1986-03

Robust empirical Bayes analyses of event rates

Gaver, Donald Paul; O'Muircheartaigh, I. G.

Monterey, California. Naval Postgraduate School

NPS55-86-007

NAVAL POSTGRADUATE SCHOOL

Monterey, California

ROBUST EMPIRICAL BAYES ANALYSES OF EVENT RATES

DONALD P. GAVER
IOGNAID G. O'MUIRCHEARTAIGH

MARCH 1986

Prepared for:
Naval Postgraduate School
Monterey, CA 93943-5000

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral R. H. Shumaker
Superintendent

D. A. Schrady
Provost

This report was prepared by:

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) NPS55-86-007	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (If applicable) Code 55	7a NAME OF MONITORING ORGANIZATION
6c ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	7b ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING/SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)

ROBUST EMPIRICAL BAYES ANALYSES OF EVENT RATES

12 PERSONAL AUTHOR(S)

Gaver, Donald P. and O'Muircheartaigh, Iognaid G.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical	FROM _____ TO _____	1986, March	53

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Poisson, likelihood, empirical Bayes, robustness

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

A number, I, of nominally similar items generate events (e.g. failures) at possibly different rates, or mean time intervals. This paper addresses the problem of appropriately pooling the data from the different sources. The approach is parametric empirical Bayes: true individual item rates are assumed to come from a fixed superpopulation. It is shown how (a) parameters of a superpopulation model can be estimated from all of the data, and (b) combined with individual unit history, can provide improved estimates of individual rates. The procedure can be robust: evidence that a particular rate is far off from the main body of rates permits that outlier to stand by itself, i.e. to resist pooling. Illustrative analyses of data are supplied.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL Donald P. Gaver	22b TELEPHONE (Include Area Code) (408)646-2605 22c OFFICE SYMBOL Code 55Gv

1. Introduction

In reliability problems, but also in studies of logistics, congestion systems and elsewhere, it is common to encounter collections of nominally similar equipments or other entities that generate point events at similar, but not identical, rates. The questions then arise as to whether evidence for differences in the rates can be elicited from event rate data on all members of such a collection, and how the data can be well utilized to provide strengthened estimates of the underlying true rates of the individual equipments. If all equipments seem to have about the same failure rate then there should be little harm in calculating a simple pooled rate and quoting it for all members, while if evidence of considerable difference between members is present, then the individual rates seem most appropriate. Some form of compromise will be worthwhile for intermediate cases. The following general setup formalizes situations and provides compromise estimates that tend to pool the data.

Consider a collection of J equipments or other units that independently generate events in accordance with Poisson processes of constant rate λ_i. Observations of these processes are available: for unit i, $s_i (=0,1,2...)$ events have been observed over an exposure time interval t_i, i=1,2,...I. To describe the possible variability between rates, characterize λ_i as the independent realization of a random variable $\underset{\sim}{\lambda}$ with fixed parametric density function $g_{\underset{\sim}{\lambda}}(\cdot;\underline{\theta})$, where $\underline{\theta}$ is a generic vector parameter. The density $g_{\underset{\sim}{\lambda}}$ can be said to describe a superpopulation of rate parameters, sample values from

which have been bestowed upon the units of interest. The first objective of the analysis will be to utilize all available data to estimate the prevailing superpopulation parameters, $\underline{\theta}$; the second is to mobilize the estimated superpopulation parameters, possibly by Bayes' formula or an alternative, to provide suitably pooled or shrunken estimates for individual rates. Both point and interval estimates are desirable. Models of the above type are called parametric empirical Bayes (PEB) models; see Morris (1983) for a review with various references. Our present approach emphasizes superpopulation specifications that lead to robust estimates in the sense that the possibility of widely discrepant rates or exponential parameters is automatically dealt with by the superpopulation form. Such performance can be called discrepancy tolerant; it resembles in various ways the behavior of modern robust location estimation and regression techniques, cf. Mosteller and Tukey (1977); we call our procedure robust parametric empirical Bayes (RPEB). General ideas of robust Bayesian analyses have been described by Berger (1980, 1984); Albert (1979) in an unpublished study considers the Poisson case. The simultaneous estimation of Poisson means has been considered by many authors; a recent high-level account is by Johnstone (1984), who provides many references. See also Martz, H. F. and Waller, R. A. (1982), which describes work in the system reliability areas.

The model described is simplistic in recognizing just two sorts of variability in point event data: the ordinary, Poissonian sampling variation of observations around a given λ-value ("within" variations) and the variation of the individual λ-values around a fixed, unknown value ("between" variation). Of course, many elaborations are possible. A

natural possibility to consider is that rate variation is controlled in part by operational factors such as temperature, vibration, maintenance frequency and adequacy, etc., describable by a regression model. Another possibility is that individual rates are themselves realizations of random processes, possibly with the addition of trends, thus requiring representation of time-dependent over-Poisson variations; see Cox and Lewis (1966) and McCullagh and Nelder (1983), pp. 131-133. The present paper does not deal with these, but extensions are in progress.

The emphasis of this paper is data-analytical. Algorithms are first constructed for estimation of superpopulation parameters; confidence regions associated with these are constructed and displayed graphically. The superpopulation parameter estimates are then applied to compute point and associated interval estimates of individual rate parameters. Much of this latter process is carried out numerically and displayed graphically as well. New shortcut and computationally economical approximate solutions to the above problems are furnished and compared to complete Bayes solutions. The procedures are applied to three sets of reliability data, and the results are discussed. Despite the formal probabilistic underpinnings described for the procedure, it seems reasonable to apply the methods in an exploratory fashion to probe for structure in data sets. This process has been briefly illustrated for one example.

2. Some Illustrative Data Sets

Here are some data sets that serve to motivate our later analyses.

2.1 Failure rates of air-conditioning equipment.

A classical data set to which our analysis appears applicable is that of failures of air conditioning equipment on 13 Boeing 720 aircraft; these data were originally provided by Proschan (1963), and have been much studied. We consider an initial analysis that takes each aircraft to have a constant individual mean time to failure, λ_i^{-1} (i=1,2,3...I=13) and an i.i.d. exponential time to failure. The data can be summarized in terms of numbers of failures over an exposure time; see Cox and Lewis (1966); for further discussion see Cox and Snell (1980).

Note that actual time-to-individual-failure data is available for each individual equipment. An initial data analysis of each unit's failure pattern failed to reveal substantial trend or evidence of departure from an exponential failure law. The likelihood function for λ_i, an individual exponential law parameter, is of the Poisson-gamma form with s_i the sufficient statistic, so the data is presented as such in Section 3, and provisionally analyzed to elicit between-λ_i variability. The columns headed r_i in the following tables include the raw quotient (individual mle) rates $r_i=s_i/t_i$. The cases in our tables are ordered by increasing raw failure rate, r_i.

Table 1.1

Air-conditioner Failures

Aircraft No. i	s_i	t_i	r_i
11	2	0.623	3.21
9	9	1.800	5.00
5	14	1.832	7.64
4	15	1.819	8.25
12	12	1.297	9.25
10	6	0.639	9.39
2	23	2.201	10.45
3	29	2.422	11.97
1	6	0.493	12.17
13	16	1.312	12.20
7	27	2.074	13.02
8	24	1.539	15.59
6	30	1.788	16.78

A maximum likelihood ratio test (Cox and Lewis, pp. 235-236) further indicates that the individual parameters are significantly different at about a 2 percent level. Thus these data can be expected to exhibit some between-rate variability, as measured by a scale (e.g. standard deviation) parameter of the superpopulation.

2.2. Loss of feedwater flow.

Table 2 presents a set of data referring to the rates of loss of feedwater flow for a collection of nuclear power generation systems; see Kaplan (1983). This class of "initiating events" is important in the probabilistic risk assessment (PRA) of nuclear plants. It has been treated in an empirical Bayesian fashion by Kaplan, although he does not use that terminology.

Table 1.2

LOSS OF FEEDWATER FLOW

System$_i$	s_i	t_i	r_i
3	0	8	0.041
19	0	2	0.17
1	4	15	0.27
7	2	5	0.4
18	1	2	0.5
8	4	4	1.0
16	3	3	1.0
25	1	1	1.0
4	10	8	1.3
10	4	3	1.3
15	4	3	1.3
5	14	6	2.3
13	10	4	2.5
27	5	2	2.5
20	3	1	3.0
2	40	12	3.3
9	13	4	3.3
26	10	3	3.3
12	14	4	3.5
14	7	2	3.5
24	12	3	4.0
28	16	4	4.0
29	14	3	4.7
21	5	1	5.0
30	58	11	5.3
17	11	2	5.5
22	6	1	6.0
6	31	5	6.2
11	27	4	6.8
23	35	5	7.0

2.3 Pump Reliability data at a pressurized water reactor (PWR) nuclear power plant.

In Table 1.3, there appears a small set of data representing failures of pumps in several systems of the nuclear plant Farley 1. The apparent variation in failure rates has several possible sources; some are mentioned later. These data may be found in an EPRI report (1982).

Table 1.3

PUMP FAILURES

System$_i$	s_i	t_i	r_i
1	5	94.320	5.3×10^{-2}
2	1	15.720	6.4×10^{-2}
3	5	62.880	8.0×10^{-2}
4	14	125.760	11.1×10^{-2}
5	3	5.240	57.3×10^{-2}
6	19	31.440	60.4×10^{-2}
7	1	1.048	95.4×10^{-2}
8	1	1.048	95.4×10^{-2}
9	4	2.096	191×10^{-2}
10	22	10.480	209.90×10^{-2}

3. Specific PEB Models

Consider two parametric families as representations of an assumed superpopulation for the event rates. These are (i) the centered and scaled log-Student t, which includes the log-normal when the degrees of freedom parameter, n, becomes infinite; and (ii) the gamma.

Form (i), the log-Student, is of potential interest in probabilistic risk analysis of nuclear power systems (PRA), where the log-normal form has long been used to characterize variability between failure rates for various equipments; see the Reactor Safety Study (1975), and Kaplan (1983). The log-Student generalizes the log-normal setup, admitting systematically heavier-than-normal/Gaussian tails and so allowing for a greater-than-Gaussian propensity for extreme outliers for the rates. The tail behavior is regulated by n, the Student degrees of freedom parameter. We do not here attempt to estimate n from data, but treat it as a tuning parameter, much in

the manner of the tuning constant, c, appearing in bi-weight regression; see Mosteller and Tukey (1977). Form (ii), the gamma, is the natural conjugate prior associated with the Poisson likelihood, and hence yields pleasant analytical simplicity.

Here are the formal descriptions of the PEB models considered.

Log-Student: Stage 1 (Sampling individual rates from the superpopulation):

$$\underset{\sim}{\lambda}_i = \exp(\underset{\sim}{\varepsilon}_i)$$

$$\underset{\sim}{\varepsilon}_i \sim g_{\underset{\sim}{\varepsilon}}(z;\mu,\tau;n) = \frac{C(n)}{\left[1 + \left(\frac{z-\mu}{\tau}\right)^2 \frac{1}{n}\right]^{(n+1)/2}} \qquad (3.1)$$

C(n) being the appropriate normalizing constant; $\{\underset{\sim}{\varepsilon}_i,\ i=1,2,\ldots,I\}$ is a sequence of independent random variables.

Stage 2 (Observations from the individual rates):

$$\underset{\sim}{s}_i \,|\, \underset{\sim}{\lambda}_i \sim \text{Poisson}\ (\underset{\sim}{\lambda}_i t_i) \qquad (3.2)$$

Apparently $\underset{\sim}{\varepsilon}_i \sim \Phi\left(\frac{z-\mu}{\tau}\right)$, the normal distribution, as $n \to \infty$, this is the log normal model favored by many PRA analysts. Note that in general

$$\mathrm{Var}[\underset{\sim}{\varepsilon}_i] = \mathrm{Var}[\ln \underset{\sim}{\lambda}_i] = \left(\frac{n}{n-2}\right)\tau^2, \quad n>2.$$

Gamma: Stage 1: $$\underset{\sim}{\lambda}_i \sim g_{\underset{\sim}{\lambda}}(w;\alpha,\beta) = e^{-\alpha w}\left(\frac{(\alpha w)^{\beta-1}\alpha}{\Gamma(\beta)}\right)$$

(3.3)

Stage 2: $$s_i \mid \underset{\sim}{\lambda}_i \sim \text{Poisson}\,(\underset{\sim}{\lambda}_i t_i). \tag{3.4}$$

There seems to be no fundamental justification for either parametric superpopulation form. Generally, the log-Student is appealing because of its controllable long tails and the ease of interpretation of normal variation, while the gamma has mathematical convenience to recommend it. Neither represents truly eccentric behavior such as multi-modality or extensive asymmetry -- features that cannot be ruled out in real data. See Laird (1978) and Copas (1984) for non-parametric approaches to this problem, and Tukey (1974) for an exploratory, totally non-probabilistic approach to a large data set of similar structure.

4. Fitting the Superpopulation Models: Stage 1.

Given data of the form exhibited in Tables 1.1, 1.2, and 1.3, it is possible to estimate the parameters in the superpopulation form by various methods. We examine two: simple moment matching and maximum likelihood.

4.1 Crude moment matching.

From the Poisson assumption and familiar conditioning arguments, one can obtain these formulas:

$$E[s_{\sim i}|\lambda_{\sim i}] = \lambda_{\sim i} t_i = \mathrm{Var}[s_{\sim i}|\lambda_{\sim i}], \tag{4.1}$$

$$E[s_{\sim i}] = E[\lambda_{\sim i}]t_i;\ \mathrm{Var}[s_{\sim i}] = E\{\mathrm{Var}[s_{\sim i}|\lambda_{\sim i}]\} + \mathrm{Var}\{E[s_{\sim i}|\lambda_{\sim i}]\}$$

So, unconditionally,

$$E[s_{\sim i}] = E[\lambda_{\sim}]t_i,$$

$$\mathrm{Var}[s_{\sim i}] = E[\lambda_{\sim}]t_i + \mathrm{Var}[\lambda_{\sim}]t_i^2. \tag{4.2}$$

Consequently if the raw rates are modelled by $r_{\sim i} = s_{\sim i}/t_i$,

$$E[r_{\sim i}] = E[\lambda_{\sim}] \tag{4.3}$$

$$\mathrm{Var}[r_{\sim i}] = \mathrm{Var}[\lambda_{\sim}] + E[\lambda_{\sim}]\frac{1}{t_i} \tag{4.4}$$

which suggests that crude estimates for $E[\lambda_{\sim}]$ and

$\mathrm{Var}[\lambda_{\sim}]$ can be obtained by matching moments:

$$\hat{E}[\underset{\sim}{\lambda}] = \bar{r}, \quad \hat{Var}[\underset{\sim}{\lambda}] = s_r^2 - \bar{r}\left(\frac{1}{I}\sum_{i=1}^{I} t_i^{-1}\right) \tag{4.5}$$

Now for the specific forms considered we know that

$$\underline{\text{log-Normal}}: \quad E[\underset{\sim}{\lambda}] = \exp\left(\mu + \frac{\tau^2}{2}\right); \quad Var[\underset{\sim}{\lambda}] = (E[\underset{\sim}{\lambda}])^2(e^{\tau^2} - 1) \tag{4.6}$$

$$\underline{\text{Gamma}}: \quad E[\underset{\sim}{\lambda}] = \frac{\beta}{\alpha}; \quad Var[\underset{\sim}{\lambda}] = \frac{\beta}{\alpha^2} \tag{4.7}$$

and so both μ and τ^2, or α and β, can be assessed, perhaps inefficiently but very conveniently, by using (4.5) in conjunction with (4.6) or (4.7). Note that $E[\underset{\sim}{\lambda}]$, and hence $Var[\underset{\sim}{\lambda}]$, is not finite for the log-Student model, and hence this simplest moment-matching procedure is inapplicable. Under the circumstance that s_i is large, accurate moment approximations for $\ln(\underset{\sim}{r}_i) = \ln(\underset{\sim}{s}_i/t_i)$ are obtainable for the Student superpopulation, provided the Student parameter n is large enough (i.e. > 2). A more refined iterative estimating procedure has been furnished for fitting the gamma in Hill, et al (1984), but the above formulas are extremely simple and useful for quick appraisals, and handy as a start for iterative likelihood maximization.

4.2 Likelihood Methods.

It is anticipated that the method of maximum likelihood will provide results superior to crude moment matching at the expense of greater

computational effort, particularly for the log-Student form. Here are the likelihoods, and comments concerning their maximization.

log-Student: The likelihood contribution of observation i is, up to irrelevant constants,

$$L_i = L_i(\mu, \tau;\ s_i, t_i; n)$$

$$= \int_{-\infty}^{\infty} e^{-\lambda(z)t_i} [\lambda(z)]^{s_i} \frac{dz}{\left[1 + \left(\frac{z-\mu}{\tau}\right)^2 \frac{1}{n}\right]^{(n+1)/2}} \qquad (4.8)$$

with $\lambda(z) = \exp(z)$, so the total likelihood is

$$L(\mu, \tau; \underline{s}, \underline{t}; n) = \prod_{i=1}^{I} L_i(\mu, \tau; \underline{s}, \underline{t}; n) \qquad (4.9)$$

The integration, and subsequent maximization, must be carried out numerically. Integration has been performed by several alternative Gauss-Hermite procedures. The first begins by approximating the integral by Laplace's Method, and concludes by Gauss-Hermite integration of a correction term remaining after the Laplace effect is removed; see Gaver (1985) for details; for short, this process will be called LGH. The second is a direct Gauss-Hermite procedure adapted from Naylor and Smith (1982); we are grateful to J. C. Naylor for furnishing a FORTRAN program that has served as the basis for this aspect of our work; call this GH. The maximization was accomplished in the first method by a refined grid search, and in the second by a quasi-Newton procedure adapted from IMSL SUBROUTINE ZXMIN, operating on

the log-likelihood surface. The classical EM algorithm discussed by Dempster, Laird and Rubin (1977) is applicable for estimating the superpopulation parameters, but does not directly produce approximate confidence regions, as obtained below.

Gamma: The likelihood contribution of observation i can be derived by integration, and is the negative binomial expression

$$L_i(\alpha,\beta;s_i,t_i) = \frac{\Gamma(s_i+\beta)}{\Gamma(\beta)} \frac{t_i^{s_i}\,\alpha^{\beta}}{(t_i+\alpha_i)^{s_i+\beta}}. \qquad (4.10)$$

In view of independence, a product of these contributions provides the total likelihood, as in (4.9). Maximization of the log likelihood has then been carried out by the IMSL procedure.

The numerical results obtained from applying the above procedures to the three illustrative data sets are given in Figs. 4.1, 4,2, and 4.3. In order to ease the comparison of the log-Student and gamma analyses, we have re-parameterized the gamma in terms of μ and τ, the latter being the parameters of a log-normal. Thus the μ and τ log-normal values that match the first two gamma moments are

$$\mu = \ln\left[\frac{\beta/\alpha}{\sqrt{1+1/\beta}}\right], \quad \tau = \sqrt{\ln(1+1/\beta)}\,; \qquad (4.11)$$

these expressions have been used to parameterize the gamma-likelihood for graphical display.

4.3 Approximate Confidence Regions for Superpopulation Parameters

The likelihood ratio test procedure has been used to define an approximate joint confidence region for μ and τ for the two superpopulation model families. The procedure specifies that all (μ, τ) values such that $-2[\ln(L(\mu,\tau;\underset{\sim}{s},\underset{\sim}{t};n)/L(\hat{\mu},\hat{\tau};n))] \leq \chi^2_2(1-\alpha)$, where $(\hat{\mu},\hat{\tau})$ is the mle, constitute an approximate $(1-\alpha)\cdot 100\%$ confidence region. The regions obtained for the three sets of data appear on the figures. The somewhat eccentric, but unimodal, shape of the log-likelihood surface is exhibited by the confidence contour plots; a bit more symmetry can in principle be obtained by re-parameterizing in terms of $\ln \tau$, but for our data sets the effect was not dramatic. The confidence contours are roughly elliptical with the ellipse semi-axes nearly parallel to the μ-τ axes; an analysis based on the simplifying assumption that $\hat{\mu}$ and $\hat{\tau}$ are independently bivariate normal works reaonably well for our data. The ellipticity tends to disappear when the data set is small and contains several $s_i=0$ values; as anticipated the region then often intersects the $\tau=0$ axis, suggesting that the data are consistent with a single underlying parameter value: $\lambda_i=\lambda$, $i=1,2,\ldots,I$, if the intersection is pronounced.

5. Individualized ("Shrunken" or "Pooled") Estimates.

If the true values of μ and τ (log-Student) or α and β (gamma) superpopulations were available, then an obvious step would be to compute the Bayes posterior of $\underset{\sim}{\varepsilon}_i = \ln \underset{\sim}{\lambda}_i$, or of $\underset{\sim}{\lambda}_i$ in the gamma case, given the value of $\underset{\sim}{s}_i$. Then any point or interval estimates desired could be

computed. Such calculations must be done numerically for the log-Student family, but are eased in the gamma case by the conjugate prior assumption. If the values of μ and τ are estimated from data, as suggested here, then approximate superpopulations can be derived by replacing (μ,τ) by $(\hat{\mu},\hat{\tau})$ and calculating the corresponding Bayes estimates; see Deely and Lindley (1981) for discussion of this empirical Bayes approach. Recent work by Morris (1983) and Hill (1984) suggests refinements to the simple procedure described. Use of the approximate individualized superpopulations (approximate Bayesian posteriors) then leads to point estimates and intervals. We have chosen to first calculate (i) the <u>means</u> $\bar{\varepsilon}_i = E[\underset{\sim}{\varepsilon}_i | \underset{\sim}{s}_i]$, of the posterior distributions for the individual unit log rates, ε_i, in the illustrative data sets; these can be compared with ordinary log raw rates, $\ln(s_i/t_i)$; (ii) the <u>standard deviations</u> $\sigma_i = [\mathrm{var}[\underset{\sim}{\varepsilon}_i | \underset{\sim}{s}_i]^{1/2}$, of the posterior distributions, (iii) approximately 95% upper tolerance limits (or 95% one-sided Bayes credibility intervals) for each unit, based on a normal approximation: $\tilde{\varepsilon}_i(0.95) = \bar{\varepsilon}_i + 1.645\sigma_i$, and (iv) upper confidence limits for the credibility intervals (iii), that recognize sampling variability in $\hat{\mu}$ and $\hat{\tau}$. We are encouraged to believe that such intervals are reasonable by looking at plots of the posterior densities of the ε_i; see Figures 4.1, 4.2, and 4.3. More exact calculations are possible by numerical integration of the posterior. Explicit expressions for the above quantities are these:

<u>Log-Student:</u> The approximate conditional means and second moments are cases of

$$E[\underset{\sim}{\varepsilon}_i^{\,k}|s_i]=\frac{1}{L_i(\hat{\mu},\hat{\tau};s_i,t_i;n)}\int_{-\infty}^{\infty} z^k\, e^{-\lambda(z)t_i}[\lambda(z)]^{s_i}\,\frac{dz}{\left[1+\left(\frac{z-\hat{\mu}}{\hat{\tau}}\right)^2\frac{1}{n}\right]^{(n+1)/2}} \tag{5.1}$$

again integrated by Gauss-Hermite quadrature. The normalized integrand of (5.1), exclusive of z^k, is the approximate Bayes posterior density of $\underset{\sim}{\varepsilon}_i$, given s_i.

Gamma: The mean and variance of the approximate gamma posterior have familiar pleasant explicit forms:

$$E[\underset{\sim}{\lambda}_i|\underset{\sim}{s}_i] = \frac{\underset{\sim}{s}_i + \hat{\beta}}{t_i + \hat{\alpha}} \tag{5.2}$$

and

$$\mathrm{Var}E[\underset{\sim}{\lambda}_i|\underset{\sim}{s}_i] = \frac{\underset{\sim}{s}_i + \hat{\beta}}{(t_i + \hat{\alpha})^2}. \tag{5.3}$$

There are no such simple expressions for $\tilde{\varepsilon}_i = \ln\underset{\sim}{\lambda}_i$ in the gamma case, but the posterior moments have been computed by Gauss-Hermite quadrature applied to the log-transformed gamma density.

5.1 Analytical Approximations to the Posterior

Although the above numerical computations are feasible, it is often useful to have relatively simple and explicit, if approximate, formulas for point estimates and posterior densities. One such can be derived for the log-Student model by writing the posterior density as

$$g_{\underset{\sim}{\varepsilon}}(z|s_i) = K\, e^{-(1/2)Q(z)} \tag{5.4}$$

and then approximating $Q(z)$ by a quadratic $q(z) = (1/2)\left(\frac{z-\mu(z)}{\sigma(z)}\right)^2$, in the manner of Laplace's method, c.f. N. G. de Bruijn (1957); for applications to Bayesian statistics see Mosteller and Wallace (1964), and Kadane and Tierney (1985). Differentiation shows that the minimum of $Q(z)$ occurs at $\hat{\varepsilon}_i = \bar{\varepsilon}_i$, where $\hat{\varepsilon}_i$ is the modal, or maximum likelihood, estimate of $\underset{\sim}{\varepsilon}_i|\underset{\sim}{s}_i$, and $\bar{\varepsilon}_i$ is the posterior mean.

Log-Student: The derivative of

$$Q(z) = -e^z t_i + s_i z - \left(\frac{n+1}{2}\right) \ln\left[1 + \left(\frac{z-\hat{\mu}}{\hat{\tau}}\right)^2 (1/n)\right] \tag{5.5}$$

set equal to zero yields an estimating equation that can be written as follows:

$$\hat{\lambda}_i = e^{\hat{\varepsilon}_i} = \left[s_i - \left(\frac{\hat{\varepsilon}_i - \hat{\mu}}{\hat{\tau}^2}\right) w_n(\hat{\varepsilon}_i)\right](1/t_i), \tag{5.6}$$

where the weight

$$w_n(\hat{\varepsilon}_i) = \left(\frac{n+1}{n}\right) \frac{1}{1 + \left(\frac{\hat{\varepsilon}_i - \hat{\mu}}{\hat{\tau}}\right)^2 (1/n)} \tag{5.7}$$

Graphical or analytical examination reveals the possibility of two solutions of $Q'(z) = 0$, one very near $\hat{\mu}$ and the other near $\ln(s_i/t_i)$, corresponding to a bimodal posterior. Convenient explicit analytical criteria for bimodality to occur are not available; but neither our present data sets, nor many others, have revealed such bimodality when the posterior was evaluated numerically and plotted. Our approach is to replace $w_n(\hat{\varepsilon}_i)$ by $\hat{w}_n = w_n(\ln(s_i/t_i))$ and by $\hat{w}_n = w_n(\ln(1/(3t_i))$ when $s_i=0$. This approximate weight leads to unique solutions of (5.6) by Newton-Raphson iteration. An interesting and interpretable formula is obtained after one iteration, starting with $\hat{\varepsilon}_i(0) = \ln(s_i/t_i)$:

$$\hat{\varepsilon}_i(1) = \frac{s_i \ln(s_i/t_i) + (\hat{\mu}/\hat{\tau}^2)\hat{w}_n}{s_i + (1/(\hat{\tau})^2)\hat{w}_n} = \ln(s_i/t_i) - \frac{(\ln(s_i/t_i)-\hat{\mu})(1/(\hat{\tau}^2))\hat{w}_n}{s_i + (1/(\hat{\tau})^2 w_n}. \tag{5.8}$$

This expression resembles the familiar Bayes normal-theory formula for combining prior mean and likelihood to obtain a linear estimator of the posterior mean. The difference is the presence of the weight $\hat{w}_n$, the effect of which is to reduce the influence of the mean of the superpopulation (prior) upon tail-discrepant observations. Discrepant observations ($\hat{w}_n$ small) are not heavily shrunken or pooled towards the estimated center, $\hat{\mu}$,

while observations close to that center (w_n large) are shrunken in that direction. Actually, the net shrinkage is caused by the factor $(1/(\hat{\tau})^2\hat{w}_n/[s_i + ((\hat{\tau})^2)\hat{w}_n]$, in which $\hat{w}_n$ plays an important but not exclusive role: the (approximate) variance s_i and $(\hat{\tau})^2$ are also significant (note that $\hat{w}_n$ depends upon $\hat{\tau}$ and upon $\ln(s_i/t_i)$, so shrinkage is not linear). Notice that as $n\to\infty$, and the log-Student approaches the log-normal, the discrepancy-tolerant effect diminishes; when $n=\infty$ there is only one solution to (5.5) and skrinkage becomes linear (provided the effect of observation i on $\hat{\mu}$ and $\hat{\tau}$ is small, as it usually is). The variance of the posterior can be assessed from the second derivative of Q(z); from (5.5)

$$\mathrm{Var}[\underset{\sim}{\varepsilon}_i | s_i] = \sigma_i^2 \approx \frac{1}{e^{\hat{\varepsilon}_i} t_i + (1/\hat{\tau}^2)\hat{w}_n} \qquad (5.9)$$

This formula again exhibits the behavior of the variance associated with the posterior encountered when normal likelihoods are combined with normal conjugate priors, <u>except</u> that wildly tail-discrepant observations are substantially downweighted: automatically in such cases $\hat{\varepsilon}_i = \ln(s_i/t_i)$ and $\sigma_i^2 = 1/s_i$ as is essentially correct for a simple mle. In other words, our approximations (5.8) and (5.9) crudely treat $\ln(s_i/t_i)$ as normal with a conditional mean that is Student t with non-negligible variance. Such approximations are very convenient, and lend themselves to simulation appraisals of the two-stage estimation procedure; see Gaver (1985), and a summary in Section 8 of this paper.

<u>Gamma</u>: In the gamma case, it can be seen that

$$Q(z) = -\alpha' e^{z} + \beta' z \tag{5.10}$$

where $\alpha' = t_i + \hat{\alpha}$, $\beta' = s_i + \hat{\beta}$. Now differentiation again shows that

$$e^{\hat{\varepsilon}_i} = (\beta'/\alpha') = \frac{s_i + \hat{\beta}}{t_i + \hat{\alpha}} \tag{5.11}$$

and

$$\sigma_i^2 = \frac{1}{\beta'} = \frac{1}{s_i + \hat{\beta}}. \tag{5.12}$$

Naturally, these formulas resemble the results for the log-Student superpopulation, but contain no weight, $\hat{w}_n$, to reduce shrinkage effects upon tail-discrepant observations.

6. Confidence Limits

Since the estimates of posterior means, variances, and tolerance limits are functions of $\hat{\mu}$ and $\hat{\tau}$, it is desirable to place confidence limits on $\bar{\varepsilon}_i(\mu,\tau)$, $\sigma_i^2(\mu,\tau)$, and $\tilde{\varepsilon}_i(\mu,\tau)$. These may be based on the confidence contours of Figs. 4.1-4.3, and are constructed numerically. We have supplied only upper 95% confidence contours, obtained by grid search over (μ,τ) space to maximize $\tilde{\varepsilon}_i(\mu,\tau)$, say, under the condition that (μ,τ) belongs to the appropriate confidence set; these limits are denoted by $\tilde{\tilde{\varepsilon}}_i$.

7. Analysis of Data Sets

The estimation procedures described have been applied to the data sets of Tables 1.1-1.3. Complete tabulations are available from the authors; here we examine only those log parameter estimates that are at the low and high extremes for each data set, and the middle or median level. Ordering of the rates is in terms of log raw rates. It is anticipated that the point estimates of the extreme individual rates will exhibit the greatest variation across estimation procedures (superpopulation models), while the middle values will be roughly in agreement. Owing to the partial pooling effect, the middle values will tend to exhibit somewhat smaller posterior variation than the extremes. The normal superpopulation model tends to shrink more extensively than the other superpopulation models. By and large, these effects are observed. For a visual notion of the posterior densities from our data sets see Figs. 7.1-7. .

7.1 Table Notation

The following notation has been used in the table headings: under Estimates,

(1) $\hat{\varepsilon}(r) = \ln(s_i/t_i) = \ln(r_i)$

(2) $\hat{\varepsilon}(1,n)$ = linearized posterior mode, Student (n) prior; (n=5) here

(3) $\bar{\varepsilon}(n)$ = posterior mean, Student (n) prior

(4) $\bar{\varepsilon}(g)$ = posterior mean, Gamma prior

(5) $\bar{\varepsilon}(\infty)$ = posterior mean, normal/Gauss prior.

The numbers in parentheses under each of the above are the standard deviations of the associated posteriors; posterior means and standard

deviations are either computed by numerical integration, in cases (3), (5), or by simple explicit approximate formulas in cases (1), (2), and (4). Under Intervals, there are included approximate upper 95% Bayes credibility intervals based on a normal approximation (mean + (1.645)(standard deviation)), these are

(6) $\tilde{\varepsilon}(r) = \hat{\varepsilon}(r) + 1.645\ \hat{\sigma}(r) = \hat{\varepsilon}_i(r) + 1.645\sqrt{1/s_i}$, using $1/s_i$, the simpliest delta-method approximation to $\text{var}[\ln(s_i/t_i)]$, while in [ ] we quote the upper limit computed making use of the chi-squared distribution of the time to accumulate s failures, an approximation to the former;

(7) $\tilde{\varepsilon}(1,n)$: same as preceding using Student (n) prior, (n=5), and linearized estimates, see (5.8) and (5.9);

(8) $\tilde{\varepsilon}(n)$: same as (6) but using (3), and associated standard deviation;

(9) $\tilde{\tilde{\varepsilon}}(n)$ upper 95% confidence limit on $\tilde{\varepsilon}(g)$, as described in Section 6;

(10) $\tilde{\varepsilon}(g)$: same as (6), using moments of log-gamma computed numerically;

(11) $\tilde{\tilde{\varepsilon}}(g)$: upper 95% confidence limit on $\tilde{\varepsilon}(g)$, as in Section 6;

(12) $\tilde{\varepsilon}(\infty)$: same as (8), using estimated normal prior;

(13) $\tilde{\tilde{\varepsilon}}(\infty)$: upper 95% confidence limit on $\tilde{\varepsilon}(g)$, as in Section 6.

7.2 Air-conditioner Data

Upward shrinkage of the smallest estimate, $\hat{\varepsilon}(r)$, is most pronounced for $\bar{\varepsilon}(\infty)$, the normal prior, less so for the gamma, and still less so for the Student (5)'s; the simple linear approximation least so. The linearized Student (5) procedure gives a small weight to the smallest observed rate. Upper interval boundaries differ less than point estimates, with the $\tilde{\varepsilon}$ levels only slightly above $\tilde{\varepsilon}$.

The middle estimate is shrunk not at all numerically by any of the point estimates, but the standard deviations of all shrunken/pooled estimates are about 70% of that of the raw estimate $\hat{\varepsilon}(r)$. Upper interval levels are correspondingly reduced.

The largest estimate is shrunk downwards slightly and consistently by all estimates, shrinkage is less extensive for the largest than the smallest; this can be partly explained by the weights: 0.52 vs 0.13.

7.3 Feedwater Flow Data

The smallest observation is a zero, and the crudely imputed rate is $\hat{\varepsilon}(r) = \ln(1/3t_i)$; it is enclosed in parentheses to signify its arbitrary nature. Here all point estimates provide some upward shrinkage: the normal prior estimate, $\bar{\varepsilon}(\infty)$, shrinking upwards the most extensively; it also exhibits the smallest standard deviation. Here the Student (5) credibility

and confidence intervals tend to be lower than the corresponding gamma and normal intervals.

The first middle estimate, (i)=15 in rank, is shrunk downwards by perhaps 10%; the most extensive shrinkage occurs for the normal model, $\bar{\varepsilon}(\infty)$. Its shrunken standard deviation is about 80% of that of the raw for the Student model. Note that this observation involves s=3 events over exposure time t=1, a short history. By contrast, its neighboring middle value (i)=16 in rank, with the more extensive history s=13 over t=4, exhibits one-half the shrinkage and very little standard deviation decrease.

The largest estimate is shrunken nearly the same by all estimates; the upper intervals agree well internally, tending to be slightly below the interval raw rate interval, $\hat{\varepsilon}(r)$.

7.4 Farley Pump Data

The smallest observation, $\hat{\varepsilon}(r)$ in this data set is shrunk towards the mean to essentially the same degree by each alternative point estimate; slightly less shrinkage occurs for the gamma and Student (5) models. The upper 95% credibility limits, $\tilde{\varepsilon}$, also agree for all models, with $\hat{\varepsilon}(1,5)$ being marginally the greatest. The upper confidence limits, $\underset{\sim}{\tilde{\varepsilon}}$, are in agreement as well.

The two median values at (i)=5,6 are individually treated consistently so far as point estimates go: all shrunken estimates reduce the log raw

rate towards the mean, with the greater shrinkage occurring for (i)=5 because of its smaller experience (failure count and exposure time). For the same reason, posterior standard deviations for (i)=5 are more than twice as great as those for (i)=6, and upper 95% credibility intervals and confidence limits reflect this difference as well.

The point estimates associated with the largest log raw rate all substantially agree in their modest downward shrinkage, and the upper credibility and confidence limits. Again, the close agreement is attributable to the relatively extensive experience embodied in unit (i)=10.

It is, however, worth notice that the estimated scale parameter, τ, is quite large for this data set. A plausible reason is that the units in the set are not truly homogeneous, and that much of the large variability is explainable by classification or regression. Our estimation procedures tend to reflect this: although weights $\hat{w}_n$ are rather similar for extreme and middle observations, the actual shrinkages are small even when there is little experience (e.g. for (i)=5). In fact, investigation reveals that the four pumps with the greatest experience (relatively large s and long t) all operate continuously, while the remaining six operate intermittently or on standby; these latter display consistently higher failure rates than the former, so a dummy variable (continuous vs intermittent) regression model should tend to reduce $\hat{\tau}$. In Fig. 7.4 we exhibit the result of a re-analysis in which the two groups' estimates of μ and τ are found separately: the two point estimate vectors are now much more consistent, the confidence regions are smaller, and only partially overlap.

Table 7.1
Air-Conditioner Data Analysis

Ranked Observations	Estimates $\hat{\xi}(r)$	$\hat{\xi}(1,5)$	$\bar{\xi}(5)$	$\bar{\xi}(g)$	$\bar{\xi}(n)$	Intervals $\tilde{\xi}(r)$	$\tilde{\xi}(1,5)$	$\tilde{\xi}(5)$	$\tilde{\tilde{\xi}}(5)$	$\tilde{\xi}(g)$	$\tilde{\tilde{\xi}}(g)$	$\tilde{\xi}(n)$	$\tilde{\tilde{\xi}}(n)$
Smallest (i=1)	1.17 (0.71)	1.94 (0.35) [0.13]	2.14 (0.24)	2.13 (0.22)	2.16 (0.20)	2.33 [2.03]	2.51	2.54	2.73	2.50	2.61	2.49	2.61
Median (i=7)	2.35 (0.21)	2.34 (0.13) [1.20]	2.34 (0.14)	2.34 (0.16)	2.34 (0.15)	2.70 [2.66]	2.56	2.58	2.67	2.59	2.67	2.59	2.82
Largest (i=13)	2.82 (0.18)	2.66 (0.16) [0.52]	2.60 (0.17)	2.60 (0.14)	2.61 (0.17)	3.12 [3.04]	2.92	2.88	3.05	2.84	3.03	2.86	3.06

Superpopulation Parameters

	$\hat{\mu}$		$\hat{\tau}$		
	MLE	MM	MLE	MM	
Stud (5)	2.35	-	0.19	-	
Normal	2.34	2.31	0.23	0.24	
Gamma	2.33	-	0.23	-	$(\hat{\alpha}=1.74, \hat{\beta}=18.41)$

Table 7.2
Loss of Feedwater Flow Data Analysis

Ranked Observations	Estimates $\hat{\varepsilon}(r)$	$\hat{\varepsilon}(1,5)$	$\bar{\varepsilon}(5)$	$\bar{\varepsilon}(g)$	$\bar{\varepsilon}(n)$	Intervals $\tilde{\varepsilon}(r)$	$\tilde{\varepsilon}(1,5)$	$\tilde{\varepsilon}(5)$	$\tilde{\tilde{\varepsilon}}(5)$	$\tilde{\varepsilon}(g)$	$\tilde{\tilde{\varepsilon}}(g)$	$\tilde{\varepsilon}(n)$	$\tilde{\tilde{\varepsilon}}(n)$
Smallest (i=1)	(-3.18) (0.57)	-1.74 (0.76) [0.17]	-2.20 (1.33)	-2.08 (0.96)	-1.30 (0.53)	-2.24 [-]	-0.49	-0.01	0.10	-0.51	0.03	-0.42	0.03
Smallest Non-zero (i=3)	-1.32 (0.50)	-0.96 (0.39)	-1.05 (0.46)	-1.13 (0.45)	-0.97 (0.38)	-0.50 [-0.66]	-0.50	-0.32	-0.25	-0.39	-0.09	-0.35	-0.09
Medians (i=15)	1.10 (0.57)	1.02 (0.45) [1.18]	0.95 (0.47)	0.98 (0.50)	0.91 (0.51)	2.04 [1.58]	1.76	1.73	1.87	1.80	1.82	1.75	1.88
(i=16)	1.18 (0.28)	1.14 (0.26) [1.17]	1.11 (0.27)	1.13 (0.27)	1.11 (0.27)	1.641 [1.58]	1.64	1.55	1.59	1.57	1.58	1.56	1.59
Largest (i=30)	1.95 (0.17)	1.90 (0.17) [0.86]	1.89 (0.17)	1.88 (0.17)	1.89 (0.17)	2.23 [2.20]	2.23	2.17	2.19	2.15	2.18	2.17	2.20

Superpopulation Parameters

	$\hat{\mu}$ MLE	$\hat{\mu}$ MM	$\hat{\tau}$ MLE	$\hat{\tau}$ MM	
Stud (5)	0.76	-	0.91	-	
Normal	0.93	0.94	0.72	0.57	
Gamma	2.33	-	0.23	-	$(\hat{\alpha}= .52, \hat{\beta}=1.53)$

Table 7.3
Pump Data Analysis

Ranked Observations	Estimates $\hat{\varepsilon}(r)$	$\hat{\varepsilon}(1,5)$	$\bar{\varepsilon}(5)$	$\bar{\varepsilon}(g)$	$\bar{\varepsilon}(n)$	Intervals $\tilde{\varepsilon}(r)$	$\tilde{\varepsilon}(1,5)$	$\tilde{\varepsilon}(5)$	$\tilde{\tilde{\varepsilon}}(5)$	$\tilde{\varepsilon}(g)$	$\tilde{\tilde{\varepsilon}}(g)$	$\tilde{\varepsilon}(n)$	$\tilde{\tilde{\varepsilon}}(n)$
Smallest (i=1)	-2.94 (0.45)	-2.75 (0.39) [0.84]	-2.84 0.42	-2.89 0.43	-2.83 0.40)	-2.20 [-2.33]	-2.10	-2.16	-2.04	-2.17	-2.08	-2.17	-2.09
Medians (i=5)	-0.56 (0.58)	-0.69 (0.54) [1.14]	-0.81 (0.56)	-0.67 (0.55)	-0.79 (0.57)	0.39 [0.18]	0.20	0.12	0.29	0.23	0.44	0.15	0.29
(i=6)	-0.50 (0.23)	-0.53 (0.23) [1.13]	-0.56 (0.23)	-0.55 (0.23)	-0.57 (0.23)	-0.12 [-0.16]	-0.15	-0.18	-0.14	-0.15	-0.10	-0.17	-0.15
Largest (i=10)	0.74 (0.21)	0.69 (0.22) [0.79]	0.67 (0.22)	0.64 (0.21)	0.67 (0.22)	1.08 [1.06]	1.05	1.03	1.06	0.99	1.08	1.03	1.08

Superpopulation Parameters

	$\hat{\mu}$ MLE	$\hat{\mu}$ MM	$\hat{\tau}$ MLE	$\hat{\tau}$ MM	
Stud (5)	-1.18	-	1.29	-	
Normal	-1.19	-0.55	1.19	0.71	
Gamma	-0.83	-	0.89	-	$(\hat{\alpha}=1.27, \hat{\beta}=0.82)$

8. Simulation Results

Limited simulation experiments have been carried out to evaluate some of the estimation procedures described. Here is the design; see Gaver (1985) for further details.

First, the superpopulation form for the Poisson log rates ε_i was taken for convenience to be a member of the controllably long-tailed Tukey h family:

$\underset{\sim}{\varepsilon}_i \sim \mu + \tau \underset{\sim}{z}_i \exp(h\underset{\sim}{z}_i^2)$ with $\underset{\sim}{z}_i \sim N(0,1)$ and h, the tail-stretching parameter, non-negative (here h=0.15); see Hoaglin (1983), and Gaver (1983) for details concerning this family. We wished to compare the treatment of the different rate-values in a random sample from the superpopulation by various estimators, so ordered λ-values were next created (and stored): $\lambda_{(i)} = \exp(\varepsilon_{(i)})$, $\varepsilon_{(i)} = \mu + \tau z_{(i)} \exp[h\, z_{(i)}^2]$, $z_{(i)}$ being the i^{th} largest order statistic in a sample of size I from N(0,1). For $h > 0$ the extremes $\lambda_{(1)}$ and $\lambda_{(I)}$ tend to be outliers, while the median, $\lambda_{\left(\frac{I+1}{2}\right)}$, is characteristic of a central value. Second, the $\lambda_{(i)}$-values were used to generate Poisson counts, $s_{(i)}$. Then Stage 1 and Stage 2 estimation processes were applied to estimate first μ, τ, and then the individual λ_i-values. The speedy LGH procedure was used to estimate μ and τ, and these values were then used in conjunction with the approximation $\lambda_i = \exp(\varepsilon_i)$ that solves (5.6) by iteration. Detailed numerical quadrature using the GH procedure is perhaps superior, but would have consumed more computer time. The squared differences of the estimated $\underset{\sim}{\lambda}_i$ values and their true

counterparts were then averaged over S(=200) simulations and quoted as mean-squared errors (MSE). Table 8.1 summarizes results for several such experiments. We have quoted the ordinary units estimate results as MLE, the results of applying the estimating formulas (5.6) as RS (Restricted Shrinkage, as governed by $\hat{w}_n$), and the results of applying (5.6) without the weight as SS (Simple Shrinkage); the latter approximately represents the effect of applying a log-Student model when n=50.

The results obtained are suggestive if not dramatic. First, estimates of the superpopulation mean μ are nearly unbiased, while those for τ^2 appear biased low. Standard errors of estimates (figures in parentheses) are, not surprisingly, substantial; apparently more simulation repetitions would be desirable. Nevertheless, comparison of the MSE figures for the various estimators implies that RS(n=4) has virtue: for the smallest and largest rates, $\lambda_{(1)}$ and $\lambda_{(15)}$, RS estimates resemble MLE performance, while SS over-shrinks and for the middle value, $\lambda_{(8)}$, RS is far superior to the simple individual, unpooled MLE. The numerical differences in MSE shown in the table are small but real because of positive correlation between estimate values on each simulation experiment.

Table 8.1

Selected Mean Squared Error Comparisons
and Estimated Superpopulation Parameters

J=15, h=0.15, 200 Simulations
Student d.f. (tuning constant) n=4,50

True Values	Estimated	Estimator	$\lambda_{(1)}$ (small)	$\lambda_{(8)}$ (median)	$\lambda_{(15)}$ (large)
	(n=4):				
	$\hat{\mu}$=-0.97(0.41)	RS	0.016	0.019	0.33
	$\hat{\tau}^2$=0.17(0.15)				
μ=-1.0					
τ^2=0.25		MLE	0.007	0.030	0.32
	(n=50):				
	$\hat{\mu}$=-0.98(0.45)				
	$\hat{\tau}^2$=0.18(0.15)	SS	0.019	0.020	0.35
	(n=4):				
	$\hat{\mu}$=-1.93(0.50)	RS	0.050	0.0060	0.28
	$\hat{\tau}^2$=0.18(0.17)				
μ=-2.0					
τ^2=0.25		MLE	0.0026	0.014	0.27
	(n=50):				
	$\hat{\mu}$=-1.93(0.52)				
	$\hat{\tau}^2$=0.20(0.18)	SS	0.0054	0.0057	0.30

9. Summary and Conclusions

This paper displays the results of analyzing several small batches (optimistically, but not realistically, random samples) of event rate data as if (i) parameters of each batch were drawn independently from a fixed superpopulation, and then (ii) the batches themselves were samples from random processes, here stationary Poisson or exponential-interval. Such is at least a pleasant fiction, to be used as a starting point. Computational methods have been used to obtain estimates of superpopulation parameters, and, from these pooled or shrunken individualized (log) rate estimates were obtained. Such parametric empirical Bayes (PEB) analyses have been described before by Hill, et al (1984), Deely and Lindley (1981), Hinde (1982), Kaplan (1983) and perhaps others. We extend these by introducing a heavy-tailed superpopulation form, the log-Student t, that allows for outliers or tail discrepancies incompatible with the log-normal/Gauss description. We call this the RPEB setup. The qualitative effect of such a generalization is revealed by appearance of the weight, $\hat{w}_n$, that selectively reduces the linear shrinkage towards a center; see (5.8). Thus $\hat{w}_n$ plays a role similar to that of an influence function in robust location estimation (see Mosteller and Tukey (1971), p. 351), although in the estimation of a single log rate it curbs the influence of the overall mean, $\hat{\mu}$, on that estimate if the data give evidence of extreme discrepancy. A more complete indication of the effect of an observation, i.e. $\ln(s_i/t_i) = \hat{\varepsilon}_i(r)$, on its own shrinkage is given by the quantity $(1/(\hat{\tau})^2)\hat{w}_n/[s_i + (1/(\hat{\tau})^2)\hat{w}_n]$ appearing in the rightmost expression in (5.8): both within variability, measured by $s_i (\approx \mathrm{var}[\hat{\varepsilon}_i(r)])$ and between variability, assessed by $(\hat{\tau})^2$, play their parts along with $\hat{w}_n$. Besides providing insight, expressions like (5.8) and (5.9) seem to agree reasonably well with more exact results, and

are easy to compute, especially if one settles for inefficient moment estimators of superpopulation parameters.

As the Tables and Figures reveal, the example data analyses performed do not show enormous differences between log-normal, log-Student(5), and gamma superpopulation (Bayes prior) specifications, especially for the median and also for the largest batch values. The smallest batch observations are treated similarly by gamma and Student(5), with the normal/Gauss representation tending to shrink a "small" (zero) value more extensively than do the others up towards the center, μ, on the log scale. As anticipated, other analyses indicate even less tail shrinkage by Student(n) for $n<5$. Estimation of n from the batch values would be of interest, but is unlikely to be done with much accuracy from scanty data. This suggests that use of a gamma form for the prior, and hence for the posterior may be relatively harmless. There is little evidence in our examples that over-shrinkage of the largest values in a data set occurs when the gamma specification is used (although a small-n Student analysis could be performed as an indication of the possible extent of over-shrinkage). Certainly the gamma is technically convenient for computing point estimates of reliabilities or availabilities of complex systems: integrations can often be carried out explicitly as Laplace transforms.

Of considerable interest would be the reduction of the apparent between variability by classification or regression, as briefly illustrated for the Farley data. Research in this area is currently in progress, with promising results. If part of the between variability could be suitably accounted for, then estimators could be constructed that legitimately pool towards

appropriate individualized centers, μ_i rather than μ, and outliers could be explained and reduced in effect. All of the above requires attention to collection of representative current data, and the monitoring of analytical results over time to check for changes, e.g. in basic parameters. Our present analysis is only a step in this direction. Further generalizations include analyses for failure-on-demand data, for which responses are binary and explanatory variables could include the time durations between inspections or serious activations. Analyses of complex redundant systems have been proposed in Gaver and Lehoczky (1985).

REFERENCES

Albert, J. H. (1979). "Robust Bayes Estimation". Purdue University Department of Statistics Mimeo. Series #79-9.

Berger, J. (1980). Statistical Decision Theory. Springer-Verlag.

Berger, J. O. (1984). "The Robust Bayesian Viewpoint"; in Robustness of Bayesian Analyses, J. B. Kadane (ed.), Elsevier.

Box, G. E. P. (1980). "Sampling and Bayes' inference in scientific modelling and robustness". Journal of the Royal Statistical Soc., 143, A, pp. 124-130.

Burridge, J. (1981). "Empirical Bayes Analysis of Surviaal Time Data". J. R. Statist. Soc. B, 43, No. 1, pp. 65-76.

Copas, J. B. (1984). "A Computer-Intensive Non-Parametric Estimate of a Mixing Distribution, with Application to Poisson Failure Data." Monterey: Naval Postgraduate School. Technical Report.

Cox, D. R. and Snell, E. J. (1980) Applied Statistics. Chapman and Hall. London.

Cox, D. R. and Isham, V. (1980). Point Processes. Chapman and Hall. London.

Cox, D. R. and Lewis, P. A. W. (1966). The Statistical Analysis of Series of Events. London: Methuen.

deBruin, N. G. (1957) Asymptotic Methods in Analysis.

Deely, J. J. and Lindley, D. V. (1981). "Bayes Empirical Bayes". Journal of the American Statistical Association. Vol 1, 76, No. 376, pp. 833-841.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Soc., B, Vol 39, pp. 1-38.

Efron, B. and Morris, C. (1973). "Stein's Estimation Rule and Its Competitors--An Empirical Bayes Approach". Journal of the American Statistical Association. Vol. 68, No. 341, pp.117-130.

Gaver, D. P. (1985). "Discrepancy Tolerant Hierarchical Poisson Event Rate Analyses". Tech. Report, Naval Postgraduate School, Monterey, California.

Gaver, D. and Lehoczky, J. (1985). "Random parameter Markov Population Process Models and Their Likelihood, Bayes, and Empirical Bayes Analysis," Naval Postgraduate School Tech. Report; to appear in the Herbert Solomon Anniversary Volume (A. Gelfand, ed.).

Gaver, D. P. (1963). "Random hazard in reliability problems". Technometrics, 5, pp. 211-226.

Hill, J. R., Heger, A. S. and Koen, B. V. (1984). "The Application of Stein and Related Parametric Empirical Bayes Estimators to the Nuclear Plant Reliability Data System". Austin, Texas: The Univ. of Texas. NUREG/CR-3637.

Hinde, J. (1982). "Compound Poisson Regression Models". In GLIM 82; Proc. of the International Conference in Generalized Linear Models (R. Gilchnist, ed.) pp. 109-121. Springer-V., Berlin.

Hoaglin, G. C. (1983). "g and h distributions". Encyclopedia of Statistics, Vol. 3, No.

Johnstone, I. (1984). "Admissability Difference Equations and Recurrence in Estimating a Poisson Mean." Annals of Statistics, Vol. 12, pp. 1173-1198.

Laird, N. (1978). "Non-parametric Maximum Likelihood Estimation of a Mixing Distribution." Journal of the American Statistical Assoc. Vol. 73, pp. 805-811.

Kaplan, S. (1983). "On a two-stage Bayesian procedure for determining failure rates from experiential data". IEEE Transactions on Power Apparatus and Systems, Vol. PAS-102, No. 1, pp. 195-202.

Martz, H. F. and Waller, R. A. (1982). Bayesian Reliability Analysis. John Wiley.

Morris, C. (1982). "Natural exponential families with quadratic variance functions: statistical theory." Annuals of Statistics, Vol II, No. 2, pp. 515-529.

Morris, C. (1983). "Parametric empirical Bayes inference: theory and applications." Journal of the American Statistical Association. Vol. 78, No. 381 (with discussion), pp. 47-65.

Mosteller, F., and Tukey, J. W. (1977). Data Analysis and Regression. Addison-Wesley: Reading, Mass.

Mosteller, F. and Wallace, D. L. (1964). Inference and Disputed Authorship: The Federalist. Addison-Wesley: Reading, Mass.

Naylor, J. C. and Smith, A. F. M. (1982). "Applications of a method for the efficient computation of posterior distributions". Applied Statistics, 31, No. 373, pp. 214-225.

Rasmussen, N. et al (1975). "Reactor Safety Study: An Assessment of Accident Risks in U. S. Commercial Nuclear Power Plants." NUREG-75/014, WASH 1400.

Reynolds, D. S., and Savage, I. R. (1971). "Random wear models in reliability theory." Adv. Appl. Prob., 3, pp. 229-248.

Robbins, H. (1983). "Some thoughts on empirical Bayes estimation." Annals of Statistics. Vol. 11, No. 3, pp. 713-723.

Tierney, L. and Kadane, J. B. (1984). "Accurate approximations for posterior moments and marginals." Tech. Reports, Department of Statistics, University of Minnesota, and Department of Statistics and Social Science, Carnegie-Mellon, Univ.

Tukey, J. W. (1974). "Named and faceless values: an initial exploration in memory of Prasanta C. Mahalanobis." Sankhya, Series A, 36, pp. 125-176.

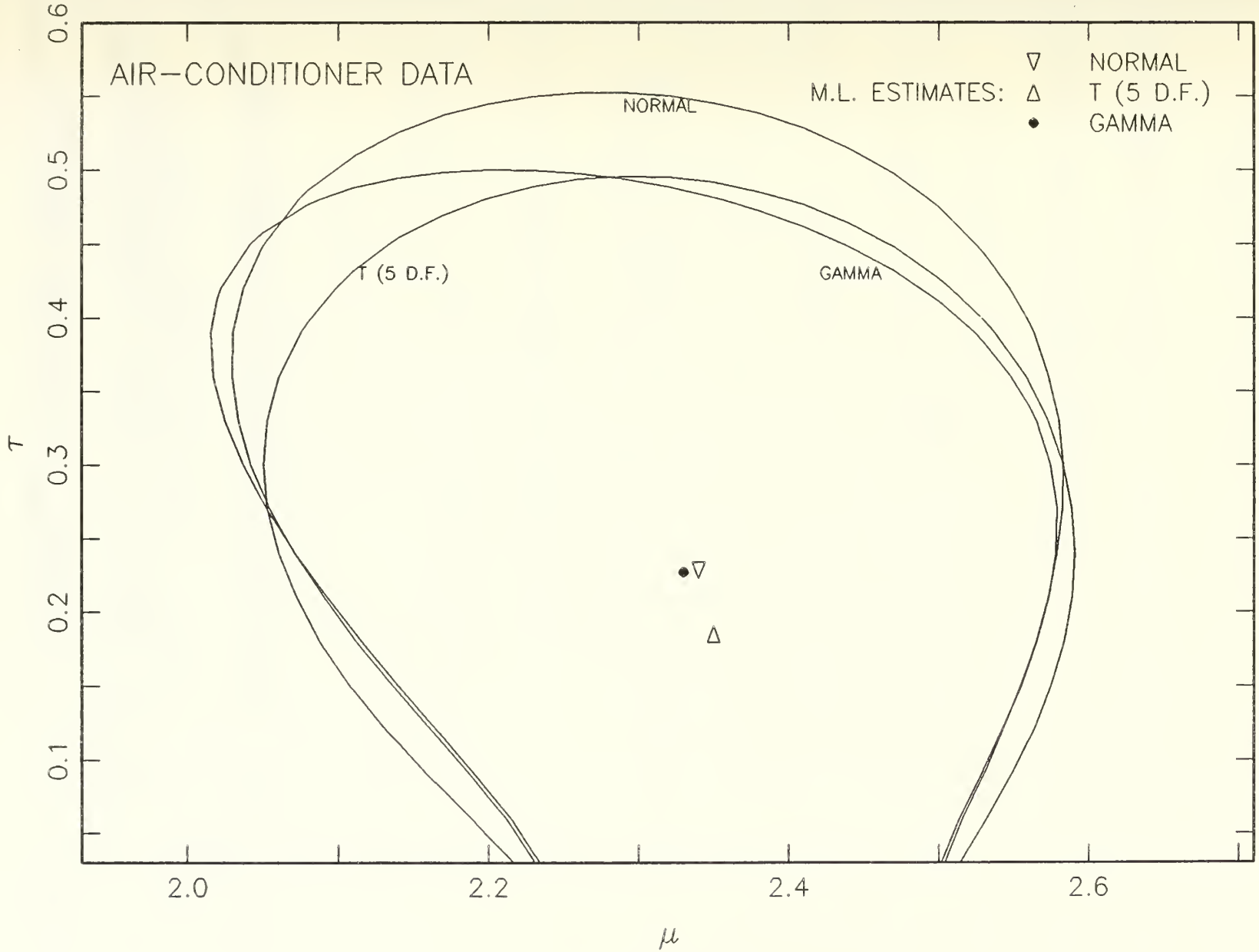

95% CONFIDENCE REGIONS FOR μ,τ
AIR–CONDITIONER DATA
M.L. ESTIMATES:
NORMAL
T (5 D.F.)
GAMMA
NORMAL
T (5 D.F.)
GAMMA
τ
μ
0.1
0.2
0.3
0.4
0.5
0.6
2.0
2.2
2.4
2.6

Figure 4.1

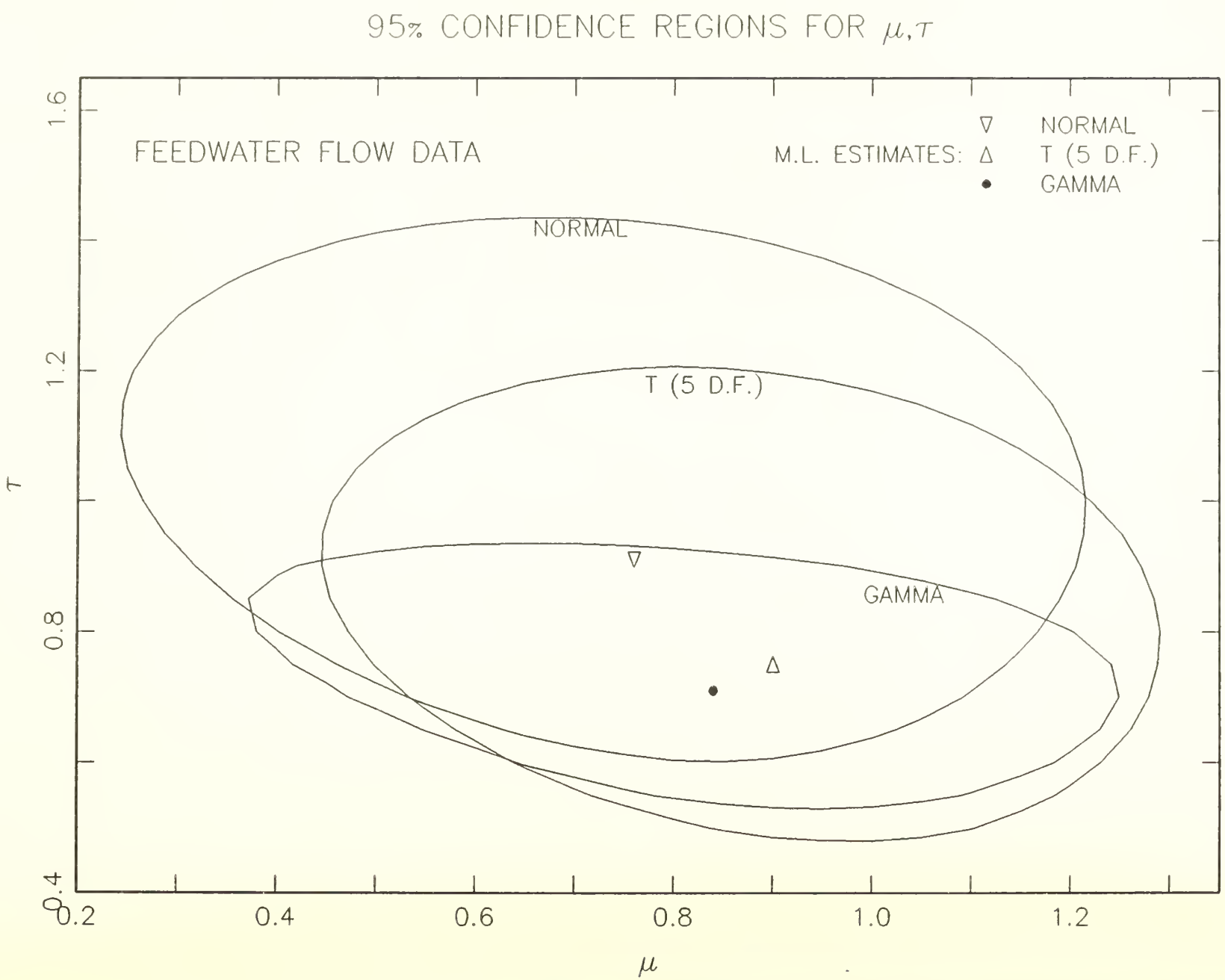

95% CONFIDENCE REGIONS FOR μ,τ
FEEDWATER FLOW DATA
M.L. ESTIMATES:
NORMAL
T (5 D.F.)
GAMMA
NORMAL
T (5 D.F.)
GAMMA
τ
μ
0.4
0.8
1.2
1.6
0.2
0.4
0.6
0.8
1.0
1.2

Figure 4.2

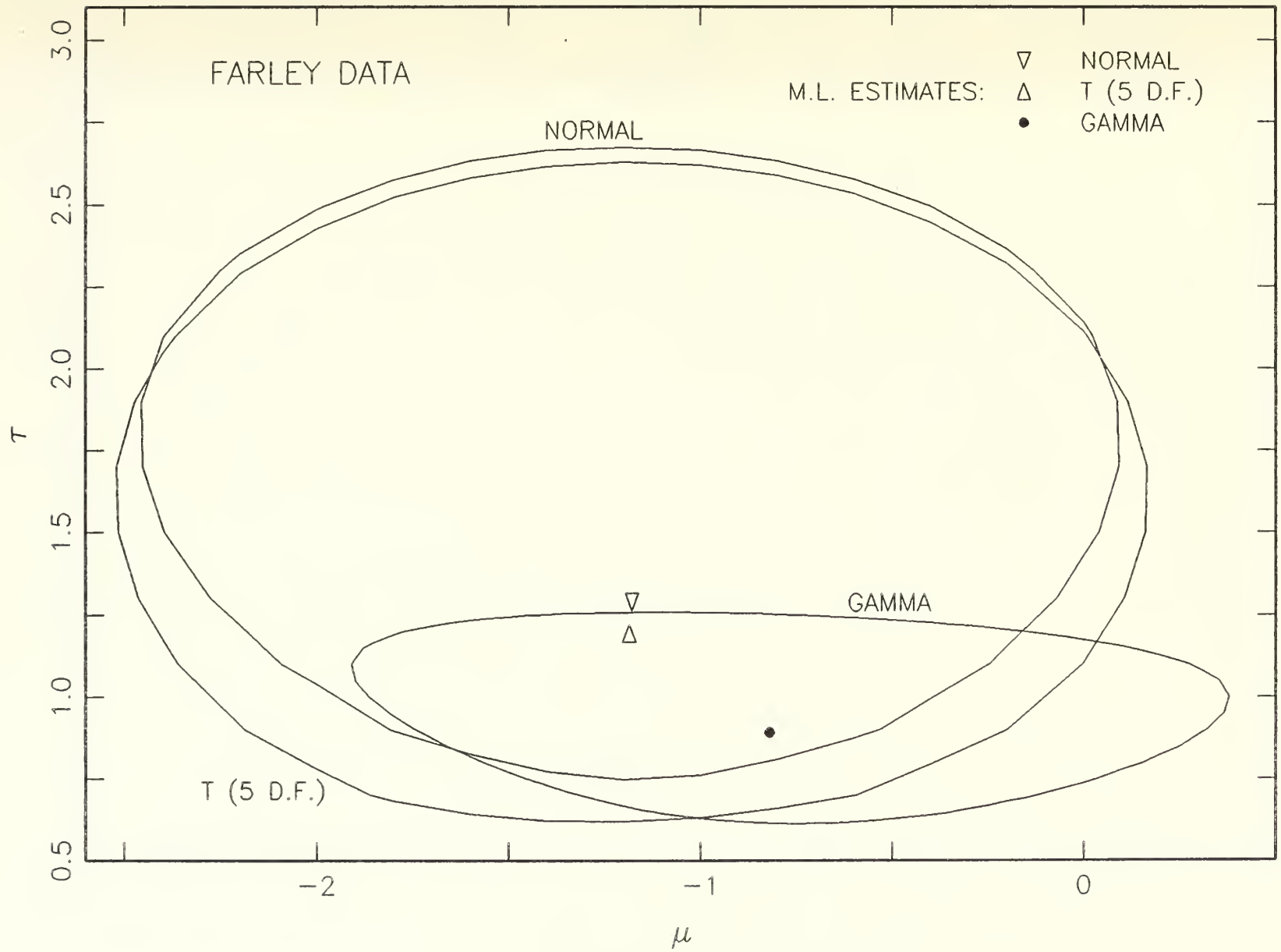
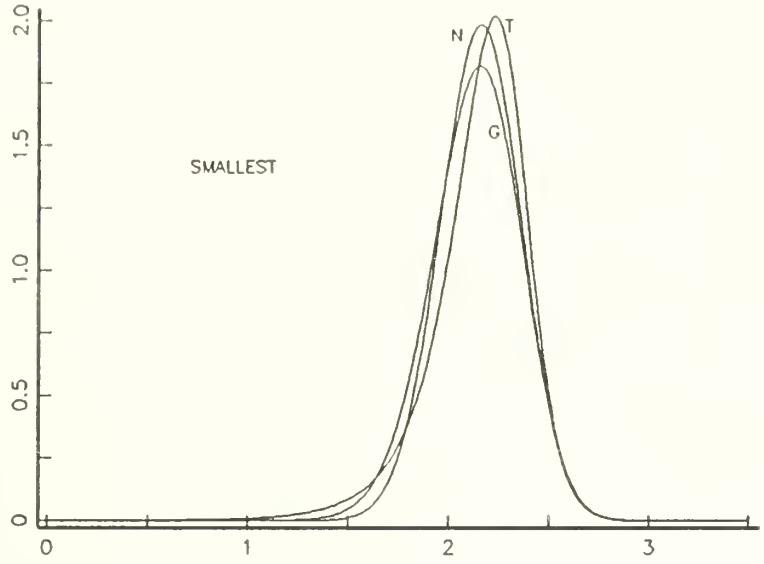
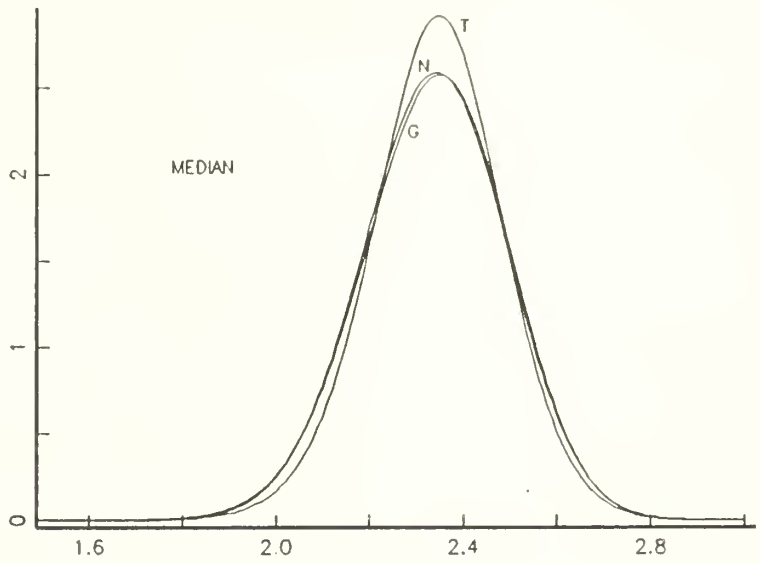

95% CONFIDENCE REGIONS FOR μ,τ
FARLEY DATA
M.L. ESTIMATES:
NORMAL
T (5 D.F.)
GAMMA
NORMAL
GAMMA
T (5 D.F.)
τ
μ
3.0
2.5
2.0
1.5
1.0
0.5
−2
−1
0

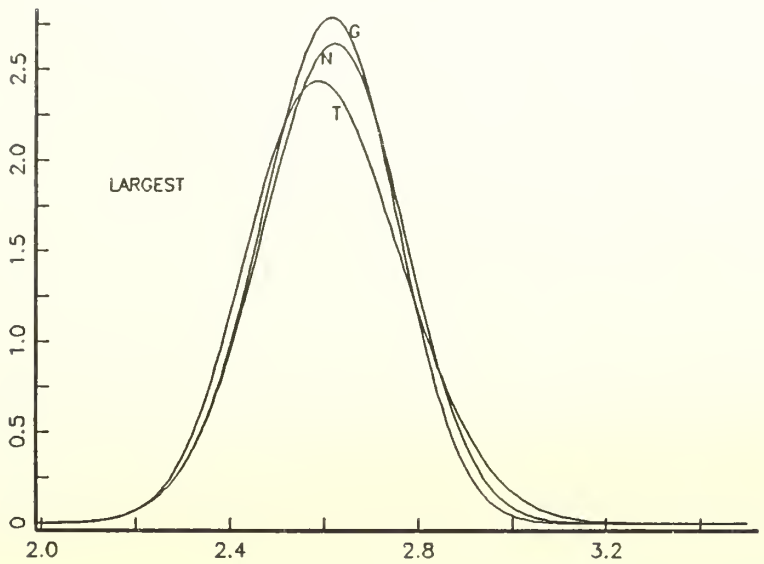
Figure 4.3

POSTERIOR DENSITIES

AIR-CONDITIONER DATA







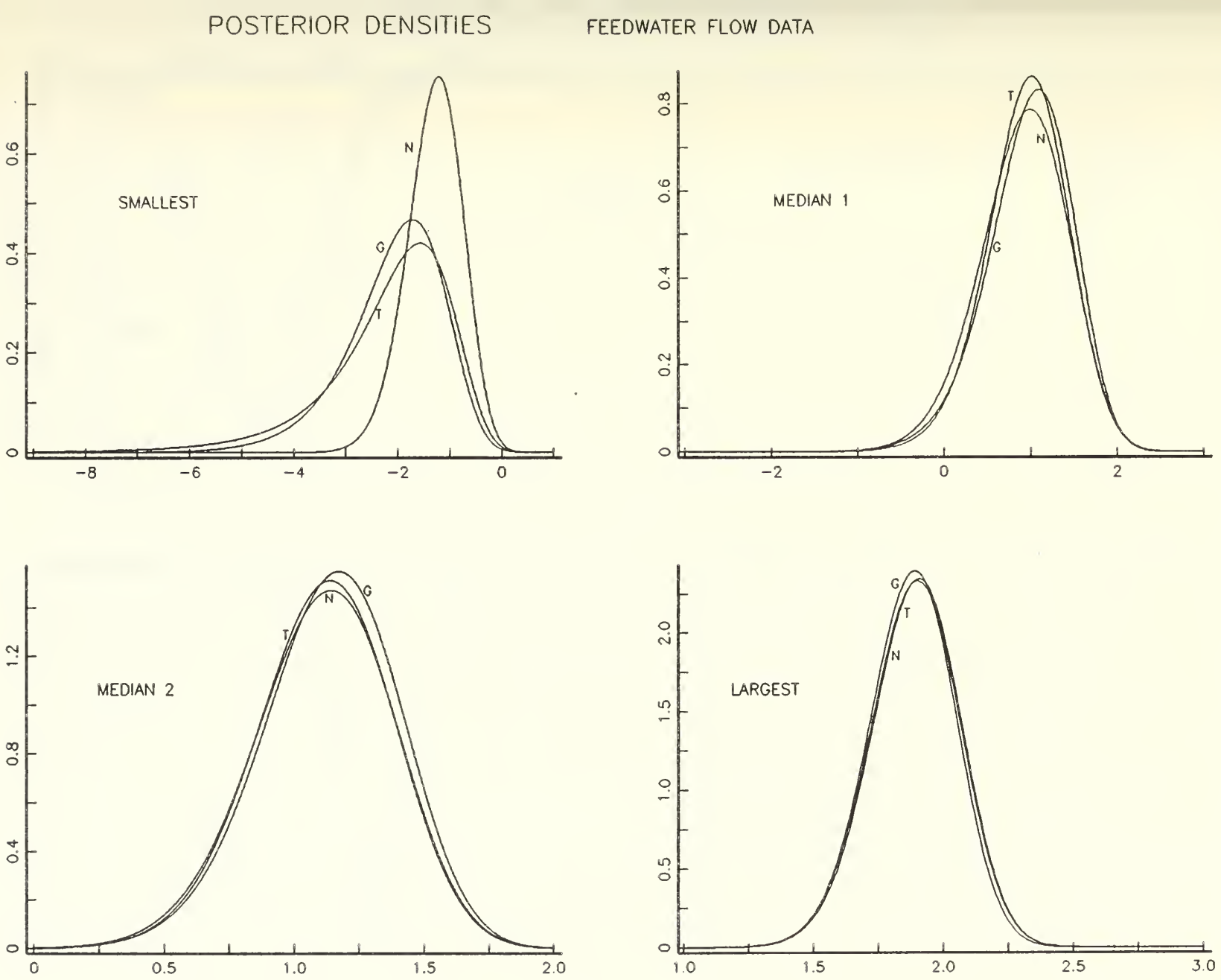
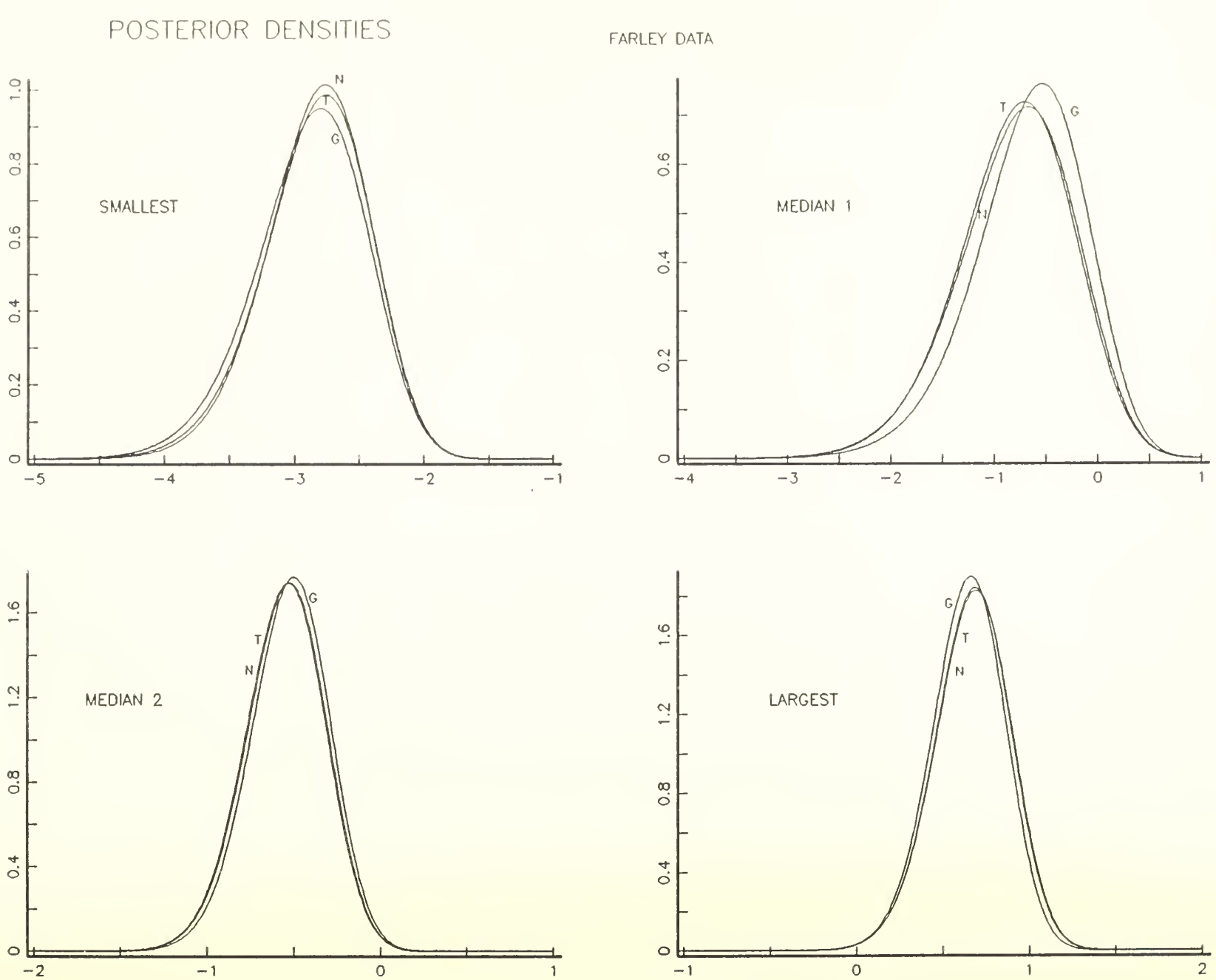

POSTERIOR DENSITIES
FEEDWATER FLOW DATA
SMALLEST
MEDIAN 1
MEDIAN 2
LARGEST
N
G
T

Figure 7.2


POSTERIOR DENSITIES
FARLEY DATA
SMALLEST
MEDIAN 1
MEDIAN 2
LARGEST
N
T
G

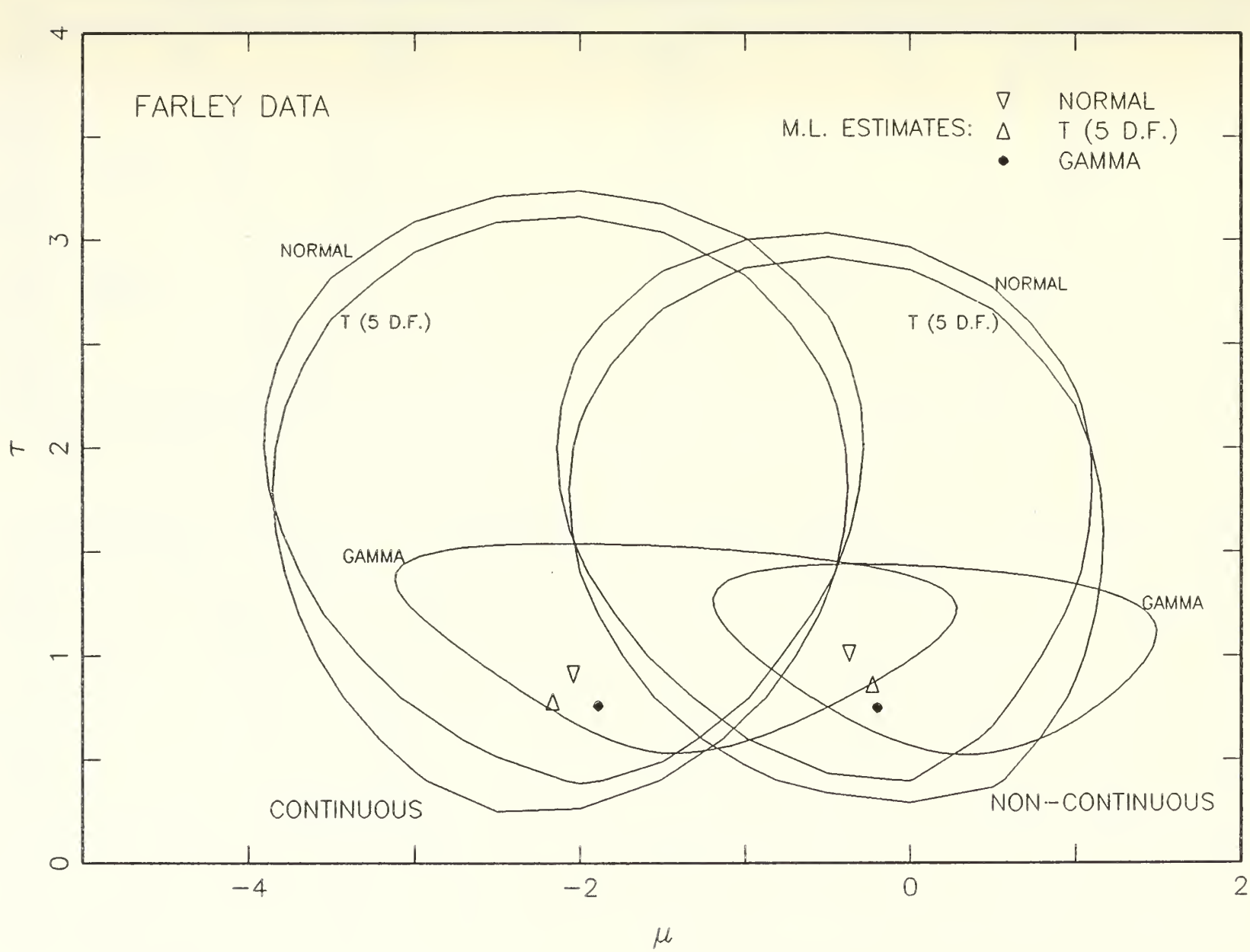

95% CONFIDENCE REGIONS FOR μ,τ
FARLEY DATA
M.L. ESTIMATES:
NORMAL
T (5 D.F.)
GAMMA
NORMAL
T (5 D.F.)
NORMAL
T (5 D.F.)
GAMMA
GAMMA
CONTINUOUS
NON-CONTINUOUS
τ
μ
0
1
2
3
4
−4
−2
0
2

Figure 7.4

Acknowledgements

The writers wish to acknowledge support and encourgement by Dr. David H. Worledge, Program Manager, Nuclear Safety and Analysis Department, Electric Power Research Institute (EPRI), during the process of writing this paper. They also gratefully acknowledge the research support of the office of Naval Research, Probability and Statistics Branch. Finally, Computations and figures in the paper were obtained using the excellent IBM developmental product GRAFSTAT, installed at the Naval Postgraduate School.

DISTRIBUTION LIST

	NO. OF COPIES
Dr. D. F. Daley Statistics Dept. (I.A.S) Australian National University Canberra A.C.T. 2606 AUSTRALIA	1
Prof. F. J. Anscombe Department of Statistics Yale University, Box 2179 New Haven, CT 06520	1
Dr. David Brillinger Statistics Department University of California Berkeley, CA 94720	1
Dr. R. Gnanadesikan Bell Core 435 South Street Morris Township, NJ 07960	1
Prof. Bernard Harris Dept. of Statistics University of Wisconsin 610 Walnut Street Madison, WI 53706	1
Dr. D. R. Cox Department of Mathematics Imperial College London SW7 ENGLAND	1
Dr. A. J. Lawrance Dept. of Mathematical Statistics University of Birmingham P. O. Box 363 Birmingham B15 2TT ENGLAND	1
Professor W. M. Hinich University of Texas Austin, TX 78712	1

Dr. John Copas
Dept. of Mathematical Statistics
University of Birmingham
P. O. Box 363
Birmingham B15 2TT
ENGLAND 1

P. Heidelberger
IBM Research Laboratory
Yorktown Heights
New York, NY 10598 1

Prof. M. Leadbetter
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514 1

Dr. D. L. Iglehart
Dept. of Operations Research
Stanford University
Stanford, CA 94350 1

Dr. D. Vere-Jones
Dept. of Mathematics
Victoria University of Wellington
P. O. Box 196
Wellington
NEW ZEALAND 1

Prof. J. B. Kadane
Dept. of Statistics
Carnegie-Mellon
Pittsburgh, PA 15213 1

Prof. J. Lehoczky
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213 1

Dr. J. Maar (R51)
National Security Agency
Fort Meade, MD 20755 1

Dr. M. Mazumdar
Dept. of Industrial Engineering
University of Pittsburgh
Pittsburgh, PA 15235 1

Prof. M. Rosenblatt 1
Department of Mathematics
University of California-San Diego
La Jolla, CA 92093

Prof. Rupert G. Miller, Jr. 1
Statistics Department
Sequoia Hall
Stanford University
Stanford, CA 94305

Prof. I. R. Savage 1
Dept. of Statistics
Yale University
New Haven, CT 06520

Dr. Paul Shaman 1
National Science Foundation
Mathematical Sciences Section
1800 G. Street, NW
Washington, D. C. 20550

Prof. W. R. Schucany 1
Dept. of Statistics
Southern Methodist University
Dallas, TX 75222

Prof. D. C. Siegmund 1
Dept. of Statistics
Sequoia Hall
Stanford University
Stanford, CA 94305

Prof. H. Solomon 1
Department of Statistics
Sequoia Hall
Stanford University
Stanford, CA 94305

Dr. Ed Wegman 1
Statistics & Probability Program
Code 411(SP)
Office of Naval Research
Arlington, VA 22217

Dr. P. Welch 1
IBM Research Laboratory
Yorktown Heights, NY 10598

Dr. Marvin Moss 1
Office of Naval Research
Arlington, VA 22217

Dr. Roy Welsch
Sloan School
M. I. T.
Cambridge, MA 02139 1

Pat Welsh
Head, Polar Oceanography Branch
Code 332
Naval Ocean Research & Dev. Activity
NSTL Station, MS 39529 1

Dr. Douglas de Priest
Statistics & Probability Program
Code 411(SP)
Office of Naval Research
Arlington, VA 22217 1

Dr. Morris DeGroot
Statistics Department
Carnegie-Mellon University
Pittsburgh, PA 15235 1

Prof. J. R. Thompson
Dept. of Mathematical Science
Rice University
Houston, TX 77001 1

Prof. J. W. Tukey
Statistics Department
Princeton University
Princeton, NJ 08540 1

Dr. Daniel H. Wagner
Station Square One
Paoli, PA 19301 1

Dr. Colin Mallows
Bell Telephone Laboratories
Murray Hill, NJ 07974 1

Dr. D. Pregibon
Bell Telephone Laboratories - AT&T
Murray Hill, NJ 07974 1

Dr. Jon Kettenring
Bell Core
435 South Street
Morris Township, NJ 07960 1

Dr. David L. Wallace
Statistics Dept.
University of Chicago
5734 S. University Ave.
Chicago, IL 60637 1

Dr. F. Mosteller 1
Dept. of Statistics
Harvard University
Cambridge, MA 02138

Dr. S. R. Dalal 1
Bell Laboratories - AT&T
Mountain Avenue
Murray Hill, NJ 07974

Prof. Donald P. Gaver 20
Code 55Gv
Naval Postgraduate School
Monterey, California 93943-5000

Prof. Patricia Jacobs 1
Code 55Jc
Naval Postgraduate School
Monterey, California 93943-5000

Dr. Guy Fayolle 1
I.N.R.I.A.
Dom de Voluceau-Rocquencourt
78150 Le Chesnay Cedex
FRANCE

Dr. M. J. Fischer 1
Defense Communications Agency
1860 Wiehle Avenue
Reston, VA 22070

Prof. George S. Fishman 1
Curr. in OR & Systems Analysis
University of North Carolina
Chapel Hill, NC 20742

Prof. Guy Latouche 1
University Libre Bruxelles
C. P. 212
Blvd De Triomphe
B-1050 Bruxelles
BELGIUM

Library 4
Code 1424
Naval Postgraduate School
Monterey, CA 93943-5000

Dr. Alan F. Petty 1
Code 7930
Naval Research Laboratory
Washington, DC 20375

Prof. Bradley Efron 1
Statistics Department
Sequoia Hall
Stanford University
Stanford, CA 94305

Prof. Carl N. Morris 1
Dept. of Mathematics
University of Texas
Austin, TX 78712

Dr. John E. Rolph 1
RAND Corporation
1700 Main Street
Santa Monica, CA 90406

Prof. Linda V. Green 1
Graduate School of Business
Columbia University
New York, NY 10027

Dr. David Burman 1
Bell Laboratories - AT&T
Mountain Avenue
Murray Hill, NJ 07974

Dr. Ed Coffman 1
Bell Laboratories - AT&T
Mountain Avenue
Murray Hill, NJ 07974

Prof. William Jewell 1
Operations Research Department
University of California, Berkeley
Berkeley, CA 94720

Dr. Tom A. Louis 1
Biostatics Department
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

Dr. Nan Laird 1
Biostatics Department
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

Dr. Marvin Zelen 1
Biostatics Department
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

Dr. John Orav
Biostatics Department
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115 1

Prof. R. Douglas Martin
Chairman
Department of Statistics, GN-22
University of Washington
Seattle, WA 98195 1

Prof. W. Stuetzle
Department of Statistics
University of Washington
Seattle, WA 98195 1

Prof. F. W. Mosteller
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, MA 02138 1

Dr. D. C. Hoaglin
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, MA 02138 1

Prof. N. D. Singpurwalla
George Washington University
Washington, D. C. 20052 1

Center for Naval Analyses
2000 Beauregard Street
Alexandria, VA 22311 1

Prof. H. Chernoff
Department of Mathematics
M. I. T.
Cambridge, MA 02139 1

Dr. T. J. Ott
Bell Core
435 South Street
Morris Township, NJ 07960 1

Alan Weiss
AT&T Bell Laboratories
Murray Hill, NJ 1

Operations Research Center, Room E40-164 1
Massachusetts Institute of Technology
Attn: R. C. Larson and J. F. Shapiro
Cambridge, MA 02139

Office of Research Administration (Code 012) 1
Naval Postgraduate School
Monterey, CA 93943-5000