10^{14}$	$.2406 \times 10^{-13}$
29	.7971	.4671	.7971	$.1245 \times 10^{14}$	$.3751 \times 10^{-13}$
32	.7852	.5042	.7852	$.6521 \times 10^{13}$	$.7732 \times 10^{-13}$
35	.7776	.5290	.7776	$.4358 \times 10^{13}$	$.1214 \times 10^{-12}$

Table 5.30. The parameter estimates of DD and DW at different sample sizes for the data in Table 5.28.

5.3.4. Summarized analysis for the remaining discrete data

5.3.4.1. System L1 data (AD-L1)

Table 5.31 shows the summarised results of System L1. This data set and the following three are obtained from the same source and each of them contains 31 observations.

According to PL ranks, the last three prediction systems (DW, DJM and DJMNHPP) will be discounted. These predictors also have the highest ranks in all other statistics, as well as giving the worst (and significant) Chi-square values.

Test Discrete Data System AD-LI

No.of Observations: = 31
Starting Sample Size: = 18

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	2.1076	7	1.8433	7	1.6378	7	36.409 .1%	35.407 .1%(11)	.4993 .1%	7	3.1221 39.2401	7
DL	1.5275	5	1.3306	5	1.3291	5	28.268 1%	28.032 1%(12)	.4042 2%	5	3.4694 36.6533	5
DLV	.8975	3	.8090	3	1.0161	2	21.620 10%		.3076 20%	3	2.9624 34.4375	2
DKL	.8838	2	.7608	2	.9631	1	20.072 10		.2440 N.S.	1	2.9184 34.4360	1
DD	.8464	1	.7592	1	1.1828	3	20.354 10%		.2989 20%	2	4.3119 36.0872	3
DJMNHPP	1.8724	6	1.6189	6	1.4767	6	31.1748 1%	31.025 .5%(12)	.4555 1%	6	3.5696 37.9247	6
DLNHPP	1.3665	4	1.2112	4	1.2460	4	26.219 5%		.3869 5%	4	3.5696 36.4031	4
DW	3.3512	8	3.2228	8	2.6982	8	100.862 .1%	68.730 .1%(7)	.6767 1%	8	1.4120 54.6116	8

TABLE 5.31. The summarized results of system L1 data. The ML routine does not always terminate normally for DL.

The remaining predictors (DL, DLNHPP, DD, DLV and DKL) which are the best in PL ranks, are close to each other in median plots (Figure 5.95). However, DL and DLNHPP exhibit more noise than the other three, and both have singificant Chi-square values and u-plots. Among the other three DKL is best in PL, u-plot, Chi-square and it is the only one with a value for the Braun statistic less than one. This is the system which would be preferred for future predictions.

5.3.4.2. System L2 data (AD-L2)

PL ranks suggest that DD, DJMNHPP and DLNHPP are best performing prediction systems for this data set (Table 5.32). Although DL, DLV and DKL have non-significant Chi-square, their PL performances have been affected by their bias.

DJM and DW have significant Chi-square at 5% and 10% level respectively, but the Chi-square value of DJM is improved by grouping (Table 5.32). Both models reflect "bias" in their predictions, which contributes to their poor PL performance.

It is difficult finally to choose between the three with best PL performance. DD is slightly better in PL and u-plot but it is noisier than DJMNHPP. More importantly, it has a very poor Braun statistic. This may be due to non-stationarity (as in section 5.3.1), but there is not sufficient data to judge with confidence. It might be safer to reject DD because of this problem and use either DJMNHPP or DLNHPP.

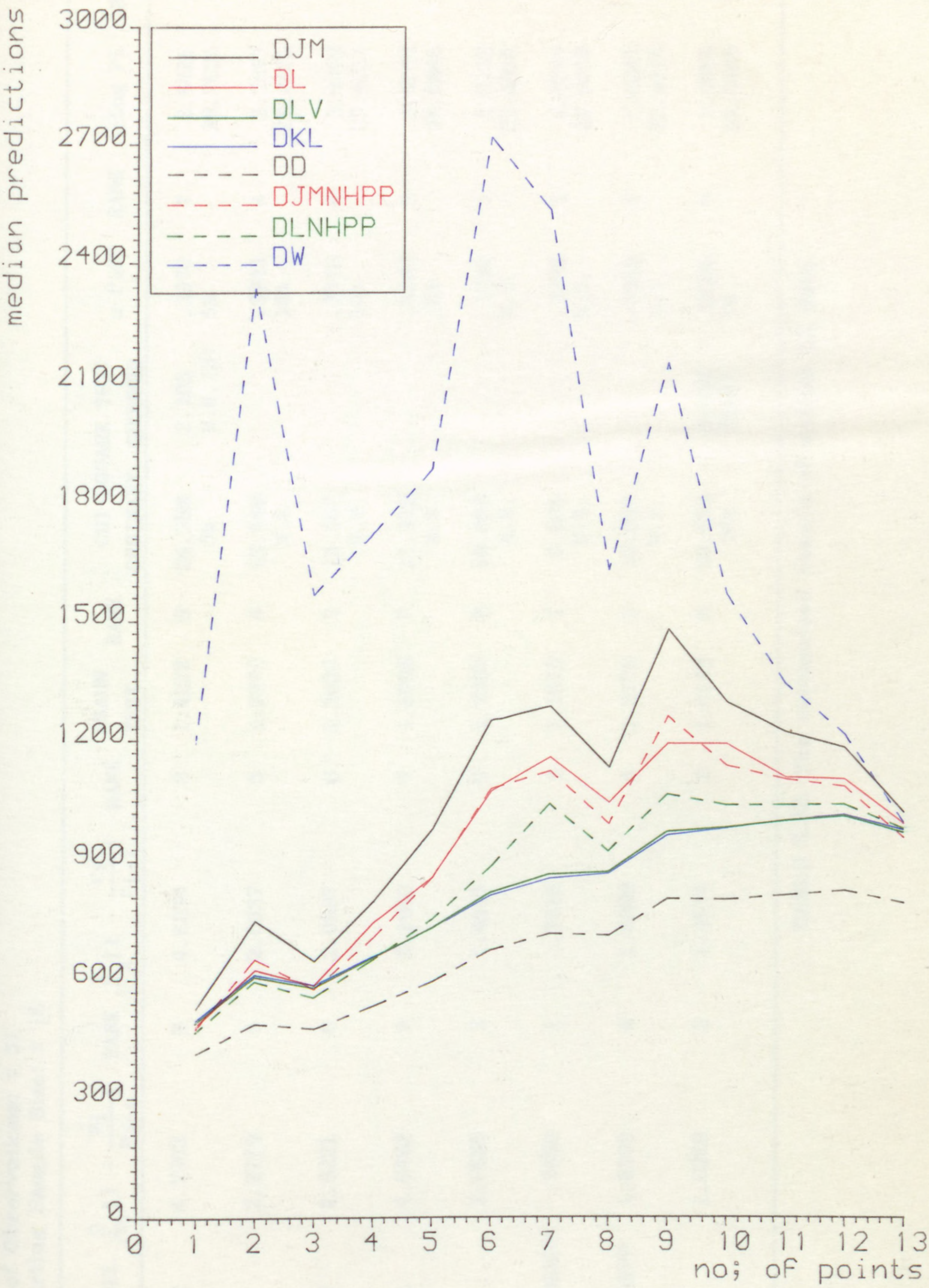


FIG.5.95. The plots of predictive median time to 1st failure in each prediction interval for system L1 data(AD-L1)

Test Discrete Data System: AD-L2

No.of Observations: = 31
Starting Sample Size: = 18

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	4.7983	8	4.6234	8	1.4222	6	24.359 5%	2.103 N.S.(3)	.3970 5%	7	3.6426 28.7623	8
DL	2.2719	5	2.0337	5	1.2470	4	12.648 N.S.		.3243 10%	4	2.7877 23.4876	4
DLV	2.6221	6	2.0997	6	1.2024	2	13.067 N.S.		.3548 10%	6	2.9259 25.4031	5
DKL	4.6462	7	3.2940	7	1.6796	7	11.912 N.S.		.3287 10%	5	2.6778 25.6845	7
DD	1.6635	3	1.4915	3	1.7420	8	10.053 N.S.		.1352 N.S.	1	2.0152 22.0815	1
DJMNHPP	.7650	1	.7616	1	1.1815	1	9.836 N.S.		.1992 N.S.	2	2.6243 22.6090	2
DLNHPP	1.8302	4	1.6889	4	1.2176	3	10.892 N.S.		.2365 N.S.	3	2.6247 22.8810	3
DW	1.6369	2	1.4514	2	1.4143	5	20.050 10%	8.666 5%(3)	.5636 1%	8	3.0884 25.5469	6

TABLE 5.32. The summarised results of system L2 data.

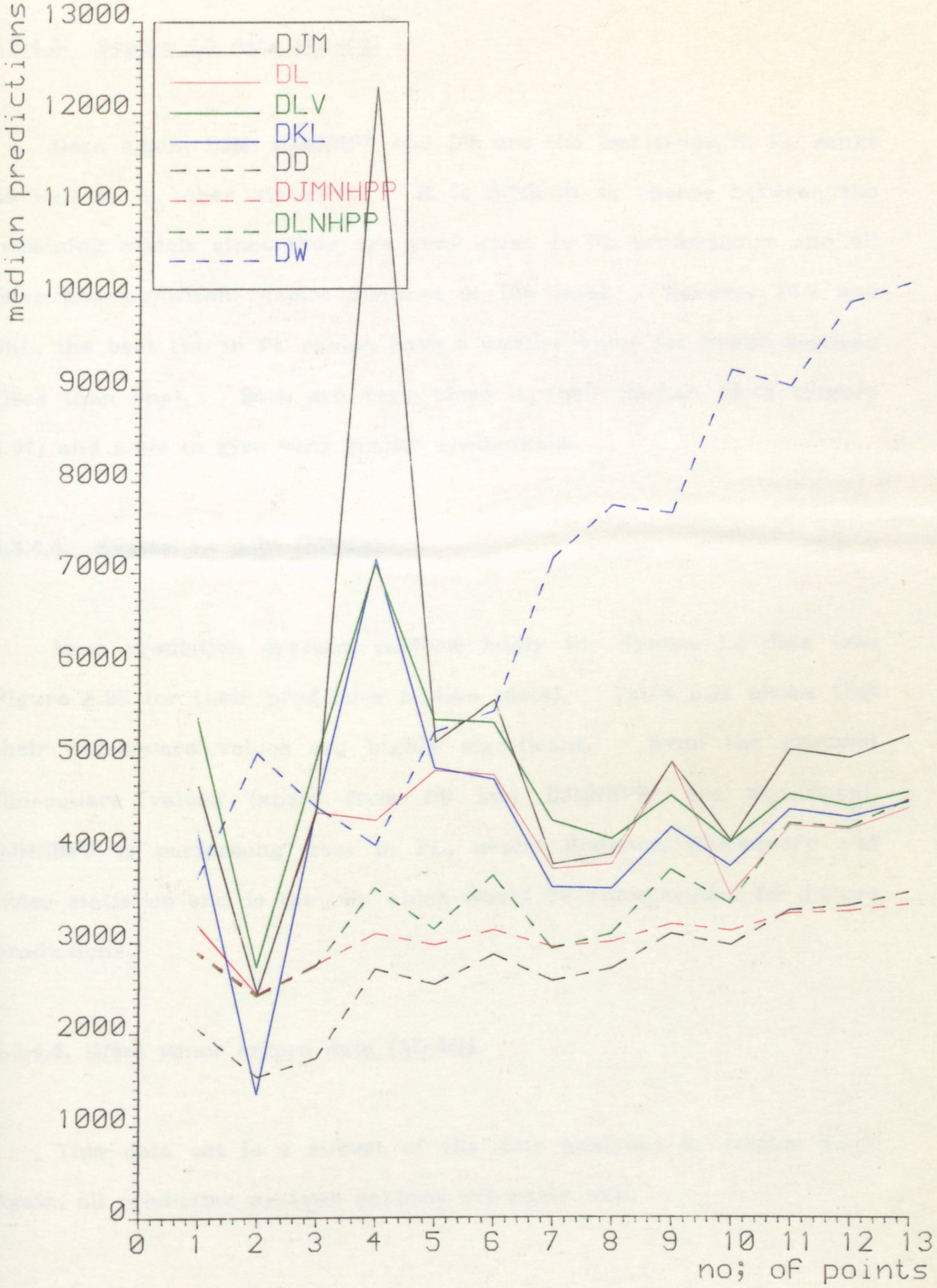


FIG.5.96. The plots of predictive median time to 1st failure in each prediction interval for system L2 data(AD-L2)

5.3.4.3. System L3 data (AD-L3)

Once again, DJM, DJMNHPP and DW are the last three in PL ranks as well as in other statistics. It is difficult to choose between the remaining models since they are very close in PL performance and all have non-significant u-plot distance at 10% level. However DLV and DKL, the best two in PL ranks, have a smaller value for Braun statistic (less than one). Both are very close in their median plots (Figure 5.97) and seem to give very similar predictions.

5.3.4.4. System L4 data (AD-L4)

Most prediction systems perform badly for System L4 data (see Figure 5.98 for their predictive median plots). Table 5.34 shows that their Chi-square values are highly significant. Even the grouped Chi-square values (apart from DD and DJMNHPP) are significant. DJMNHPP is performing best in PL, u-plot distance, Chi-square and noise statistics and is the one which would be recommended for future predictions.

5.3.4.5. Misra minor errors data (AD-MI)

This data set is a subset of the data analysed in section 5.3.3. Again, all prediction systems perform extremely well.

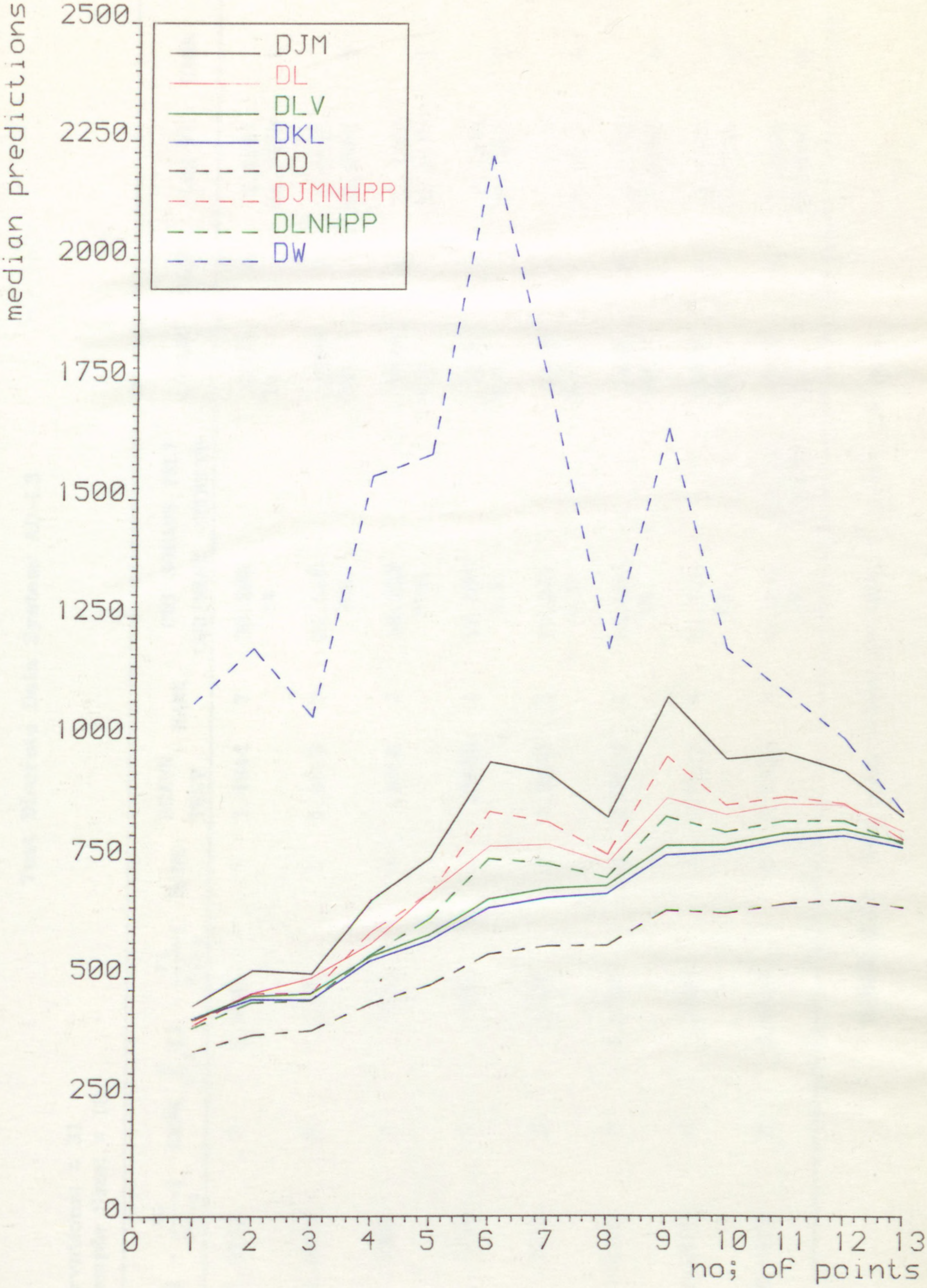


FIG.5.97. The plots of predictive median time to 1st failure in each prediction interval for system L3 data(AD-L3)

Test Discrete Data System: AD-L3

No.of Observations: = 31
Starting Sample Size: = 18

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	1.5742	7	1.3801	7	1.4844	7	30.206 1%		.4065 2%	7	2.1044 38.3579	7
DL	1.0483	5	.9379	5	1.0823	3	21.596 10%		.2950 20%	4	2.2775 35.4055	3
DLV	.7340	3	.7030	3	.9566	2	18.943 N.S.		.2309 N.S.	2	2.3252 35.0177	1
DKL	.7264	1	.6447	1	.9448	1	17.669 N.S.		.2078 N.S.	1	2.3348 35.2692	2
DD	.7310	2	.6632	2	1.2260	5	19.341 N.S.		.2448 N.S.	3	2.8857 36.8023	5
DJMNHPP	1.4298	6	1.2390	6	1.3579	6	26.669 5%		.3605 10%	6	2.3378 37.4262	6
DLNHPP	1.0456	4	.9180	4	1.1112	4	21.778 5%		.3043 20%	5	2.3577 35.8363	4
DW	2.6422	8	2.6413	8	2.8126	8	87.695 .1%	59.465 .1%(8)	.7111 1%	8	2.2664 54.9906	8

TABLE 5.33. The summarised results of system L3 data.

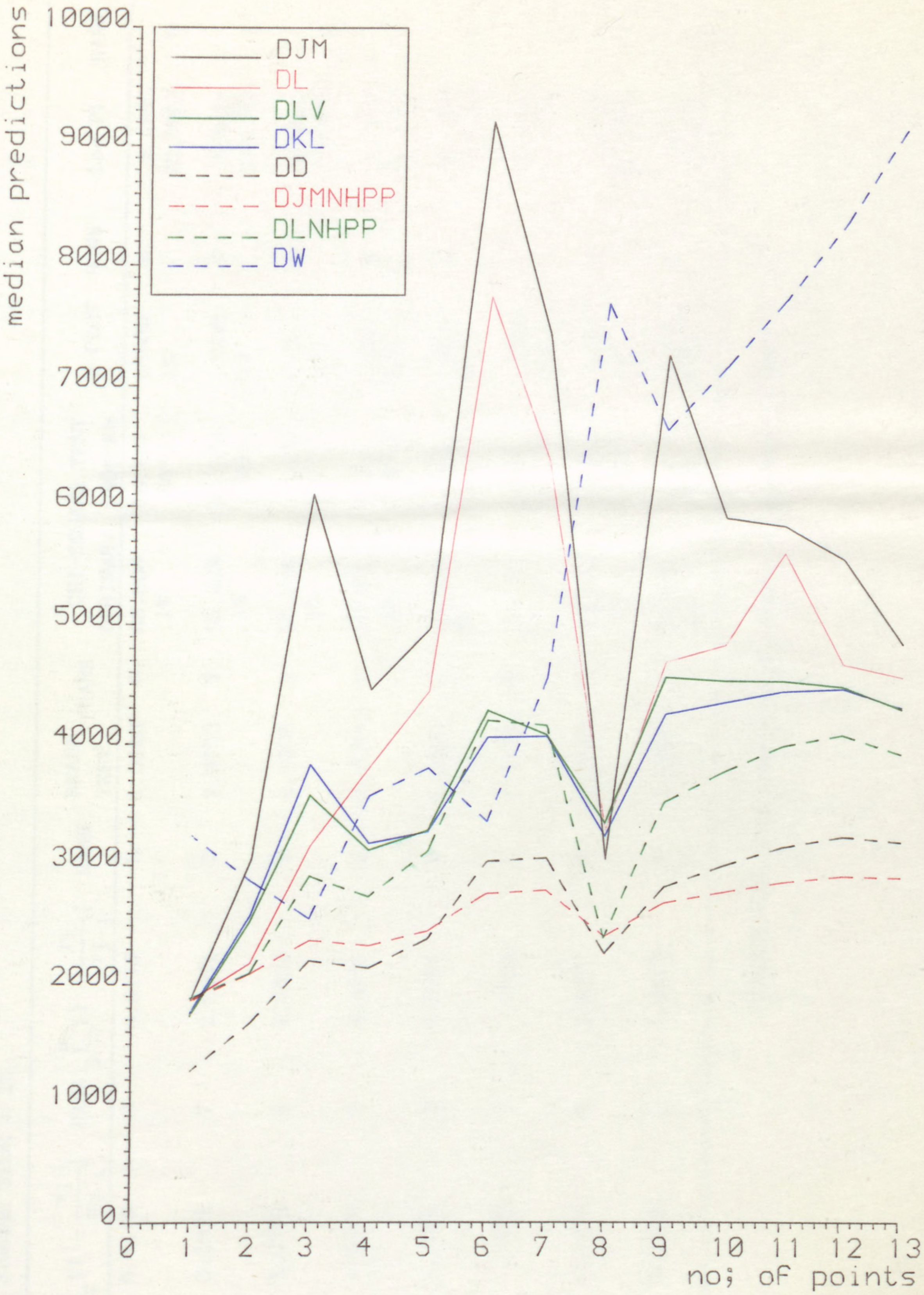


FIG.5.98. The plots of predictive median time to 1st failure in each prediction interval for system L4 data(AD-L4)

Test Discrete Data System: AD-L4

No.of Observations: = 31
Starting Sample Size: = 18

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	5.4110	8	4.5178	8	1.5103	8	65.806 .1%	16.187 .5%(3)	.4630 1%	8	1.6946 39.2963	8
DL	3.1921	7	2.9931	7	1.3759	7	53.829 .1%	25.724 .1%(4)	.3792 5%	7	1.6946 35.3144	7
DLV	1.9958	4	1.4009	3	1.1598	3	37.286 .1%	9.371 10%(4)	.2865 20%	4	1.8069 29.6641	4
DKL	1.9926	3	1.3065	2	1.1372	2	36.262 .1%	9.084 10%(4)	.2983 20%	5	1.7693 28.9673	2
DD	1.7248	2	1.5297	4	1.3590	6	30.956 1%	7.741 N.S.(5)	.2031 N.S.	2	2.5378 29.5933	3
DJMNHPP	.8079	1	.7704	1	1.1703	4	26.443 5%	5.758 N.S.(5)	.1399 N.S.	1	1.6954 26.5933	1
DLNHPP	2.1374	6	2.0177	6	1.3172	5	39.189 .1%	10.410 5%(4)	.2465 N.S.	3	1.6713 31.6486	5
DW	2.3034	6	1.9479	5	1.1222	1	76.042 .1%	11.931 1%(3)	.3365 10%	6	1.0022 33.1682	6

TABLE 5.34. The summarised results of system L4 data.

Apart from DJM and DJMNHPP, all predictors have non-significant u-plot distances at 20% level (Table 5.35). All models are close in their predictive median plots (Figure 5.99). Again, the Braun statistic values are remarkably small (less than .7) for all predictors. Clearly, it is hard to choose a particular prediction system for future use. However DD is giving a slightly better performance in all measures.

5.3.4.6. System 8 data (AD-8)

The prediction systems which have the worst PL ranks (DLV, DKL, DD and DW) also have significant Chi-square and u-plot distances (Table 5.36). The u-plots of DLV, DKL and DD imply that these models are too pessimistic, (they produce too many small u's). Thus they are not able to capture the mean of the data accurately, so they have significant Chi-square values. Conversely, DW is too optimistic (too many large u's). Again the model is not accurate in predicting the mean. It is this bias in the predictions which probably contributes most to their poor PL performance.

The remaining four models (DJM, DL, DJMNHPP and DLNHPP) are the best performing models in PL. Their predictions seem very close (see, for example, median plots in Figure 5.100). Table 5.37 shows why DL and DJM are close: they have similar estimates of N and $\hat{\alpha}$, $\hat{\beta}$ are large with $\hat{\alpha}/\hat{\beta}$ approximately equal to $\hat{\Phi}$.

Test Discrete Data System: AD-MI

No.of Observations: = 38
Starting Sample Size: = 20

MODEL	$\sum_{i=2}^{n} \left\| 1 - \frac{m_i}{m_{i-1}} \right\|$	RANK	$\sum_{i=2}^{n} \left\| 1 - \frac{r_i}{r_{i-1}} \right\|$	RANK	BRAUN TEST	RANK	CHI SQUARE TEST ORDINARY	CHI SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	1.1177	7	1.0703	8	.6717	8	23.534 N.S.		.2679 20%	8	1.0619 35.9744	8
DL	.9849	5	.9402	5	.6178	6	21.672 N.S.		.2466 N.S.	6	1.0688 35.3669	6
DLV	.5979	2	.5830	2	.5350	3	19.626 N.S.		.2085 N.S.	3	1.1051 34.5518	3
DKL	.6294	3	.6159	3	.5364	4	19.713 N.S.		.2109 N.S.	4	1.1139 34.5414	2
DD	.4729	1	.4604	1	.4233	1	16.497 N.S.		.1715 N.S.	1	1.0599 34.0949	1
DJMNHPP	1.0573	6	1.0069	7	.6464	7	22.352 N.S.		.2492 20%	7	1.0475 35.7834	7
DLNHPP	1.7180	8	.9485	6	.5760	5	20.215 N.S.		.2196 N.S.	5	1.0582 35.0473	5
DW	.7697	4	.7535	4	.4983	2	17.943 N.S.		.2007 N.S.	2	1.0599 34.5842	4

TABLE 5.35. The summarised results of Misra's minor errors data.
The ML routine does not always terminate normally for DL, DLV, DKL and DLNHPP.

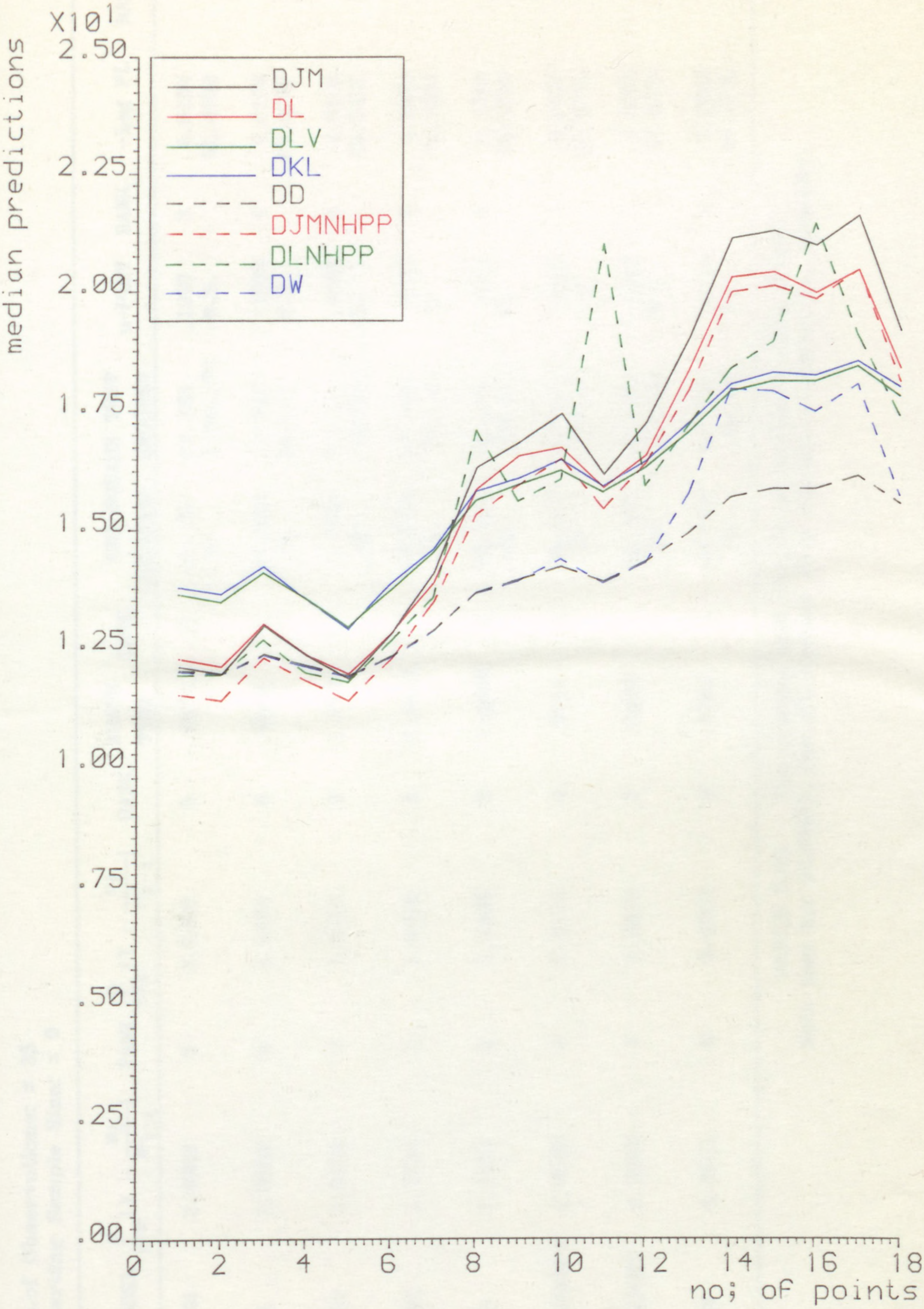


FIG.5.99. The plots of predictive median time to 1st failure in each prediction interval for Misra minor errors data(AD-MI)

Test Discrete Data System: AD-8

No.of Observations: = 35
Starting Sample Size: = 9

MODEL	$\sum_{i=2}^{n} \|1 - \frac{m_i}{m_{i-1}}\|$	RANK	$\sum_{i=2}^{n} \|1 - \frac{r_i}{r_{i-1}}\|$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	2.8868	5	2.5406	5	.9374	2	66.121	33.261	.1997	3	4.5104	
							.1%	2.5%(17)	N.S.		62.4980	1
DL	2.8984	6	2.5444	6	.9365	1	66.729	28.520	.2021	4	4.5204	
							.1%	5%(17)	N.S.		62.5373	2
DLV	1.3475	3	1.1757	2	1.3339	6	55.649	55.346	.4299	6	4.8195	
							.1%	.1%(23)	1%		65.7492	5
DKL	1.1735	1	1.0918	1	1.4111	7	57.610	57.346	.5145	7	4.9338	
							.1%	.1%(24)	1%		71.7481	7
DD	1.3431	2	1.2556	3	1.8484	8	66.424	66.232	.5672	8	4.7703	
							.1%	.1%(24)	1%		78.5236	8
DJMNHPP	2.8548	4	2.5179	4	.9714	3	64.225	55.917	.1680	2	4.7703	
							.1%	.1%(18)	N.S.		62.6649	3
DLNHPP	2.9850	7	2.6476	7	.9924	5	63.939	55.169	.1210	1	4.8062	
							.1%	.1%(19)	N.S.		62.8533	4
DW	4.0611	8	3.4954	8	.9753	4	104.091	55.181	.2769	5	2.7846	
							.1%	.1%(12)	5%		69.1776	6

TABLE 5.36. The summarized results of System 8 data (AD-8)
Note that for DLNHPP, the ML routine does not always terminate normally.

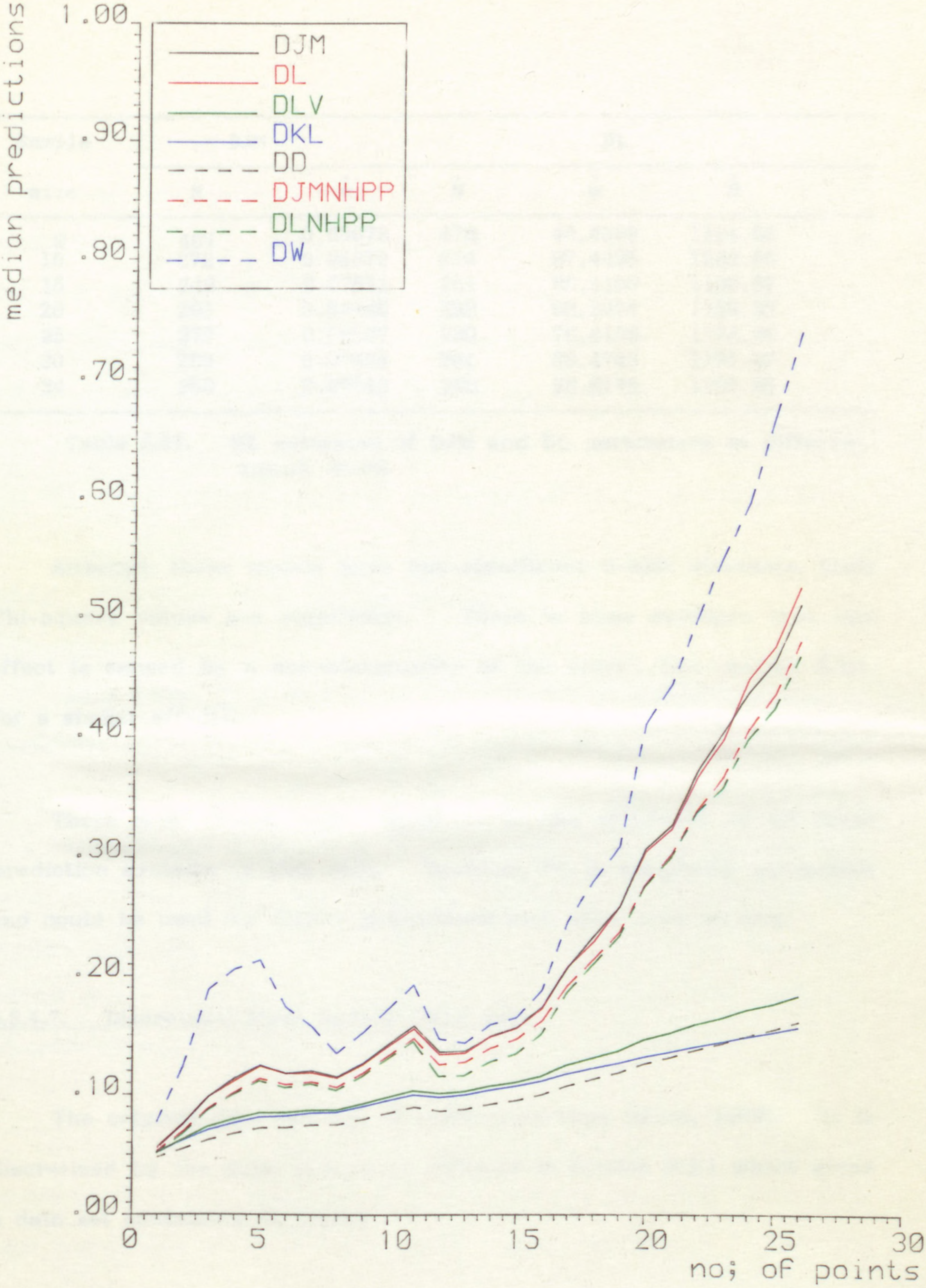


FIG.5.100. The plots of predictive median time to 1st failure in each prediction interval for system 8 data(AD-8)

Sample	DJM		DL		
size	$\hat{N}$	$\hat{\Phi}$	$\hat{N}$	$\hat{\alpha}$	$\hat{\beta}$
9	467	0.03679	478	40.4369	1124.62
10	272	0.06879	274	87.4495	1283.56
15	249	0.07573	251	85.1100	1130.57
20	295	0.05925	298	68.2474	1159.33
25	279	0.06527	280	76.4139	1174.86
30	260	0.07434	261	88.4743	1191.47
34	259	0.07513	259	88.8179	1191.85

Table 5.37. ML estimates of DJM and DL parameters at different sample sizes.

Although these models have non-significant u-plot distances, their Chi-square values are significant. There is some evidence that this effect is caused by a non-stationarity of the errors (see section 5.3.1. for a similar effect).

There must remain some doubt as to the efficiency of all these prediction systems on this data. However, DL is marginally acceptable and could be used for future predictions with some reservations.

5.3.4.7. Discretized Musa System SSIC data

The original data set was of continuous type [Musa, 1979]. It is discretized by the same procedure followed in section 5.3.1 which gives a data set containing 41 points.

Table 5.38 shows the summarized results of this data set. It is clear that Chi-square values are highly significant for all the prediction systems. All predictions seem to produce too small u's in the first half and too large ones in the later half so that overall u's reflect unbiased prediction for all (DJM and DD apart). This behaviour affects both Chi-square and Braun statistics.

No prediction system can be recommended for future predictions.

5.3.4.8. Discretized Musa System SS3 data

Apart from DLV and DKL, all prediction systems seem too pessimistic as judged by inspection of their u-plots.

Although DLV and DKL have non-significant u-plot distances they have significant Chi-square values. There is evidence that they predict small u's in the first half of the data and large u's in the second half. This effect is averaged in the u-plot, but neither is accurate in capturing the mean of the data.

A conservative approach here might be to use DL or DLNHPP for future predictions. The poor PL performance is probably due to their bias: but at least this bias errs on the side of pessimism.

Test Discrete Data System: AD-SS1C

No.of Observations: = 41
Starting Sample Size: = 21

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	.7863	3	.7546	3	1.1865	4	50.275 .1%		.2635 20%	7	1.8677 63.2704	3
DL	1.1650	5	1.0948	5	1.1656	2	66.099 .1%	62.235 .1%(19)	.1914 N.S.	4	1.8874 63.4469	4
DLV	.7514	2	.7417	2	1.1501	1	53.468 .1%		.1645 N.S.	1	2.0923 55.6383	1
DKL	.6302	1	.5096	1	1.1938	5	52.298 .1%		.1850 N.S.	3	2.1290 55.7785	2
DD	.8195	4	.7858	4	1.2521	6	53.623 .1%		.2669 10%	8	1.9327 64.8102	6
DJMNHPP	1.9760	7	2.0800	7	1.3898	8	117.189 .1%	95.652 .1%(17)	.2002 N.S.	5	2.0116 73.9538	8
DLNHPP	1.1929	6	1.1218	6	1.1786	3	64.861 .1%	61.069 .1%(17)	.2043 N.S.	6	1.9736 64.3615	5
DW	2.3800	8	2.3189	8	1.3351	7	94.661 .1%	71.195 .1%(17)	.1646 N.S.	2	1.9327 70.2248	7

TABLE 5.38. The summarised results of discretized Musa's data system SS1C.

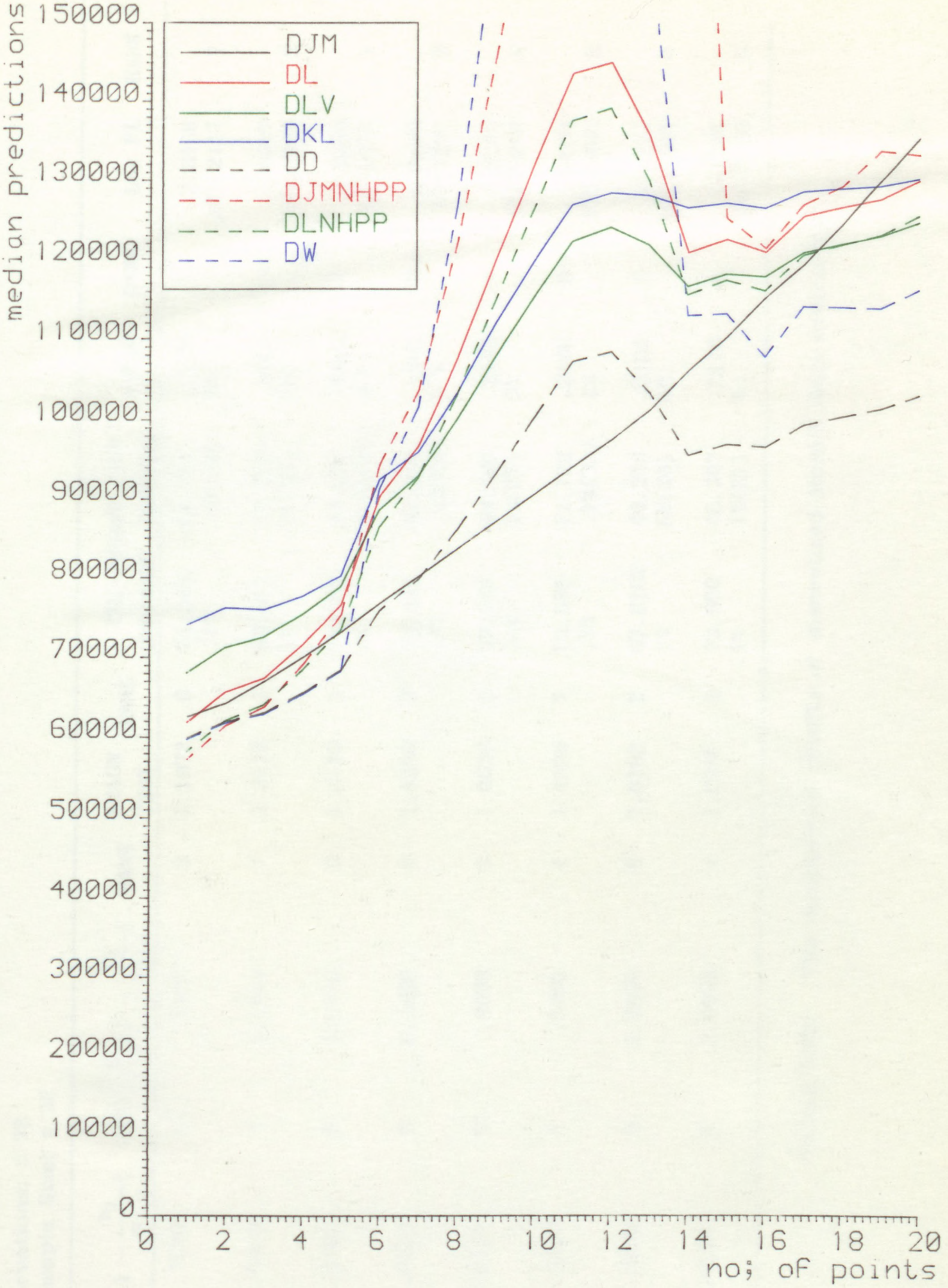


FIG.5.101. The plots of predictive median time to 1st failure in each prediction interval for discretized Musa system SS1C data

Test Discrete Data System: AD-SS3

No.of Observations: = 78
Starting Sample Size: = 38

MODEL	$\sum_{i=2}^{n} \lvert 1 - \frac{m_i}{m_{i-1}} \rvert$	RANK	$\sum_{i=2}^{n} \lvert 1 - \frac{r_i}{r_{i-1}} \rvert$	RANK	BRAUN TEST	RANK	CHI-SQUARE TEST ORDINARY	CHI-SQUARE TEST GROUPED	u-PLOT	RANK	-Log PL	RANK
DJM	.7590	2	.7497	2	1.1972	6	74.348 .1%	54.862 .5%(30)	.2224 5%	7	2.8468 107.2117	7
DL	1.6980	7	1.6416	7	1.0113	1	69.657 1%	40.777 5%(26)	.2102 10%	3	2.8417 100.9830	3
DLV	1.6846	6	1.6170	6	1.0546	3	70.882 1%	43.083 2.5%(26)	.0886 N.S.	1	2.9334 94.6565	1
DKL	2.6320	8	2.1928	8	1.5259	8	78.239 .1%	57.008 .1%(28)	.1567 N.S	2	2.5129 97.3297	2
DD	1.0131	3	.9998	3	1.0636	5	73.787 .1%	48.666 1%(27)	.2110 5%	4	3.2284 103.5810	6
DJMNHPP	.5515	1	.5456	1	1.4390	7	79.585 .1%	71.132 .1%(35)	.2586 1%	8	2.6654 113.6486	8
DLNHPP	1.3399	5	1.2898	5	1.0183	2	67.8182 1%	40.256 5%(26)	.2191 5%	6	2.8761 101.0194	4
DW	1.0200	4	1.0063	4	1.0566	4	73.600 .1%	48.343 1%(27)	.2110 5%	4	3.2284 101.7256	5

TABLE 5.39. The summarised results of discretized Musa's data system SS3.

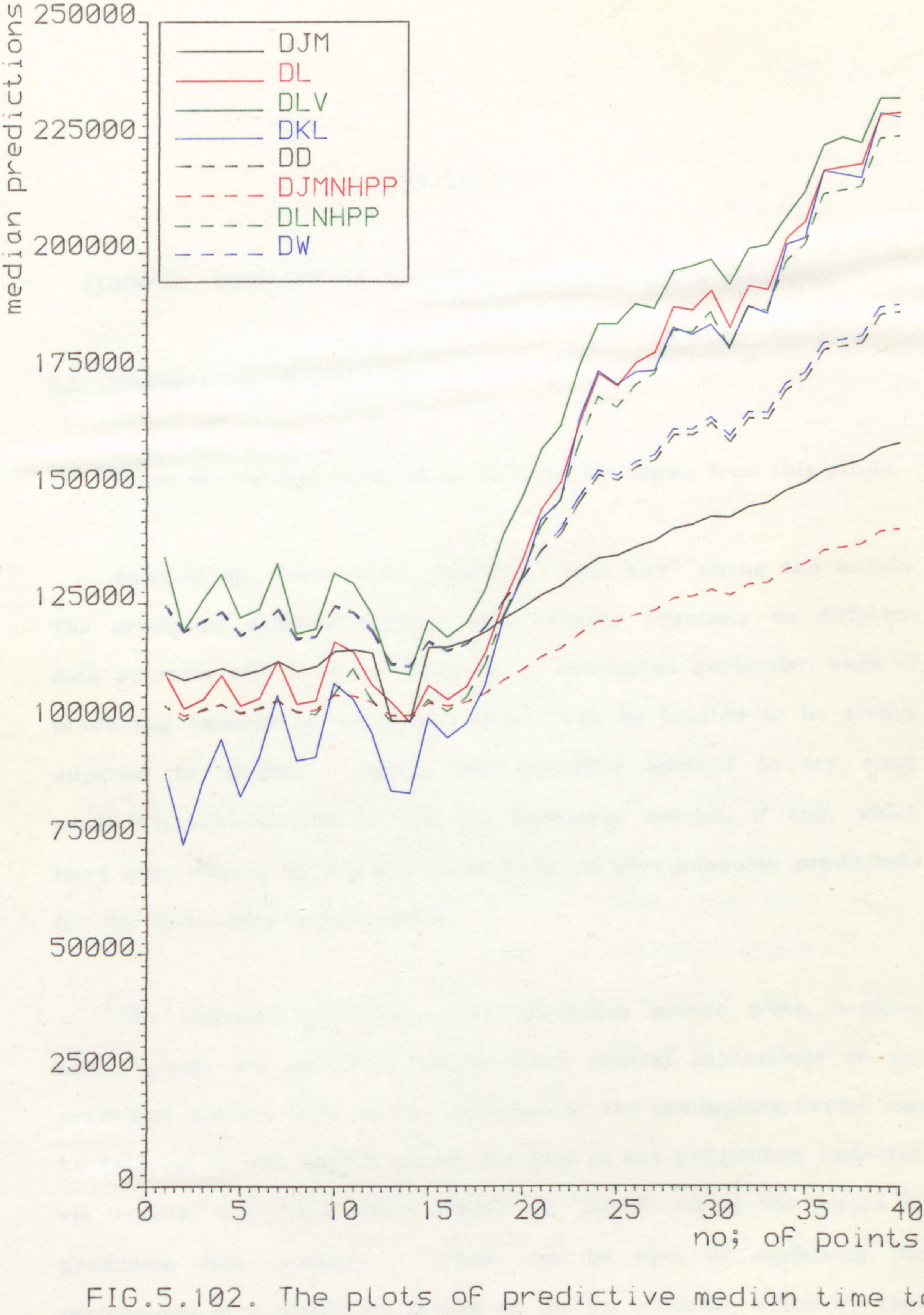


FIG.5.102. The plots of predictive median time to 1st failure in each prediction interval for discretized Musa system SS3 data

CHAPTER VI

SUMMARY, CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

6.1. Summary and Conclusions

There are several conclusions that can be drawn from this study.

First of all, there is no universal "best buy" among the models. The prediction systems perform with varying adequacy on different data sources, and it is not possible to recommend particular ways of predicting because no prediction system can be trusted to be always superior to others. Users are therefore advised to try many prediction systems and to use the reliability metrics, if any, which have been shown, by our analytical tools, to give adequate predictions for the data under consideration.

The graphical techniques (the predictive median plots, u-plots, and y-plots) are useful in giving some general impressions of the prediction quality, such as the noisiness of the predictions (which can be detected by the median plots), the bias in the predictions (detected via u-plots) and the general pattern or "trend" among the errors of prediction data (y-plots). These can be seen as explaining the performance of a prediction system on the PL criterion, which we view as a general tool for comparison.

The prediction systems considered in this study can be divided, by their performance, into two different groups:

(i) the optimistic models consisting of JM, W, JMNHPP, L and LNHPP (in discrete and continuous time) as well as BJM and BL (continuous time). Although L, LNHPP, BL and W (in large sets only) are considered in the group, their performances have been very close to reality in most data sets.

(ii) the pessimistic group includes LV, KL and D (in both cases, discrete and continuous). The predictions of these models have been very close to each other in most cases.

There are strong doubts about the assumption that all faults contribute the same amount to the overall failure rate of the program. This is clear from the performance of JM model family (JM, BJM, JMNHPP, DJM and DJMNHPP) where these models have been discredited by all tools in favour of other models for most data sets. This seems to be evidence that the deficiency arises from the modelling assumptions rather than the inference or prediction procedure. It seems much more plausible that the program starts life containing faults of different sizes, and this observation is supported by recent empirical studies [Nagel and Skrivan, 1981].

The Weibull model (discrete and continuous) seems to perform badly on the small data sets, while its performance in the large data sets has been one of good predictions. On the other hand, the L model tends to perform best of the order statistic models. This seems to suggest that the more extreme Pareto density best describes the distribution of times to detect faults.

The generalisation of the continuous models to cope with the discrete and summarized data will be useful, given the difficulty in obtaining full inter-event time data. The performance of these new discrete versions have been very close to the original ones in the cases when we have discretized a data set which has been analysed in its original continuous form.

Finally, it must be admitted that the predictive analysis tools which are reported here, are not the complete answer to the problem.

In the first place, PL is a relative tool, rather than an absolute one. That is, it allows us to pick the best predictive system, but does not (of itself) allow us to decide whether the predictions are objectively accurate. On the other hand, the u-plots and y-plots are not relative tools. So a poor u-plot tells us that predictions are, in some sense, objectively biased.

Secondly, the y-plot is a fairly crude measure of "stationarity of prediction error". In order to justify an adaptive procedure (see next section) we need to be sure that predictions are "only biased". Other tests of stationarity in the $\{u_i\}$ sequence should be investigated.

We have not succeeded in developing measures which tell, on the one hand, when unwarranted noise is present and, on the other, when predictions are too smooth compared with reality. Judicious use of our noise measures together with PL sometimes allow us to overcome this problem, however, and it seems reasonable to assume that nature here is usually smooth.

Finally, it is not obvious whether there are <u>other</u> deviations between prediction and reality which are not captured by the tools reported here. By analogy with conventional statistics, it seems reasonable that we have captured at least the first order effects.

6.2. Suggestions for future work

There are certain predictions for which the techniques reported here are not useful, such as the prediction of the time needed to achieve a specified target reliability (e.g. the reliability which must be achieved before the product can be shipped). Problems of this kind require further study, since it cannot be assumed that good predictions of one type (e.g. current reliability) necessarily imply good prediction in a different context.

All the analyses reported here have been carried out for real data sets. It might be useful to conduct a similar study using simulated data. The efficacy of the tools in selecting the best predictions could be evaluated when the true relationship between prediction and reality is known. The difficulty would be in selecting "suitable" data generating mechanisms.

The u-plot procedure examines the closeness of the distribution of the u's to uniformity. The y-plot procedure examines trend. This leaves the independence of the u's to be examined. It might be useful to examine serial correlations of the u's.

In some situations, it is found that predictors with good y-plots have very poor u-plots (which measures "bias"). Such a predictor could be a good candidate for an adaptive procedure: essentially measuring the bias on earlier predictions in order to remove it on the current one. Some early work in this area is reported by Keiller and Littlewood (1984). This area is one which invites future work, particularly on the PL performance of adapted prediction systems.

Along with bias, the main source of error is unwarranted noisiness in some prediction systems. A prediction system which is unbiased but too noisy might be a good candidate for smoothing techniques. However, even if such smoothing techniques can be developed, we need to be able to test that the noise in a prediction system really is unwarranted (i.e. it is not truly reflecting a noisy "nature"). We do not have such a test, although the PL can sometimes be used to make this decision.

In section 4.3., we saw that there are sometimes reversals of relative performance between different prediction systems for a given data set. For example, for the data in Table 4.1., LV is better overall than JM, while JM is slightly superior at the early stages of analysis. Such behaviour suggests it might be better to use combinations of two or more prediction systems using PL as an objective function.

An attempt was made to generalize BJM and BL model to cope with discrete and summarized software reliability data. Although the mathematical form of the prediction system has been obtained, computational difficulties have frustrated attempts to obtain numerical results. In view of the wider availability of this type of data, it may be worth persevering with this work (particularly in the case of BL).

The adaptive procedure mentioned earlier can be seen as a non-parametric method for <u>local</u> prediction. By assuming stationarity of errors, the shape of an individual predictive density can be estimated from previous predictions. This observation opens up the possibiity of a completely non-parametric prediction system. If we had a non-parametric method of estimating the <u>trend</u>, we could use the adaptive procedure to make complete probability predictions for the future. Obviously, this trend estimation problem is harder than the density estimation implicit in the adaptive procedure. Miller (1984) has suggested one approach, based on generalisations of isotonic regression. Another approach might be to fit a sequence of means $\{m_i\}$ by minimising the y-plot distance of $\{t_i/m_i\}$ subject to suitable constraints.

APPENDIX A

Condition for finite estimates of DJM Model

Littlewood and Verrall (1981) proved a condition for finite estimates of the Jelinski-Moranda Model (JM) in the continuous case. This condition has an equivalent form:

APPENDIX

Condition for finite estimates of Discrete Jelinski-Moranda Model (DJM)

[illegible] A.1

where C_{i-1} = [illegible] in the discrete case.

Thus the likelihood function of DJM model given by equation (3.2.2) has a unique maximum at finite N and [illegible] if and only if the condition (A.1) holds. Otherwise the likelihood has its maximum at finite λ = [illegible] for infinite N.

Proof

Let x = 1/N (3.2.2) becomes:

[illegible]

[illegible] A.2

APPENDIX A

Condition for finite estimates of DJM Model

Littlewood and Verrall (1981) proved a condition for finite estimates of the Jelinski-Moranda Model (JM) in the continuous case. This condition has an equivalent form:

$$\frac{\sum_{i=1}^{k} C_{i-1} t_i}{\sum_{i=1}^{k} n_i C_{i-1}} > \frac{\sum_{i=1}^{k} t_i}{\sum_{i=1}^{k} n_i} \qquad \text{A.1.}$$

where $C_{i-1} = \sum_{j=1}^{k} n_i$, in the discrete case.

Thus the likelihood function of DJM model given by equation (3.2.2.) has a unique maximum at finite N and non-zero Φ if and only if the condition (A.1) holds. Otherwise the likelihood has its maximum at finite $\lambda = N\Phi$ for infinite N.

Proof

Let $X = 1/\Phi$; (3.2.2) becomes:

$$L(n_1, n_2, \ldots n_k/\lambda, X) = \prod_{i=1}^{k} \left[\left(\frac{\lambda X - C_{i-1}}{X}\right)^{n_i} \frac{t_i^{\,n_i}}{n_i!} \right]$$

$$\exp.\left(-\lambda \sum_{i=1}^{k} t_i + \frac{1}{X} \sum_{i=1}^{k} C_{i-1} t_i\right) \qquad \text{A.2}$$

i.e.

$$\ell = \log L(n_1, n_2, \ldots n_k/\lambda, \chi)$$

$$= \sum_{i=1}^{k} n_i \log\left(\frac{\lambda\chi - C_{i-1}}{\chi}\right) - \lambda \sum_{i=1}^{k} t_i + \frac{\sum_{i=1}^{k} C_{i-1} t_i}{\chi}$$

$$+ \sum_{i=1}^{k} n_i \log t_i - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log j$$

So,

$$\frac{\partial}{\partial\chi} \ell = \sum_{i=1}^{k} \frac{n_i \lambda}{\lambda\chi - C_{i-1}} - \frac{\sum_{i=1}^{k} n_i}{\chi} - \frac{\sum_{i=1}^{k} C_{i-1} t_i}{\chi^2}$$

$$= \frac{\sum_{i=1}^{k} n_i}{\chi} + \frac{\sum_{i=1}^{k} n_i C_{i-1}}{\lambda\chi^2} + O(\chi^{-3}) - \frac{\sum_{i=1}^{k} n_i}{\chi} - \frac{\sum_{i=1}^{k} C_{i-1} t_i}{\chi^2}$$

Clearly $\partial\ell/\partial\chi \to 0$ as $\chi \to 0$;

$\frac{\partial\ell}{\partial\chi}$ approaches zero from above if:

$$\lambda < \frac{\sum_{i=1}^{k} n_i C_{i-1}}{\sum_{i=1}^{k} C_{i-1} t_i} \qquad \text{A.3.}$$

If the above inequality is reversed, $\partial\ell/\partial\chi$ approaches zero from below as $\chi \to \infty$. We thus obtain the situation depicted in Figure A.1. The parameter space can be divided into two regions. In Region 1 the likelihood has its largest value at finite N,χ. In Region 2 there will be a maximum at infinite N,χ. Consider the likelihood on this "arc at infinity", it takes values:

$$\lim_{\chi\to\infty} L(n_1, \ldots n_k/\lambda, \chi) = \lambda^{\sum_{i=1}^{k} n_i} \left[\prod_{i=1}^{k} \frac{t_i^{n_i}}{n_i!}\right] \exp\{-\lambda \sum_{i=1}^{k} t_i\}$$

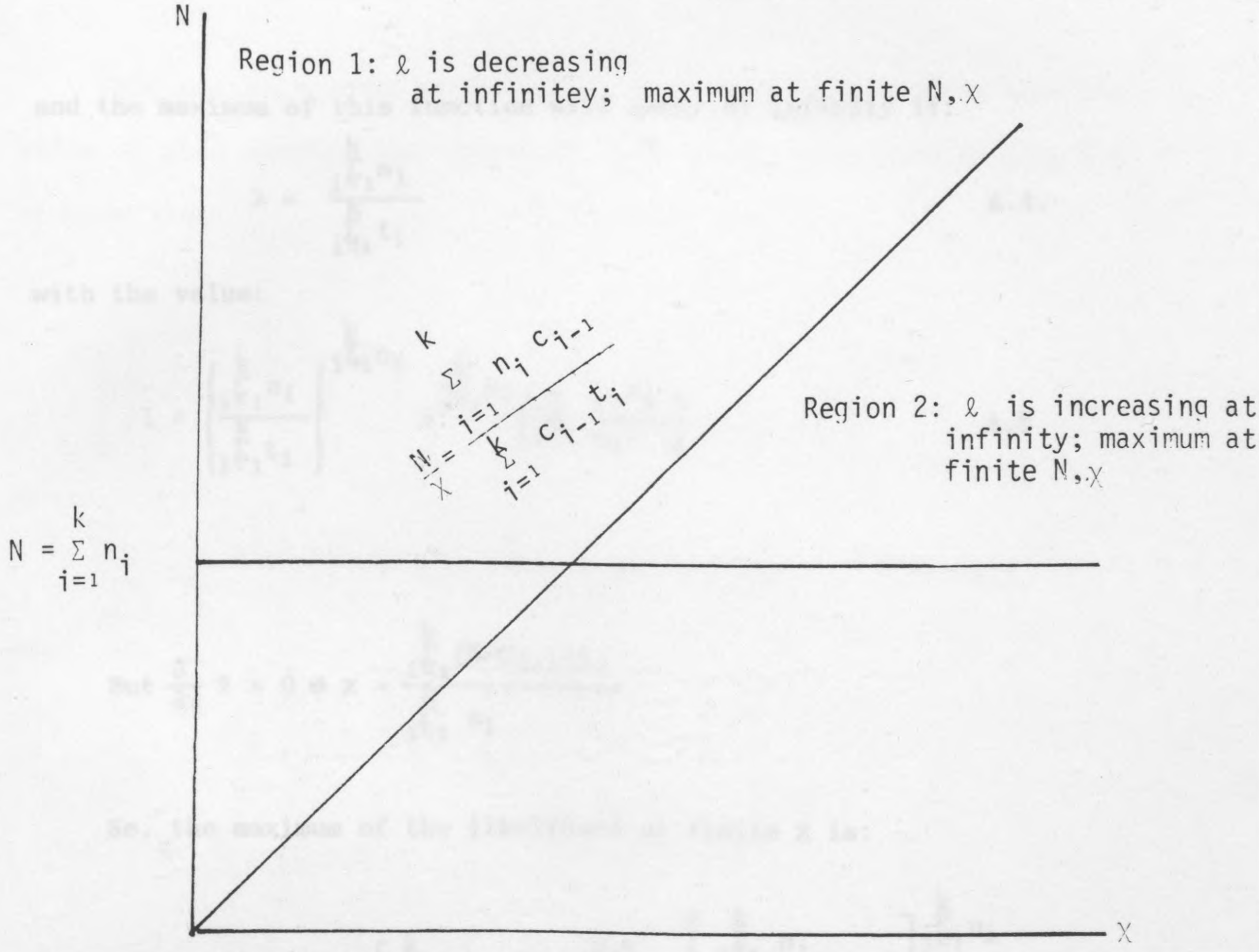


FIGURE A.1. The parameter space can be divided into two regions and the maximum of this function will occur at infinity if:

$$\lambda = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} t_i} \qquad \text{A.4.}$$

with the value:

$$L = \left\{\frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} t_i}\right\}^{\sum_{i=1}^{k} n_i} e^{-\sum_{i=1}^{k} n_i}\left[\prod_{i=1}^{k} \frac{t_i^{\,n_i}}{n_i!}\right] \qquad \text{A.5}$$

But $\frac{\partial}{\partial \chi} \ell = 0 \Rightarrow \chi = \frac{\sum_{i=1}^{k}(N-C_{i-1})t_i}{\sum_{i=1}^{k} n_i}$

So, the maximum of the likelihood at finite χ is:

$$L(N,\chi(N)) = \left[\prod_{i=1}^{k}(N-C_{i-1})^{n_i}\right] \left[\frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k}(N-C_{i-1})t_i}\right]^{\sum_{i=1}^{k} n_i}$$

$$.e^{-\sum_{i=1}^{k} n_i}\left[\prod_{i=1}^{k} \frac{t_i^{\,n_i}}{n_i!}\right] \qquad \text{A.6.}$$

From (A.3) and (A.4) the global maximum of the likelihood must occur at finite N,χ, if:

$$\frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} t_i} > \frac{\sum_{i=1}^{k} n_i C_{i-1}}{\sum_{i=1}^{k} C_{i-1} t_i} \qquad \text{A.7.}$$

This can be proved as follows. It is sufficient to show that the value of (A.5) exceeds the values of (A.6) at all finite (N,X) points, i.,e. to show that:

$$\left[\frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} t_i}\right]^{\sum_{i=1}^{k} n_i} > \left[\prod_{i=1}^{k} (N - C_{i-1})^{n_i}\right] \left[\frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} (N - C_{i-1}) t_i}\right]^{\sum_{i=1}^{k} n_i}$$

or

$$\sum_{i=1}^{k} t_i \sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} (N - C_{i-1})^{n_i}} < \sum_{i=1}^{k} (N - C_{i-1}) t_i \qquad \text{A.8.}$$

Now

$$\frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} t_i} < \frac{\sum_{i=1}^{k} n_i C_{i-1}}{\sum_{i=1}^{k} C_{i-1} t_i}$$

$$\Rightarrow \sum_{i=1}^{k} t_i \sum_{i=1}^{k} n_i C_{i-1} > \sum_{i=1}^{k} n_i \sum_{i=1}^{k} C_{i-1} t_i$$

$$\Rightarrow \quad N \sum_{i=1}^{k} n_i \sum_{i=1}^{k} t_i - \sum_{i=1}^{k} t_i \sum_{i=1}^{k} n_i C_{i-1} <$$

$$N \sum_{i=1}^{k} n_i \sum_{i=1}^{k} t_i - \sum_{i=1}^{k} n_i \sum_{i=1}^{k} C_{i-1} t_i$$

$$\Rightarrow \quad \left(\sum_{i=1}^{k} t_i\right)\left(\sum_{i=1}^{k} n_i (N - C_{i-1})\right) < \left(\sum_{i=1}^{k} n_i\right)\left(\sum_{i=1}^{k} (N - C_{i-1}) t_i\right)$$

$$\Rightarrow \quad \sum_{i=1}^{k} (N - C_{i-1}) t_i > \left(\sum_{i=1}^{k} t_i\right) \left[\frac{\sum_{i=1}^{k} n_i (N - C_{i-1})}{\sum_{i=1}^{k} n_i}\right] \qquad \text{A.9.}$$

But, it is well known that:

$$\frac{\sum_{i=1}^{k} n_i (N - C_{i-1})}{\sum_{i=1}^{k} n_i} > \sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} (N - C_{i-1})^{n_i}}$$

So (A.9) $\Rightarrow$

$$\sum_{i=1}^{k} (N - C_{i-1}) t_i > \left(\sum_{i=1}^{k} t_i\right) \sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} (N - C_{i-1})^{n_i}} \qquad \text{A.8.}$$

Uniqueness can be shown as follows. It is clear that the likelihood has a unique maximum at infinite (N,χ) when:

$$\hat{\lambda} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} t_i} = (\hat{N}\Phi) = (\hat{N}/\chi) \qquad \text{A.10.}$$

But the maximum of the likelihood at finite (N,χ) is given by (A.6) which contains four terms. The first term is a polynomial in N with no roots in the parameter space ($N \geq \sum_{i=1}^{k} n_i$) and the remaining three terms form a decreasing function in N.

Thus (A.6) has at most one turning point, in fact the maximum of (A.6) could occur at $N = \sum_{i=1}^{k} n_i$. This proves that there is only one maximum of likelihood for finite N,χ (non-zero Φ), since $\hat{\chi}$ is uniquely determined by $\hat{N}$.

REFERENCES AND BIBIOGRAPHY

ABDEL-GHALY, A.A., CHAN, P.Y. and LITTLEWOOD, B. (1985a)
Tools for the analysis of the accuracy of software reliability predictions. In NATO Advanced Study Institute, The Challenge of Advanced Computing Technology to System Design Methods, Durham, U.K., (to be published by Springer Verlag, Heidelberg).

ABDEL-GHALY, A.A., CHAN, P.Y. and LITTLEWOOD, B. (1985b)
Evaluation of Competing software reliability predictions. CSR Research Report, The Centre for Software Reliability, The City University, London, U.K., (to appear in IEEE Trans. Software Engineering).

ADAMS, E.N. (1984)
Optimizing preventive service of software products. IBM Journal of Research and Development, Vol. 28, No.1, pp. 2-14.

AITCHISON, J, and DUNSMORE, I.R. (1975).
Statistical Prediction Analysis. Cambridge University Press, Cambridge, U.K.

AKAIKE, H. (1982)
Prediction and Entropy. Technical Summary Report. Mathematics Research Centre, University of Wisconsin, Madison, U.S.A.

ASCHER, H., and FEINGOLD, H. (1984)
Repairable systems reliability. Lecture notes in Statistics, No.7, Marcel Dekker, New York, U.S.A.

BARLOW, R.E. and PROSCHAN, F. (1975)
Statistical Theory of Reliability and Life Testing. Holt, Reinhart and Winston Inc., New York, U.S.A.

BARLOW, R.E. and PROSCHAN, F. (1965)
Mathematical Theory of Reliability. New York Wiley, New York, U.S.A.

BARLOW, R., BARTHOLOMEW, D.J., BREMMER, J.M. and BRUNK, H.D. (1972)
Statistical Inference under Order Restrictions. John Wiley & Sons, Inc., New York, U.S.A.

BRAUN, H., and PAINE, J.M. (1977)
A Comparative study of model for reliability growth. Technical report No. 126, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey, U.S.A.

BRAUN, H.,and SCHENKER, N. (1980)
New models for reliability growth. Technical Report No. 174. Department of Statistics, Princeton University, Princeton, New Jersey, U.S.A.

BROOKS, W.D. and MOTLEY, R.W. (1980).
Analysis of discrete software reliability models. Technical Report distributed by Defence Technical Information Centre. Defense Logistics Agency, Cameron Station, Alexandria, Virginia 22314, U.S.A.

COCHRAN, W.G. (1952)
The X^2 test of goodness of fit. Ann. Math. Statist. 23, pp. 315-345.

COX, A.P. and LEWIS, P.A.W. (1966)
Statistical analysis of series of events. Methuen, London, U.K.

CROW, L.H. (1979)
Construction of idealised growth curves. Class Notes, US Army Material Systems Analysis Activity, Aberdeen, Md.

CROW, L.H. (1977)
Confidence interval procedures for reliability growth analysis. Technical Report No. 107, U.S.Army Materials Systems Analysis Activity, Aberdeen, Md.

DALE, J.C. (1982)

Software reliability evaluation methods. British Aerospace Dynamics Group, ST-26750.

DAVID, H. (1981)

Order Statistics. 2nd edition, John Wiley & Sons Inc., New York, U.S.A.

DAWID, A.P. (1985)

Probability forecasting. Encylopedia of Statistical Sciences. Vol. 6., (S.Kotz, N.L.Johnson and C.B.Read, Eds.) Wiley-Interscience.

DAWID,A.P. (1984a)

Statistical Theory, the prequential approach. J. Royal Statist. Soc. A. 147. pp. 278-292.

DAWID, A.P. (1984b)

Calibration based empirical probability. Research Report 36, Department of Statistical Science, University College London, U.K.

DAWID, A.P. (1982)

The well calibrated Bayesian, with discussion. J. Amer. Statist. Assoc., 77, pp. 605-613.

DUANE, J.T. (1964)

Learning curve approach to reliability monitoring. IEEE Trans. on Aerospace, Vol. 2, pp. 563-566.

DURBIN, J. (1975)

Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika, Vol. 62, No.1. pp. 5-22.

FELLER, W. (1971)

An introduction to probability theory and its application. John Wiley & Sons Inc., New York, USA.

FINKELSTEIN, J.M. (1976)
Confidence bounds on the parameters of the Weibull process. Technometrics, Vol. 18, No. 1. pp. 115-117.

FORMAN, E.H. and SINGPURWALLA, N.D. (1977)
An empirical stopping rule of debugging and testing computer software. J. Amer. Statist. Assoc., Vol. 72 pp. 750-757.

GIBBONS, J.D., OLKIN, I. and SOBEL, M. (1977)
Selecting and Ordering Populations, a new statistical methodology. John Wiley & Sons Inc., New York, U.S.A.

GOEL, A.L. (1980)
Software error detecting model with applications. The Journal of Systems and Software 1, pp. 243-249.

GOEL, A.L. and OKUMOTO, K. (1979)
Time-dependent error-detection rate model for software reliability and other performance measures. IEEE Trans. Reliability, Vol. R28. pp. 206-211.

JELINSKI, Z. and MORANDA, P. (1972)
Software reliability research. In: Statistical Computer Performance Evaluation. Academic Press Inc., London, U.K. pp. 465-484.

JEWELL, S.W. (1984)
A general framework for learning curve reliability growth models. Operations Research, Vol. 32, No. 3. pp. 547-558.

JOE, H. and REID, N. (1985)
Estimating the number of faults in a system. J. Amer. Statist. Assoc.,Vol. 80, pp. 222-226.

JOHNSON, N.L. and KOTZ, S. (1969)
Distribution in Statistics: Discrete distribution. John Wiley & Sons Inc., New York, U.S.A.

JOHNSON, N.L. and KOTZ, S. (1970a)
Distribution in Statistics: Continuous Univariate Distributions I. John Wiley & Sons Inc., New York, U.S.A.

JOHNSON, N.L. and KOTZ, S. (1970b)
Distribution in Statistics: Continuous Univariate Distributions II. John Wiley & Sons Inc., New York, U.S.A.

KARLIN, S. and TAYLOR, H.M. (1981)
A second course in Stochastic Processes. Academic Press, New York, U.S.A.

KEILLER, P.A. and LITTLEWOOD, B. (1984)
Adaptive software reliability modelling. Digest FTCS-14 (14th International Conference on Fault-Tolerant Computing), IEEE Computer Society, Silver Spring, pp. 108-113.

KEILLER, P.A., LITTLEWOOD, B., MILLER, D.R. and SOFER, A. (1983a)
On the quality of software reliability prediction. NATO ASI Series, Vol. F3. Electronic system effectiveness and life cycle costing, Edited by J.K.Skwirzynski, Springer-Verlag, Heidelberg.

KEILLER, P.A., LITTLEWOOD, B., MILLER, D.R. and SOFER, A. (1983b)
Comparison of software reliability predictions. Digest FTCS-13 (13th International Conference on Fault Tolerant Computing), IEEE Computer Society, Silver Spring, pp. 128-134.

KENDALL, M. and STUART, A. (1977)
The advanced theory of statistics. Vol. 1. "Distribution Theory", Fourth Edition, Charles Griffin & Co. Ltd., London, U.K.

KOPETZ, H. (1979)
Software Reliability. The Macmillan Press Ltd., London, U.K.

LANCASTER, H.O. (1969)

The Chi-Square distribution. John Wiley & Sons Inc., New York, U.S.A.

LANGBERG, N. and SINGPURWALLA, N.D. (1981)

A unification of some software reliability models, via the Bayesian approach. Technical memorandum serial TM-66571. The George Washington University, Washington, U.S.A.

LAPRIE, J.C. (1984)

Dependability evaluation of software systems in operation. IEEE Trans. Software Engineering, Vol. SE.10, No. 6, pp. 701-714.

LITTLEWOOD, B. (1984)

Rationale for a modified Duane model. IEEE Trans. on Reliability, Vol. R-33, No. 2, pp. 157-159.

LITTLEWOOD, B. (1982)

Stochastic reliability growth: A model with applications to computer software faults and hardware design faults. Unpublished manuscript. The Centre for Software Reliability, The City University, London, U.K.

LITTLEWOOD, B. (1981)

Stochastic reliability growth: A model for fault removal in computer programs and hardware design. IEEE Trans. on Reliability, Vol. R-30, No. 4. pp. 313-320.

LITTLEWOOD, B. (1980)

Theories of software reliability: How good are they and how can they be improved? IEEE Trans. on Software Enginering, Vol. SE-6, No. 5. pp. 489-500.

LITTLEWOOD, B. (1979a)
How to measure software reliability and how not to. IEEE Trans. on Reliability, Vol. R-28, No. 2. pp.103-110.

LITTLEWOOD, B. (1979b)
Software reliability model for modular program structure. IEEE Trans. on Reliability, Vol. R-28, No. 3. pp. 241-246.

LITTLEWOOD, B. (1978)
On the modelling of computer software reliability and quality. Ph.D. Thesis, The City University, Department of Mathematics, London, U.K.

LITTLEWOOD, B. (1976)
A semi-Markov model for software reliability with failure costs. In: M.R.I. Symposium on Computer Software Engineering, Vol. XXIV, Polytechnic Press, Polytechnic Institute of New York, pp. 281-300.

LITTLEWOOD, B. (1975)
A reliability model for sytems with Markov structure. The Journal of the Royal Statistical Society Series C (Applied Statistics), Vol. 24, No. 2. pp. 172-177.

LITTLEWOOD, B. and SOFER, A. (1981)
A Bayesian modification to the Jelinski-Moranda software reliability growth model. Unpublished manuscript, The Centre for Software Reliability, The City University, London, U.K.

LITTLEWOOD, B. and VERRALL, J.L. (1981)
Likelihood function of a debugging model for computer software reliability. IEEE Trans. on Reliability, Vol. R-30, No. 2, pp. 145-148.

LITTLEWOOD, B. and VERRALL, J.L. (1973)
A Bayesian reliability growth model for computer software. The Journal of The Royal Statistical Society, Series C, (Applied Statistics), Vol. 22., No. 3. pp. 332-346.

MELLAR, P. (1984)
Private Communication

MILLER, D.R. (1983)
Private Communication

MILLER, D.R. (1984)
Exponential order statistic models of software reliability growth. Technical Report T-496/84, George Washington University, Washington, U.S.A.

MISRA, P.N. (1983)
Software reliability analysis. IBM Systems Journal, Vol. 22, No. 3. pp. 262-270.

MOEK, M. (1984)
Software reliability model on trial: selection, improved estimation and practical results. Unpublished paper, National Aerospace Laboratory, NLR, Informatics Division, P.0. Box 90502, 1006 BM, Amsterdam, The Netherlands.

MUSA, J.D. (1980a)
The measurement and management of software reliability. Proceedings of the IEEE, Vol. 68, No. 9. pp. 1131-1143.

MUSA, J.D. (1980b)
Software reliability measurement. The Journal of Systems and Software 1, pp. 223-241.

MUSA, J.D. (1979)
Software reliability data. Report available from Data and Analysis Centre for Software, Rome Air Development Centre, Rome, N.Y.

MUSA, J.D. (1975)
A theory of software reliability and its application. IEEE Trans. Software Engineering, Vol. SE-1 pp. 312-327.

MUSA, J.D. and OKUMOTO, K. (1984)
A logarithmic Poisson execution time model for software reliability measurement. Proceedings of 7th International Conference on Software Engineering, Orlando, Florida, U.S.A. IEEE Computer Society, Silver Spring, pp. 230-238.

NAGEL, P.M. and SKRIVAN, J.A. (1981)
Software reliability; repetitive run experimentation and modelling. BCS-40399, Boeing Computer Services Company, Seattle, Washington, U.S.A.

NELDER, J.A. and MEAD, R. (1965)
A simplex method for function minimization. Computer Journal, Vol. 7. pp. 308-313.

NIELSEN, B.O. (1978)
Information and exponential families in statistical theory. John Wiley & Sons, New York, U.S.A.

OBHA, M. (1984)
Software reliability analysis models. IBM Journal Research Develop. Vol. 28, No. 4. pp. 428-443.

SHOOMAN, M.L. (1973)
Operational testing and software reliability estimation during program development. Record IEEE Symp. Computer Software Reliability New York, N.Y. (April 30- May 2), pp. 51-57.

SHOOMAN, M.L. (1972)

Probability models for software reliability prediction. Statistical Computer Performance Evaluation, Academic Press Inc., London, U.K, pp. 485-502.

WALBERG, J.R. (1967)

Prediction Analysis. D.Van Nostrand Company Ltd., London, U.K.

WILKES, S.S. (1962)

Mathematical Statistics. John Wiley & Sons Inc., New York, U.S.A.