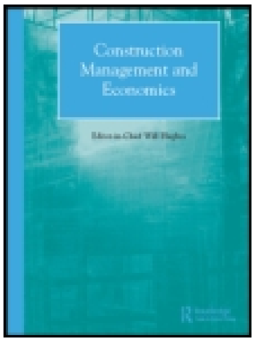

Construction Management and Economics

ISSN: 0144-6193 (Print) 1466-433X (Online) Journal homepage: www.tandfonline.com/journals/rcme20

The construction contract bidder homogeneity assumption: An empirical test

Martin Skitmore

Construction Management and Economics, 1991, **9**, 403–429

The construction contract bidder homogeneity assumption: An empirical test

MARTIN SKITMORE
Department of Civil Engineering, University of Salford, Salford M5 4WT, UK

This paper describes an empirical study to test the proposition that all construction contract bidders are homogeneous, i.e. they can be treated as behaving collectively in an identical (statistical) manner. An examination of previous analyses of bidding data reveals a flaw in the method of standardizing bids across different size contracts and a new procedure is proposed which involves the estimation of a contract datum.

Three independent sets of bidding data were then subjected to this procedure and estimates of the necessary distributional parameters obtained. These were then tested against the bidder homogeneity assumption, resulting in the conclusion that the assumption may be appropriate for a three-parameter log-normal shape, but not for scale and location.

Keywords: Bidding, tendering, statistical models.

Introduction

Statistical models of construction contract bids have been used to indicate the likely effects of changed bidding strategies (Capen *et al.*, 1971; Fine and Hackemar, 1970; Flanagan and Norman, 1985; Friedman, 1956; Grinyer and Whittaker, 1973; Hossein, 1977; McCaffer, 1976; Mitchell, 1977; Oren and Rothkopf, 1975; Park, 1966; Shaffer and Micheau, 1971; Skitmore, 1982; Weverbergh, 1982; Whittaker, 1970), estimating practices (Barnes, 1971; Beeston, 1982; Morrison, 1984; Morrison and Stevens, 1980; Skitmore, 1981), procurement policies (AICBOR, 1967; Johnston, 1978; Wilson and Sharpe, 1988) and contractor selection (Cauwelaert and Heynig, 1978).

Many theoretical models have been proposed to represent the probability distribution of construction contract bids but relatively few have been derived empirically (Skitmore, 1989). Most of the empirical studies examine the effect of contract characteristics (type, size, number of bidders, procurement system, etc.) on the frequency distribution of pooled bids under the assumption that all bidders can be treated as behaving collectively in an identical (statistical) manner (Runeson, 1988; Whittaker, 1970; Wilson and Sharpe, 1988). This assumption is quite crucial in the analysis and, if inappropriate, could easily invalidate the results and proposals contained in these studies. Studies of individual bidding behaviour (Flanagan and Norman, 1985; Harvey, 1979; McCaffer, 1976), on the other hand, have made no attempt to test the suitability of the collective model.

In contemplating the design of such a test, it is clear that, as construction contracts are rarely of a similar size, some method is needed by which bids can be compared between contracts irrespective of the contract size. The popular approach to this problem is to take

0144–6193/91 $03.00+.12

the ratio of the bids to some measure of contract size (e.g. mean bid, lowest bid or designer's estimate) as the observational measure of analysis. This has been found wanting, however, on the grounds that the distribution of data standardized in this way is not directly comparable with the distribution of the unstandardized bids (Johnston, 1978). Also, as the statistical properties of ratios have received relatively little treatment in theoretical statistics, formulation of the relationship between the distribution of ratio standardized bids and the unstandardized bids is certain to be a major task. Furthermore, in order to simplify any possible future applications of the ensuing models, it would seem best to avoid ratio distributions altogether.

One alternative approach is to take the *difference* between the contract size and bids as the observational measure, a method used by Weverbergh (1970) in his analysis of consortium bids for oil tracts. Weverbergh's measure of contract size in this case was the value of the lowest bid. The criticism of this is that the lowest bid is a variable itself, rather than the constant term needed as a contract datum.

Another approach suggested by Skitmore (1982) overcomes the ratio and datum problem by taking the difference between logarithmic values of bids of individual bidders as the observational measure. This would seem to lead to estimates of scale and relative location parameters for each bidder but leaving the shape parameters an unknown quantity.

This paper describes the development of these ideas into a more explicit treatment of the subject area by an examination of some potential models of bids, the proposition of a new method of determining the contract datum, the estimation of the necessary distributional parameters and, finally, the application of some empirical tests to the bidder homogeneity assumption implied by previous research.

The data

Three separate sets of bidding data were obtained, referred to here as cases 1, 2 and 3. Case 1 data were donated by a construction company operating in the London area. The data covered all the company's building contract bidding activities during a 12-month period in the early 1980s for a total of 86 contracts. Details of the type of projects were available but not used in the analysis. Some of the data were incomplete, i.e. the value of some bids or the identity of bidders were not known by the company. In several cases, it was possible to supplement these data from the case 3 source. The resulting number of contracts for which a full set of bids, together with the identity of the bidder, were available for analysis totalled 51. A total of 93 bidders entered 318 bids.

Case 2 data were donated by a North of England County Council for building contract bids over approximately 4 years prior to July 1982. Details of 258 contracts were provided in a precoded format. In some cases, the codes were missing or no tenders had been received. In other cases, the codes or bids were illegible. The resulting number of contracts for which a full set of bids, together with the identity of the bidder, were available for analysis totalled 218. A total of 187 bidders entered 1235 bids.

Case 3 data were obtained from the records of a bidding information agency in the London area. The agency held details of most bids for most bidding contracts in the London area in card form. A period of 1 week was spent copying a sample of contracts data for the period November 1976 to February 1977. The bids and associated bidders' names were recorded and the names later encoded for analysis. The resulting number of contracts for which a full set of bids, together with the identity of the bidders, were available for analysis totalled 373. A total of 356 bidders entered 1915 bids.

Full details of all the case data are available from the author, subject to the approval of the source holders.

The model

An obvious specific model for contract bidding is:

$$x_{ij} \in X_{ij} \sim f_{ij}(\mu_{ij}, \sigma_{ij})$$

where x_{ij} is bidder i's bid ($i = 1, 2, \ldots r$) for contract j ($j = 1, 2, \ldots c$), and $f_{ij}(\ )$, μ_{ij} and σ_{ij} are the various shape, location and scale parameters respectively. Thus each bid is conceived as being drawn from a probability distribution unique to each bidder for each contract. The difficulty with this model is, of course, that it is impossible to provide empirical estimates of each of the parameters $f_{ij}(\ )$, μ_{ij} and σ_{ij} with effectively only one x_{ij} observation available. Clearly, for sufficient data to be available, some form of generalized model is necessary.

Nine possibilities for generalization exist for this model:

(a) *Shape generalization*

(i) by contracts:

$$f_{i.}(\ ) = f_{i1}(\ ) = f_{i2}(\ ) = \cdots = f_{ic}(\ )$$

(ii) by bidders:

$$f_{.j}(\ ) = f_{1j}(\ ) = f_{2j}(\ ) = \cdots = f_{rj}(\ )$$

(iii) by both contracts and bidders:

$$f_{..}(\ ) = f_{i.}(\ ) = f_{.j}(\ )$$

(b) *Scale generalization*

(i) by contracts:

$$\sigma_{i.} = \sigma_{i1} = \sigma_{i2} = \cdots = \sigma_{ic}$$

(ii) by bidders:

$$\sigma_{.j} = \sigma_{1j} = \sigma_{2j} = \cdots = \sigma_{rj}$$

(iii) by both contracts and bidders:

$$\sigma_{..} = \sigma_{i.} = \sigma_{.j}$$

(c) *Location generalization*

(i) by contracts:

$$\mu_{i.} = \mu_{i1} = \mu_{i2} = \cdots = \mu_{ic}$$

(ii) by bidders:

$$\mu_{.j} = \mu_{1j} = \mu_{2j} = \cdots = \mu_{rj}$$

(iii) by both contracts and bidders:

$$\mu_{..} = \mu_{i.} = \mu_{.j}$$

The nature of construction contracts is such that the assumption of a common contract size parameter, generalization (c)(i) and thus (c)(iii), is inappropriate generally (with the exception of Beeston's, 1974, Telephone Exchanges), and certainly for the case data considered in the present study. On the other hand, little empirical or theoretical evidence exists to contradict generalizations (a)(i) and (b)(i) – that distribution shape and spread is unaffected by any contract characteristics independently of bidders.

In view of these considerations, the aims of the study (to test a model for *bidders*) and the limitations of the data collected, it was proposed to test the common bidder generalizations contained in (a)(iii), (b)(ii) and (c)(ii).

The first attempt at this was to investigate the relative bidder approach by which the differences of the logarithmic values of bids are used instead of the more popular, but suspect, ratio values. This implies the model:

$$Y_{ij} \sim f_{..}(\mu_{i.}, \sigma_{i.}^2) \tag{1}$$

where $y_{ij} = \ln(x_{ij})$.

Thus, after taking the natural logs of the bids, the resulting values are taken to have been drawn from a probability distribution with a constant shape, $f_{..}(\ )$, a unique bidder variance, $\sigma_{i.}^2$, and a unique bidder mean, $\mu_{i.}$ (the subscript dots are omitted from the later analysis for notational ease).

To test the assumptions that the bidder location parameters $\mu_1 = \mu_2 = \cdots = \mu_i = \cdots = \mu_r$ and the bidder scale parameters $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_i^2 = \cdots = \sigma_r^2$, it was first necessary to estimate their values.

Parameter estimation: Relative bidder method

Consider a bid x_{lj} entered by bidder l ($l = 1, 2, \ldots, r$; $l \neq i$) for contract j; then, assuming bids are independent, estimates of the parameters $\mu_i - \mu_l$ and σ_i^2 in (1) may be obtained from:

$$\begin{gathered} \bar{y}_i - \bar{y}_l \simeq \mu_i - \mu_l \\ s_i^2 \simeq \sigma_i^2 \end{gathered}$$

by solving the two sets of equations:

$$\begin{aligned} E[Y_1] - E[Y_1] &= \bar{z}_{11} \\ E[Y_1] - E[Y_2] &= \bar{z}_{12} \\ \vdots \quad & \quad \vdots \\ E[Y_1] - E[Y_r] &= \bar{z}_{1r} \\ \vdots \quad & \quad \vdots \\ E[Y_r] - E[Y_r] &= \bar{z}_{rr} \end{aligned} \tag{2}$$

and

$$
\begin{aligned}
\mathrm{Var}[Y_1]+\mathrm{Var}[Y_1] &= s_{11}^2\\
\mathrm{Var}[Y_1]+\mathrm{Var}[Y_2] &= s_{12}^2\\
\vdots \qquad & \quad \vdots\\
\mathrm{Var}[Y_1]+\mathrm{Var}[Y_r] &= s_{1r}^2\\
\vdots \qquad & \quad \vdots\\
\mathrm{Var}[Y_r]+\mathrm{Var}[Y_r] &= s_{rr}^2
\end{aligned}
\tag{3}
$$

where

$$z_{ilj}=(y_{ij}-y_{lj})$$

$$z_{il}=\frac{1}{n_{il}}\sum_{j=1}^{c}\delta_{ilj}z_{ilj}$$

$$s_{il}^2=\frac{1}{(n_{il}-1)}\sum_{j=1}^{c}\delta_{ilj}(z_{ilj}-\bar{z}_{il})^2$$

$$n_{il}=\sum_{j=1}^{c}\sigma_{ilj}$$

$\delta_{ilj}=1$ when bidders i and l both enter bids for contract j,
0 otherwise

(after Skitmore, 1982)

If $f(\,)$ is normally distributed, $N(\,)$, the problem can, theoretically, be solved by the standard regression procedure as follows.

Letting the event that bidder i bids against bidder l be denoted by:

$$
W=\begin{pmatrix}
11 & 21 & \ldots & i1 & \ldots & r1\\
12 & 22 & \ldots & i2 & \ldots & r2\\
\vdots & \vdots & & \vdots & & \vdots\\
1l & 2l & \ldots & il & \ldots & rl\\
\vdots & \vdots & & \vdots & & \vdots\\
1r & 2r & \ldots & ir & \ldots & rr
\end{pmatrix}
$$

which is indexed by:

$$\begin{array}{lllll}
k=1 & \text{for } W^{12}, k=2 & \text{for } W^{13}, k=l-1 & \text{for } W^{1l}, k=r-1 & \text{for } W^{1r} \\
k=r & \text{for } W^{23}, k=r+1 & \text{for } W^{24}, k=r+l-3 & \text{for } W^{2l}, k=2r-3 & \text{for } W^{2r} \\
k=2r-2 & \text{for } W^{34}, k=2r-1 & \text{for } W^{35}, k=2r+l-6 & \text{for } W^{3l}, k=3r-6 & \text{for } W^{3r} \\
\vdots & \vdots & \vdots & \vdots &
\end{array}$$

$$k=\left\{(i-1)r-\sum_{p=1}^{i-1} p\right\}+1 \text{ for } W^{i,i+1} \ldots k=\left\{(i-1)r-\sum_{p=1}^{i-1} p\right\}+l \text{ for } W^{il}$$

$$k=ir-\sum_{p=1}^{i} p \text{ for } W^{ir}$$

$$n=k=\left\{(r-1)r-\sum_{p=1}^{r} p\right\} \text{ for } W^{r-1,r}$$

Then, letting

$X_{kij}=1$ when the event indexed by k occurs, which includes the lower numbered bidder i, on contract j, and

$X_{kij}=-1$ when the event indexed by k occurs, which includes the higher numbered bidder i, on contract j, otherwise

$X_{kij}=0$

and z_{kj} is the difference in bids between the two bidders when event k occurs on contract j, the normal equations are:

$$\begin{array}{l}
b_1X_{111}+b_2X_{121}+b_3X_{131}+\cdots+b_rX_{1r1}=z_{11} \\
b_1X_{211}+b_2X_{221}+b_3X_{231}+\cdots+b_rX_{2r1}=z_{21} \\
\vdots \\
b_1X_{k11}+b_2X_{k21}+b_3X_{k31}+\cdots+b_rX_{kr1}=z_{k1} \\
\vdots \\
b_1X_{m11}+b_2X_{m21}+b_3X_{m31}+\cdots+b_rX_{mr1}=z_{m1} \\
\hdashline
b_1X_{112}+b_2X_{122}+b_3X_{132}+\cdots+b_rX_{1r2}=z_{12} \\
b_1X_{212}+b_2X_{222}+b_3X_{232}+\cdots+b_rX_{2r2}=z_{22} \\
\vdots \\
b_1X_{k12}+b_2X_{k22}+b_3X_{k32}+\cdots+b_rX_{kr2}=z_{k2} \\
\vdots
\end{array}$$

$$\begin{array}{l}
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
b_1X_{m12} + b_2X_{m22} + b_3X_{m32} + \cdots + b_rX_{mr2} = z_{m1} \\
\hline
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\hline
b_1X_{11c} + b_2X_{12c} + b_3X_{13c} + \cdots + b_rX_{1rc} = z_{1c} \\
b_1X_{21c} + b_2X_{22c} + b_3X_{23c} + \cdots + b_rX_{2rc} = z_{2c} \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
b_1X_{k1c} + b_2X_{k2c} + b_3X_{k3c} + \cdots + b_rX_{krc} = z_{kc} \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
b_1X_{m1c} + b_2X_{m2c} + b_3X_{m3c} + \cdots + b_rX_{mrc} = z_{mc}
\end{array}$$

Estimates of $E[Y_1]$, $E[Y_2]$, . . . will therefore be provided by the vector:

$$\mathbf{B} = \mathbf{C}^{-1}\mathbf{D}$$

where

$$\mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_r \end{pmatrix} \qquad \mathbf{D} = \begin{pmatrix} \sum_k^m \sum_j^c X_{k1j}z_{kj} \\ \sum_k \sum_j X_{k2j}z_{kj} \\ \cdot \\ \cdot \\ \sum_k \sum_j X_{krj}z_{kj} \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} \sum_k^m \sum_j^c X_{k1j}^2 & \sum_k \sum_j X_{k1j}X_{k2j} \ldots & \sum_k \sum_j X_{k1j}X_{krj} \\ \sum_k \sum_j X_{k2j}X_{k1j} & \sum_k \sum_j X_{k2j}^2 & \ldots \sum_k \sum_j X_{k2j}X_{krj} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \sum_k \sum_j X_{krj}X_{k1j} & \sum_k \sum_j X_{krj}X_{k2j} \ldots & \sum_k \sum_j X_{krj}^2 \end{pmatrix} \tag{4}$$

And the variance of b_1 is estimated by:

$$\text{Var } b_1 = \frac{S}{N-r} c^{ii}$$

where

$$S = \sum_{k}^{m} \sum_{j}^{c} \{z_{kj} - (b_1 X_{k1j} + b_2 X_{k2j} + \cdots + b_r X_{krj})\}^2$$

and N is the total number of paired observations.

The major difficulty with this approach is in the sparseness of the matrix system. In each row in equation (4) there are $r-2$ empty cells. Afifi and Elashoff (1966) have reviewed the literature on the problem of handling multivariate data with observations missing for some or all of the variables under study, noting that the estimation problems can often be simplified if the missing data follow certain patterns. Hocking and Smith (1968) have used estimates of parameters from one part of a (multivariate normal) data structure to insert into the other parts prior to using an iterative procedure. Elman (1982) has considered the use of direct and iterative methods of solving large sparse non-symmetric systems of linear equations finding difficulties with direct methods due to the factoring process generating many more non-zeros than the coefficient matrix, thereby increasing the computational storage size needed. A further problem encountered was that the number of arithmetic operations could become excessive. His general conclusion was that '. . . although progress has been made in the development of orderings for the unknowns that decrease the complexity of directness for solving sparse problems . . . many large sparse problems cannot be solved by direct methods on present day computers'.

Some early tests on the data using matrix methods confirmed Elman's view that direct methods were unsuitable. The extreme sparseness of the data under study, in the proportion cr to $c(r-2)$, produced results severely distorted by computational rounding errors. It was, therefore, considered that an iterative procedure would be more appropriate.

The iterative procedure

A contract datum parameter was introduced into the model (1) resulting in the revised model:

$$\ln(x_{ij}) = y_{ij} \in Y_{ij} \sim f(\alpha_i + \beta_j, \sigma_i^2) \tag{5}$$

where α_i is a bidder location parameter, β_j is a contract datum parameter and $\alpha_i + \beta_j = \mu_{ij}$. For the purposes of parameter estimation, this requires the solution of:

$$y_{ij} = \alpha_i + \beta_j + \varepsilon_{ij} \tag{6}$$

where ε_{ij} is $f(o, \sigma_i^2)$. [It was noted that $y_{ij} - y_{lj} = \alpha_i - \alpha_l + \varepsilon_{ij} - \varepsilon_{il}$, where $\varepsilon_{ij} - \varepsilon_{il}$ is $f(o, \sigma_i^2 + \sigma_l^2)$ and that, although appropriate for differences, equation (6) was preferred as less information is lost.]

Assuming $f(o, \sigma_i^2)$ is $N(o, \sigma_i^2)$:

$$y_{ij} \text{ has a pdf } \frac{1}{\sigma_i^2} \exp\left\{-\frac{1}{2\sigma_i^2}(y_{ij}-\alpha_i-\beta_j)^2\right\}$$

So the log-likelihood is:

$$\ln L = -\sum_{i=1}^{r} (n_i/2) \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{r} (1/\sigma_i^2) \sum_{j=1}^{c} \delta_{ij}(y_{ij}-\alpha_i-\beta_j)^2$$

where $\delta_{ij}=1$ if bidder i bids for contract j
$=0$ if bidder i does not bid for contract j

$n_i = \sum_{j=1}^{c} \delta_{ij}$ = number of bids made by bidder i

The MLL over α's, β's and σ^2 is

$$\frac{\delta \ln L}{\delta \beta_j} = \sum_{i=1}^{r} \delta_{ij}(y_{ij}-\alpha_i-\beta_j)/\sigma_i^2 = 0$$

$$\Rightarrow \beta_j = \sum_{i=1}^{r} \delta_{ij}(y_{ij}-\alpha_i)/n_i \qquad (7)$$

$$\frac{\delta \ln L}{\delta \alpha_i} = (1/\sigma_i^2) \sum_{j=1}^{c} \delta_{ij}(y_{ij}-\alpha_i-\beta_j) = 0$$

$$\Rightarrow \alpha_i = \sum_{j=1}^{c} \delta_{ij}(y_{ij}-\beta_j)/n_i \qquad (8)$$

$$\frac{\delta \ln L}{\delta \sigma_i^2} = \frac{-n_i}{2\sigma_i^2} + \frac{1}{2\sigma_i^4} \sum_{j=1}^{c} \delta_{ij}(y_{ij}-\alpha_i-\beta_j)^2$$

$$\Rightarrow \sigma_i^2 = \frac{1}{n_i} \sum_{j=1}^{c} \delta_{ij}(y_{ij}-\alpha_i-\beta_j)^2$$

The procedure used was to initialize all $\alpha_i=0$ and iterate equations (7) and (8) to convergence. The estimates of σ_i^2 were adjusted for bias by the approximation:

$$\sigma_i^{2\prime} = \sigma_i^2 \left\{ \frac{n_i}{(n_i-1)\left(1-\dfrac{c-1}{N-r}\right)} \right\}$$

where $N = \sum_{j=1}^{c} n_i$, the total number of observations.

For computational purposes, it is unnecessary to introduce once only bidders, $n_i = 1$, until after convergence of the iteration procedure. Convergence was taken to have occurred when the largest change in estimated value of any α_i in consecutive iterations was less than e_α, where e_α is small (an appropriate value of e_α was found, after various trials, to be 10^{-7}). A sample of the estimated values of β_j, α_i and σ_i^2 obtained by this iterative method for case 1 data is given in Table 1.

Table 1. Sample of predictions from iterative procedure for case 1

(a) α_i and σ_i^2 values

Bidder (i)	No. of bids (n_i)	α_i	σ_i^2
6	2	0.01667	0.00052
8	1	0.00717	
12	1	−0.00871	
20	1	−0.07646	
24	7	−0.04071	0.00099
31	1	−0.03508	
55	20	0.03154	0.00253
60	2	0.06268	0.00292
64	1	−0.03523	
72	1	−0.02970	
73	1	0.00579	
75	2	−0.02239	0.00273
79	4	−0.04906	0.00042
83	2	0.09766	0.00361

(b) β_j values

Project (j)	β_j
1	14.18677
2	13.15789
3	14.09286
4	13.42515
5	12.88743
6	14.58104
7	14.95664
8	15.82349
9	13.70479
10	13.90194

Having now reached the point where estimates of the bidders' scale parameters' σ_i and location parameters α_i were available subject to the assumption that the shape parameter $f(\ )$ is normally distributed $N(\ )$, we were now in a position to start testing the bidder shape, scale and location generalizing assumptions described earlier by inspection of the residuals from the iterative procedure.

Following the procedure adopted in most other previous studies of bidding data, the next step taken was to examine the pooled residuals in order to gain some impression of the overall distribution shape and the validity of the normal model prior to testing the common bidder shape assumption $f(\ )$. The results of this analysis are described in the next section.

Distribution of pooled residuals

Shape

Figures 1–3 show the frequency distribution of the pooled residuals $z_{ij}=y_{ij}-\hat{\alpha}_i-\hat{\beta}_j$, $y_{ij}=\ln(x_{ij})$ for each of the three cases. Visual inspection of the shapes of these frequency distributions suggested that the Normal distribution may be an appropriate model in all cases. The cumulative probability plots (Figs 4–6), however, provide evidence to the contrary, the data being rather heavy-tailed. Superimposition of Normal curves on the histograms (Figs 1–3) illustrates the position.

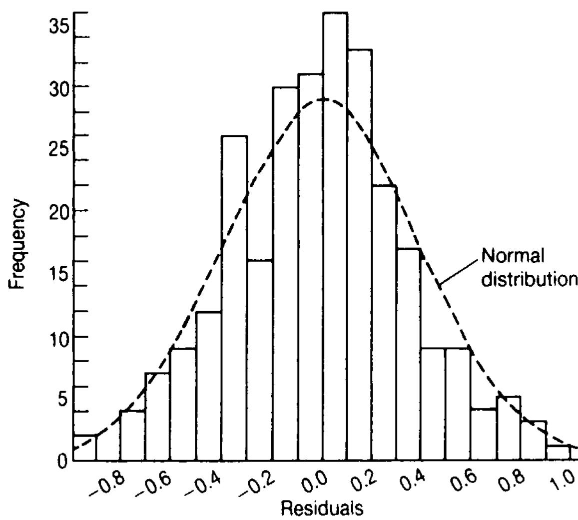

Fig. 1. Case 1: Distribution of residuals.

The plots do seem to indicate that the distribution of residuals may be similar for each case, however. This was tested by comparing the frequency distribution of residuals for each case with the frequency distribution of the pooled residuals for all cases. On the assumption that the pooled residuals represent the universal population of residuals, a chi-square test was applied to test the hypothesis that the standardized residuals for each case were samples from the same universal population. The results of this test indicated that the hypothesis should not be rejected, chi-square values of 2.9 (8 d.f.), 3.6 (10 d.f.) and 4.0 (14 d.f.) being recorded for cases 1, 2 and 3 respectively.

Having concluded that the distribution of the (standardized) residuals could be considered to be the same for all cases, the residuals were pooled and some tests applied to determine the shape of this distribution.

Normal model

As mentioned above, a visual inspection of the frequency distribution of the pooled residuals suggested the Normal model to be a possible approximation. On attempting to fit a Normal distribution of zero mean and unity variance, it was immediately apparent that a smaller

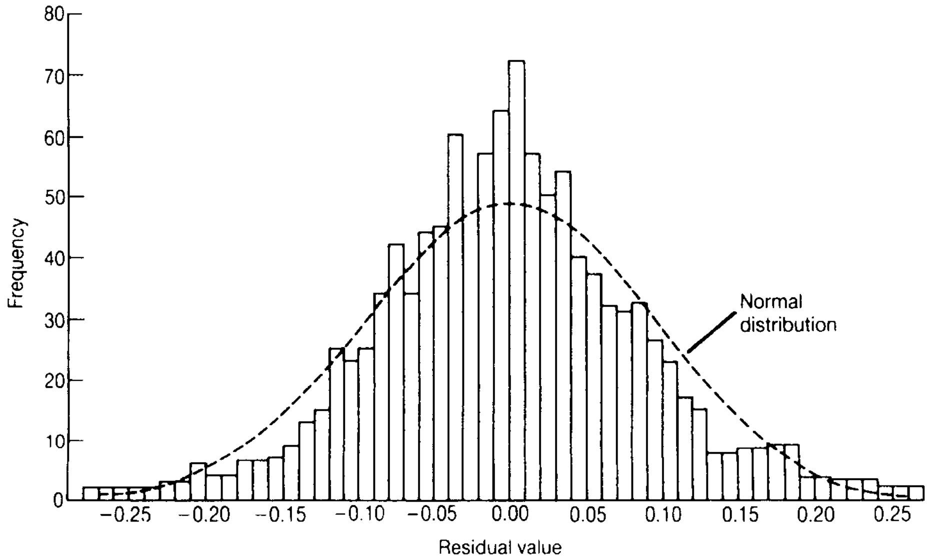

Fig. 2. Case 2: Distribution of residuals.

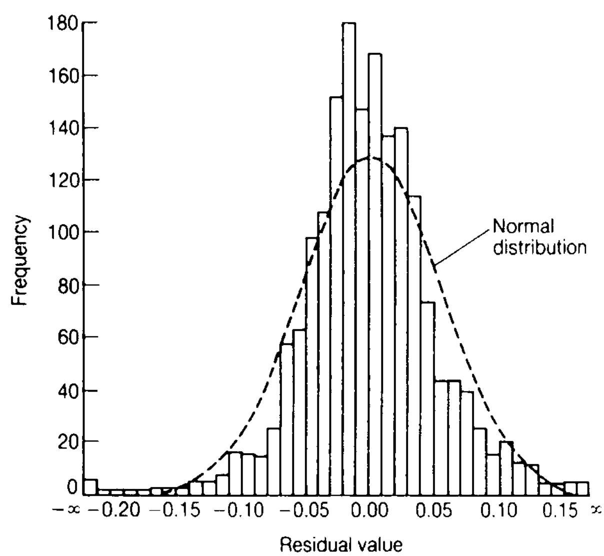

Fig. 3. Case 3: Distribution of residuals.

variance would provide a better visual fit. Several variances were, therefore, attempted (Fig. 7). Plots of Normal order statistics against the frequency of the pooled residuals were made. Figure 8 shows the plot against $N(0, 1)$ and Fig. 9 against $N(0, 0.6)$. This visual inspection suggested the $N(0, 0.6)$ model to be the best approximation. This visual fit was not

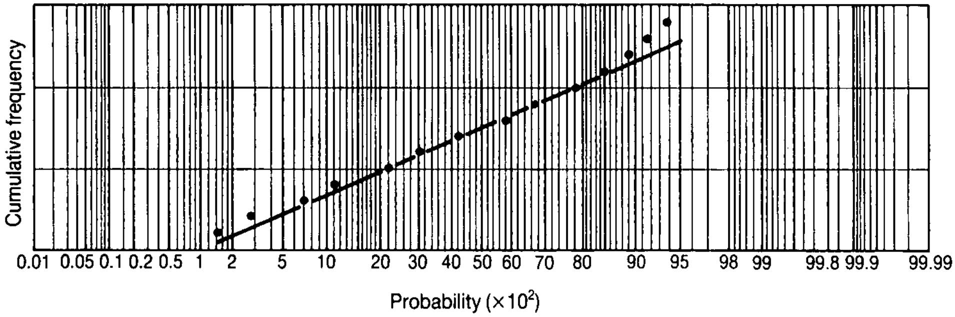

Fig. 4. Case 1: Probability plot of residuals.

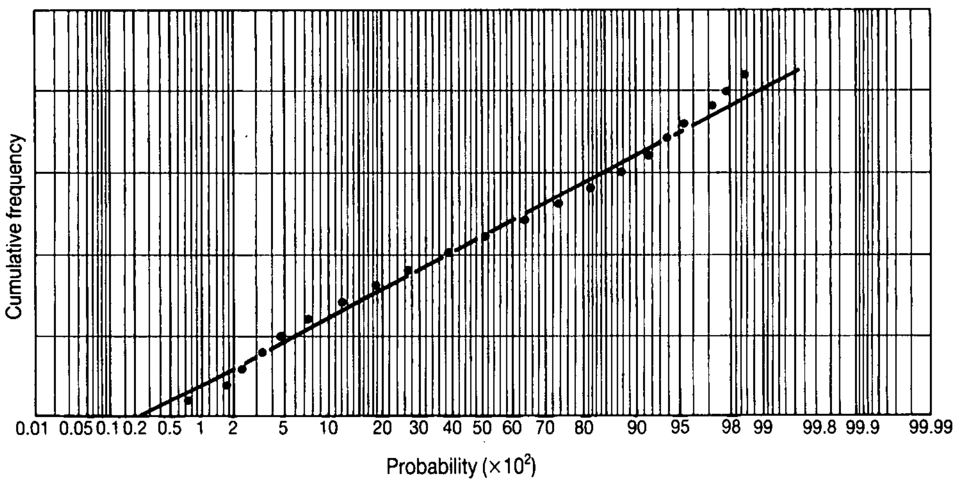

Fig. 5. Case 2: Probability plot of residuals.

confirmed, however, by the chi-square and Kolmogorov-Smirnov goodness of fit tests (Table 2).

As the critical values at the 5% significance level are $\chi^2_{(23)} < 35.17$ and $D < 0.024$, none of the Normal distributions attempted were judged to be of a sufficiently good fit. The results for $\sigma^2 = 1.0$ are, of course, not surprising, as $\sigma^2 = 1.0$ is the best estimate for standardized values.

The visual closeness of the distribution of the pooled residuals to a Normal distribution suggested that some function of the Normal distribution would be the best approach. Pearson's distributions were first consulted to check the possibility that a relatively simple unique function may suffice.

Pearson's distributions

The criterion k was calculated from the formula

$$k = \frac{\beta_1(\beta_2+3)^2}{4(2\beta_2-3\beta_1-6)(4\beta_2-3\beta_1)}$$

and found to be between zero and unity, suggesting the Type-IV distribution may be the most appropriate. However, as the Type-IV distribution requires repeated numerical integration of the pdf in its application, it was decided that a computationally more efficient distribution should be adopted. One distribution of this type is the Gram-Charlier Type A series or the Edgeworth expansion.

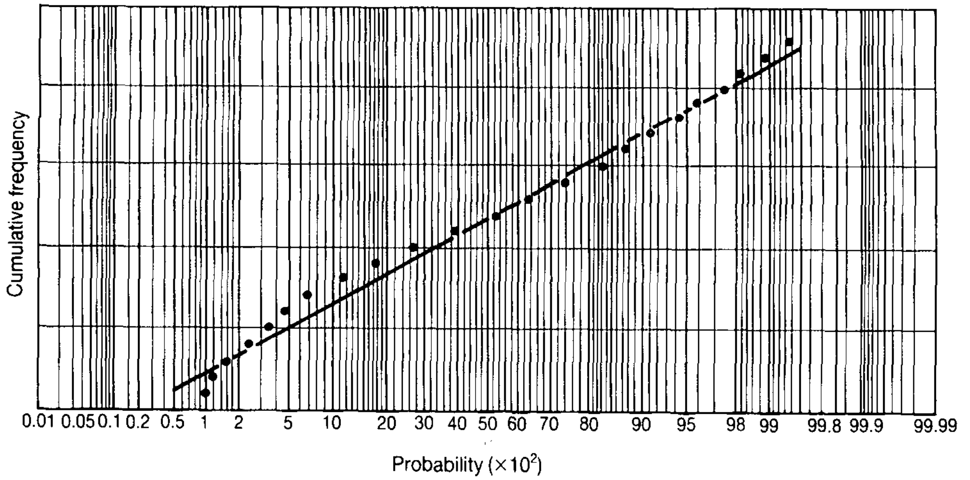

Fig. 6. Case 3: Probability plot of residuals.

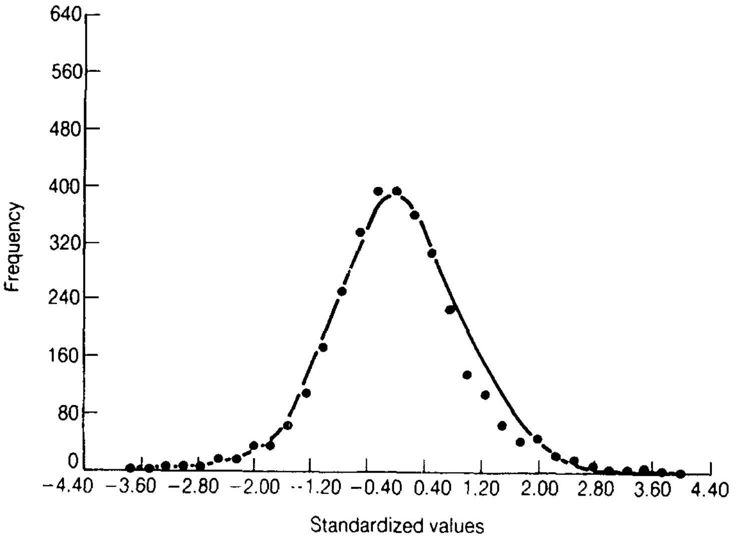

Fig. 7. Distribution of pooled residuals.

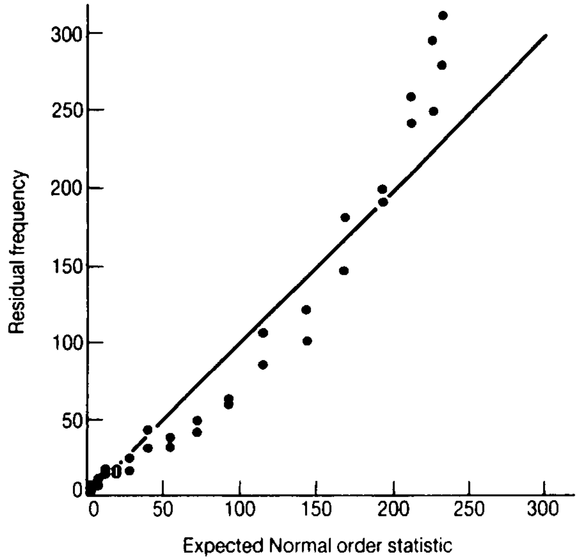

Fig. 8. Plot of Normal order statistics $N(0, 1)$ against the frequency of the standardized pooled residuals.

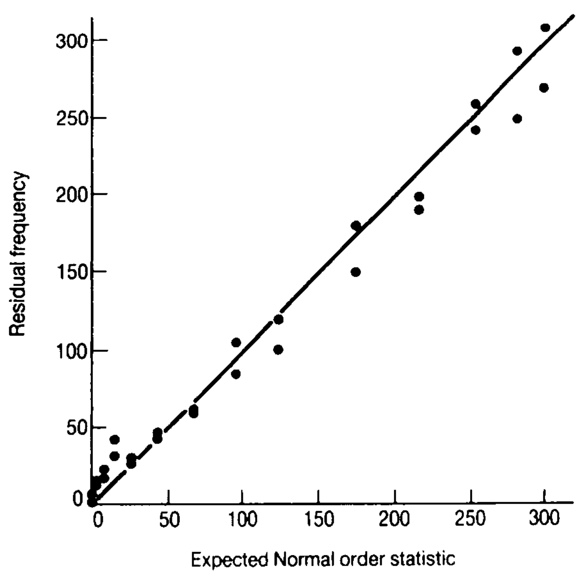

Fig. 9. Plot of Normal order statistics $N(0, 0.6)$ against the frequency of the standardized pooled residuals.

Table 2. Goodness of fit tests for various Normal distributions to the standardized pooled residuals

$N(0, \sigma^2)$	$\chi^2_{(23)}$	K-S D statistic
$\sigma^2 = 1.0$	204	0.047
0.9	229	0.063
0.8	322	0.126
0.7	547	0.203
0.6	876	0.298

Gram-Charlier Type A series and Edgeworth's form

Kendall and Stuart (1963, p. 162) note that the Type A series can encounter difficulties when more than the first four cumulants are included, unless the skewness coefficient (β_1) is 'close enough' to zero. A coefficient of $|\beta_1| \geq 0.25$ will produce a non-unimodal distribution with Edgeworth's form. Similarly, $|\beta_2| \geq 0.50$ produces negative frequencies. The Gram-Charlier series has a wider range of acceptability, but certainly non-unimodal if $|\beta_1| \geq 0.7$.

As the skewness coefficient of the pooled residuals was 0.057, it was considered that both Gram-Charlier and Edgeworth's forms would be appropriate for more than the fourth cumulant. Both the Gram-Charlier and Edgeworth forms being effectively the same, the Gram-Charlier Type A series was chosen to model the residuals.

The first four terms of the Gram-Charlier Type A series give a Cdf as follows:

$$\mathrm{Cdf}(x) = \{1/(2\pi)^{1/2}\} \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}x^2)\, dx - [\{1/(2\pi)^{1/2} \exp(-\tfrac{1}{2}x^2)\}\{(\mu_3/6)(x^2-1) + (1/24)(\mu_4-3)(x^3-3x)\}]$$

where x is standardized. The population moments μ_3 and μ_4 were estimated from the data as 0.16 and 6.03 respectively.

After a series of trials, it was found that a value of 4.5 for the fourth moment produced a more satisfactory fit (Fig. 10). Tests of goodness of fit resulted in $\chi^2_{(27)} = 32.7$ and Kolmogorov's $D = 0.012$, the data not being significantly different from the theoretical distribution (at 5% significance). A casewise check was made to establish that the model was appropriate in all cases, by fitting the model to the standardized residuals of each of the three sets of residuals (Figs 11–13). Table 3 provides the results of the goodness of fit tests.

Table 3. Gram-Charlier Type A series: Goodness of fit tests

Case	χ^2	d.f.	Crit. value	Kolmogorov's D	Crit. value
1	13.3	15	25.0	0.035	0.081
2	19.8	23	35.0	0.016	0.040
3	31.5	24	36.0	0.019	0.032
All cases	32.7	27	—	0.012	—

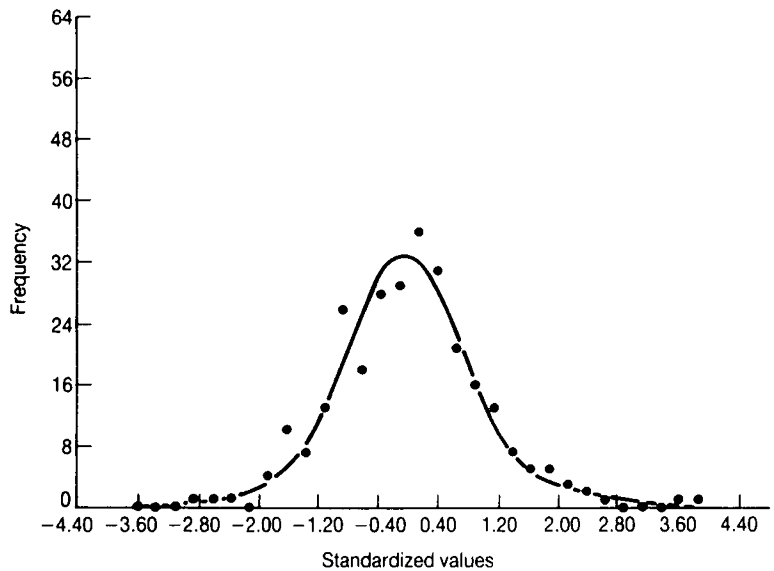

Fig. 10. Distribution of residuals 1.

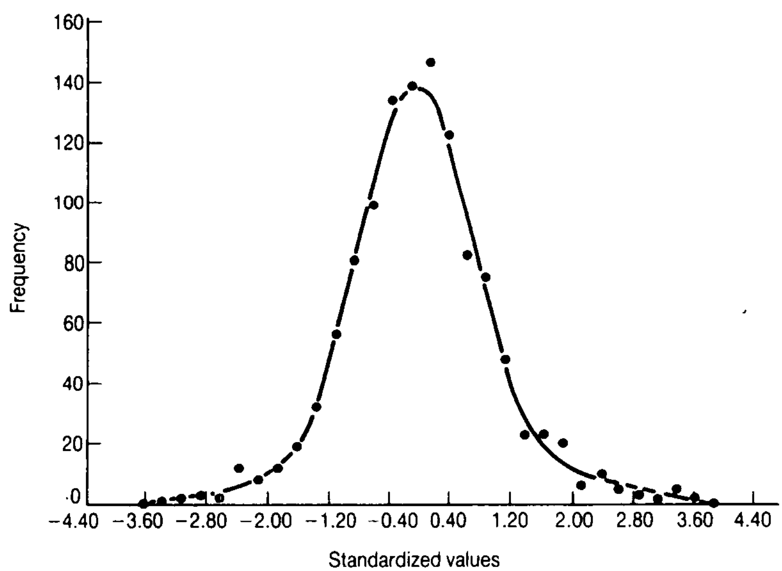

Fig. 11. Distribution of residuals 2.

As a final check, a simulation exercise was conducted in which random standardized values were generated for each bidder from a Gram-Charlier Type A distribution for $\mu_3 = 0.16$ and $\mu_4 = 4.5$. The chi-square and K-S tests both failed to reject the null hypothesis in each and every case. It was, therefore, concluded that the Gram-Charlier series with parameters $\mu_3 = 0.16$ and $\mu_4 = 4.5$ was a reasonable model of the standardized residuals.

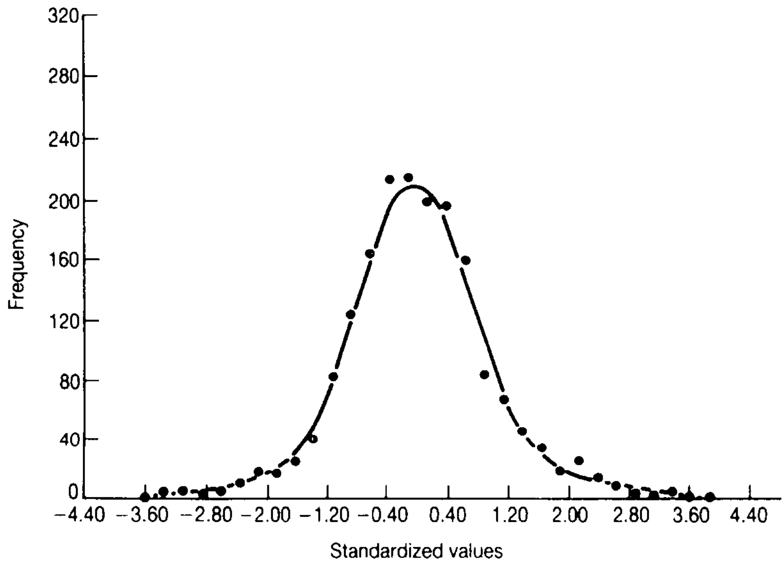

Fig. 12. Distribution of residuals 3.

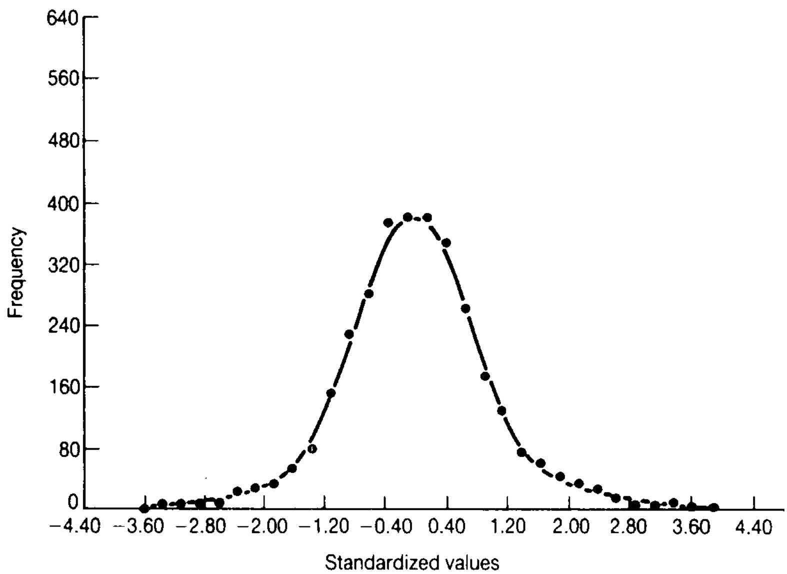

Fig. 13. Distribution of pooled residuals.

Discussion

The discovery of high-peaked, heavy-tailed distributions is not a new phenomenon in empirical studies of data of these kind. Ali and Giacotto (1982), for instance, in their study of stock market prices, found that '. . . empirical distributions of price changes are usually high peaked with heavy tails when compared with the normal distribution'. Studies by Clark (1973) and Hsu *et al.* (1974) of similar data suggest that '. . . if the price changes are normal but *not* identically distributed, it is likely that the empirical distributions would be highly peaked and heavy tailed compared to the Normal distribution' (Ali and Giacotto, 1982, p. 19), the major differences being attributed to the existence of scale differences. Thus it is possible that individual price changes are really normally distributed but with different scale and location parameters, the pooling process obscuring the analysis.

On this evidence, therefore, it was clearly necessary to avoid this problem by examining the distribution of individual bidders' bids prior to pooling.

Distribution of individual bidders' bids

Shape

The aim of this analysis was to test the assumption that the shape of the distribution of bids entered by each bidder was the same for each bidder, having first removed the contract 'size' by the iterative procedure described above, but without pooling. In order to carry out this test, it was necessary to find a parametric shape that would reasonably model these distributions of bids. More formally, it is assumed that 'each of a set of observations $x=\{x_1, \ldots, x_n\}$ is independently drawn from a density $p(x_i|\mu, \sigma, \Omega)$, where μ and σ are location and scale parameters, and Ω is a shape parameter that indexes a parametric family that contains the simplifying assumption as a special case Ω_0' (Spiegelhalter, 1983, p. 401).

In the present context, where there are several sets of observations (one set for each bidder) for each case, the notation $z_i=\{z_{1i}, \ldots, z_{ni}\}$ is adopted and the density is $p(z_{qi}|\mu_i, \sigma_i, \Omega)$ for the i^{th} bidder ($i=1, \ldots, r$), where $z_i=y_{ij}-\hat{\beta}_j$, obtained after the iteration procedure. Thus the distribution shape under consideration is that of the values $y_{ij}-\hat{\beta}_j$ around $\hat{\alpha}_i$ for each i.

No global procedure appears to be available as yet by which estimates of Ω can be obtained from the data. It has been suggested that 'we can go a long way in the process of approximation if we make the distribution used have the correct values [estimated from the data] of its first four moments' (Pearson, 1963, p. 109). In determining Ω, this implies the use of the third and fourth moments only and in particular the coefficients of skewness (Y_1) and kurtosis (Y_2), which are independent of any linear transformation of z. Two significant problems exist in this approach to these data in the absence of pooling. First, there are several sets of observations and therefore several estimates of Y_1 and Y_2, that is $Y_{11}, Y_{12}, \ldots, Y_{1r}$ and $Y_{21}, Y_{22}, \ldots, Y_{2r}$ for each bidder. Secondly, each set of observations contains different, and few, values. It is clear, therefore, that some knowledge is needed of the distribution of small sample estimates of the standardized third and fourth moments of potential models. In addition, a procedure is also needed to combine the moments obtained from different sample sizes.

McCaffer and Pettitt (1976), in testing the distribution of contract bids for normality against uniformity, were able to overcome these problems by testing the uniformity of the distribution of the approximate probability of the Anderson-Darling A^2 statistic obtained for each contract. Although this method can clearly be extended to testing the validity of distributions other than normal and uniform, it does rely on the independence of the shape and scale parameters of the distribution being tested, a property not readily found in other distributions.

In view of these considerations, it was necessary to limit the tests to the bidder distributions most commonly proposed in the literature, the Normal (Alexander, 1970; Cauwelaert and Heynig, 1978; Emond, 1971; McCaffer, 1976; Mitchell, 1977; Morrison and Stevens, 1980) and Log-normal (Brown, 1974; Capen *et al.*, 1971; Crawford, 1970; Klein, 1976; Pelto, 1971; Weverbergh, 1982; Wilson and Sharpe, 1988).

On the evidence of the statistical literature (Shapiro *et al.*, 1968, for instance), it appears that the unpredictability of small sample test statistics, certainly for the almost symmetrical distributions anticipated with these data, precludes the selection of one particularly most powerful statistic to test the distributions under consideration. In the event, a battery of five different test statistics was used: (1) the sample coefficient of skewness, Y_1; (2) the sample coefficient of kurtosis, Y_2; (3) Geary's (1935, 1936) '*a*' statistic; (4) the sample studentized range, W/S; and (5) the Anderson-Darling A^2 statistic.

The exact small sample distribution of these five statistics is as yet unknown, and only Pettitt (1975), for the Anderson-Darling statistic, has obtained an adequate approximation. The necessary percentage points for the other four statistics were therefore estimated by a lengthy simulation procedure involving the computer generation of 20 000 pseudo-random samples of size $n=3(1)14$ [$n=4(1)14$ for the Y_2 statistic].

As it was expected that the distribution of $y_{ij}-\hat{\beta}_j$ around the $\hat{\alpha}_i$ estimated values would be modified by the estimation process, a simulation study of Log-normally distributed bids was conducted by generating appropriate values of y_{ij} from a $N(\mu_j, \sigma_j^2)$ distribution, where μ_j and σ_j^2 were obtained from the known relationship for the Log-normal distribution (Aitchison and Brown, 1957):

$$\mu_j = \ln(\bar{x}_j) - \tfrac{1}{2}\sigma_j^2$$
$$\sigma_j^2 = \ln\{(\bar{x}_j^2/s_j^2) - 1\}$$

x_j and s_j^2 being obtained from the raw data. The resulting simulated bids were then subjected to the iterative procedure, values for $\hat{\alpha}_i$ and $\hat{\beta}_j$ computed and thence the values $z_{ij}=y_{ij}-\hat{\beta}_j$. The z_{ij} values were then subjected to the five shape tests, the probabilities of each of the resulting statistics tested for uniform distribution by the chi-square test, and the probability of the chi-square statistic obtained for the case. The process was repeated 1000 times for each case.

The critical values estimated from these simulations were found to be largely unaffected by the iterative procedure, the probabilities being very close to those expected.

Three classes of transformations of the data were made: (1) $y_{ij}=\ln(x_{ij}-m\bar{x}_j)$, where $\bar{x}_j$ is the sample mean of the bids; (2) $y_{ij}=\ln(x_{ij}-mx_j)$, where x_j is the value of the lowest bid; and (3) $y_{ij}=x_{ij}^{1/p}$. Thus $m=0$ in transformations (1) and (2) tests the usual two-parameter Log-normal assumption, and $p=0$ in transformation (3) tests the Normal assumption.

Table 4 gives the results of the shape tests for all cases. The numbers in columns 2–4 indicate the tests which failed at the 5% significance level. As the table shows, only transformation (2) with $0.5<m<0.9$ produces a Normal distribution for all the three cases examined, indicating a three-parameter Log-normal distribution to be a satisfactory model throughout.

Spread

Having obtained, via the iterative procedure, estimates of the variance associated with each bidder, it was now proposed to test the homoscedascity assumption that each bidder has the

Table 4. Results of tests of distribution shape for each bidder

Transformation	Test results (*no fails) Case 1	2	3
$y_{ij}=\ln(x_{ij}-m\bar{x}_j)$			
$m=0.0$	*	*	2, 3, 5
$m=0.1$	*	*	2, 3
$m=0.2$	*	5	2, 3
$m=0.3$	*	5	2, 3, 5
$m=0.4$	*	2, 5	2, 3
$m=0.5$	*	—	2, 3, 5
$m=0.6$	*	—	2, 3, 4
$y_{ij}=\ln(x_{ij}-mx_j)$			
$m=0.0$	*	*	2, 3, 5
$m=0.1$	*	*	2, 3
$m=0.2$	*	*	2, 3
$m=0.3$	*	*	2, 3
$m=0.4$	*	*	2, 3, 4, 5
$m=0.5$	*	*	2, 5
$m=0.6$	*	*	*
$m=0.7$	*	*	*
$m=0.8$	*	*	*
$m=0.9$	*	1	*
$m=0.95$	*	1, 5	*
$y_{ij}=x_{ij}^{1/p}$			
$p=1$	2	1, 2, 3, 4, 5	1, 2, 3, 4, 5
$p=2$	1	2, 5	1, 2, 3, 4, 5
$p=3$	*	*	1, 2, 3, 5
$p=4$	*	3, 4	1, 2, 3
$p=5$	*	*	2, 3
$p=6$	*	*	2, 3
$p=7$	*	3	2, 3
$p=8$	*	*	2, 3, 4

same scale parameter, $s_1^2=s_2^2=\cdots=s_i^2=\cdots=s_r^2$. To examine the validity of the homoscedascity assumption with the data used in this study, Bartlett's test was applied to the z_{ij} values obtained from the iterative procedure.

As with the shape tests, some problems were anticipated with presence of small sample sizes. The possibility also existed that the distribution of the Bartlett test statistic may be affected by the iterative procedure itself. A further simulation study was, therefore, conducted by generating Log-normal bid values for each bidder with an equal variance (estimated from the data). The iterative procedure was implemented and the probability of Bartlett's statistic computed from the residuals $y_{ij}-\hat{\beta}_j$ for each bidder. This was repeated 1000 times for cases 1 and 2, 100 times for case 3, due to the length of time for each trial (20 min for case 3). The values at the critical percentage points are shown in Table 5. The low values obtained for case 2 were unaccounted for but, as the same computer program was used for all cases, the values were accepted as being correct.

Table 5. Simulated variance tests: Critical values of Bartlett's probability (simulation of log values obtained by iteration, average variance, Normal distribution)

Test statistics at	Case 1	Case 2	Case 3
1%	0.010030	0.001980	0.003340
$2\frac{1}{2}$%	0.029890	0.003500	0.019610
5%	0.051730	0.008940	0.036155
$7\frac{1}{2}$%	0.084520	0.016180	0.051650
10%	0.117130	0.021650	0.066625
95%	0.953750	0.771060	0.935750
$97\frac{1}{2}$%	0.976570	0.828170	0.978470
99%	0.990170	0.932930	0.993800

The results of the analysis of the actual case data are given in Table 6 for the various transformations employed earlier. As Table 6 shows, the homoscedascity assumption is not generally valid for cases 2 and 3.

Location

It was now proposed to test the assumption that each bidder has the same location parameter $\alpha_1 = \alpha_2 = \cdots = \alpha_i = \cdots = \alpha_r$. The approximation test chosen was the analysis of variance (ANOVA). The ANOVA relies on the assumption that the data are Normally distributed and the variances are equal, although the test is known to be robust, certainly as far as the Normal assumption is concerned (Kendall and Stuart, 1963, p. 465).

The ANOVA, in this case, involves the test statistic

$$F = \frac{\mathrm{SSQ}_B/(r-2)}{\mathrm{SSQ}_w/(N-c-r+1)}$$

where

$$\mathrm{SSQ}_w = \sum_i \sum_j \delta_{ij}(y_{ij} - \hat{\alpha}_i - \hat{\beta}_j)^2$$

$$\mathrm{SSQ}_B = \sum_i n_i(\hat{\alpha}_i - \bar{\alpha})^2$$

$$\bar{\alpha} = \hat{\alpha}^i \Big/ \sum_i n_i$$

And F follows the F distribution with $N-r-c+1$ and $r-2$ degrees of freedom. An example for case 1 is given in Table 7.

The results obtained for the various transformations are provided in Table 8. Apart from the raw data for case 3 (which is considered to be highly non-Normal and heterogeneous

Table 6. Results of tests for homoscedascity of bidders (Bartlett's test)

	Case					
	1		2		3	
Transformation	$\chi^2_{(49)}$	P	$\chi^2_{(127)}$	P	$\chi^2_{(231)}$	P
$y_{ij}=\ln(x_{ij}-m\bar{x}_j)$						
$m=0.0$	60.6	0.124	257.0	0.000	446.0	0.000
$m=0.1$	60.1	0.133	258.3	0.000	446.0	0.000
$m=0.2$	59.5	0.145	260.6	0.000	446.5	0.000
$m=0.3$	58.8	0.159	264.7	0.000	448.0	0.000
$m=0.4$	58.1	0.176	272.9	0.000	451.9	0.000
$m=0.5$	57.2	0.197	—	—	462.9	0.000
$m=0.6$	56.1	0.225	—	—	485.7	0.000
$y_{ij}=\ln(x_{ij}-mx_j)$						
$m=0.0$	60.6	0.124	257.0	0.000	446.0	0.000
$m=0.1$	59.6	0.143	253.5	0.000	438.7	0.000
$m=0.2$	58.5	0.166	249.7	0.000	431.0	0.000
$m=0.3$	51.4	0.192	245.5	0.000	422.5	0.000
$m=0.4$	56.2	0.223	240.8	0.000	415.2	0.000
$m=0.5$	55.0	0.259	235.6	0.000	398.2	0.000
$m=0.6$	53.6	0.300	229.8	0.000	358.4	0.000
$m=0.7$	52.4	0.344	223.8	0.000	358.4	0.000
$m=0.8$	51.9	0.361	218.5	0.000	330.9	0.000
$m=0.9$	52.6	0.335	219.5	0.000	290.3	0.005
$m=0.95$	42.1	0.745	234.7	0.000	260.5	0.089
$y_{ij}=x_{ij}^{1/p}$						
$p=1$	125.5	0.050	1400.1	0.000	1733.1	0.000
$p=2$	76.9	0.007	339.3	0.000	709.9	0.000
$p=3$	62.1	0.098	204.6	0.000	540.6	0.000
$p=4$	59.6	0.143	183.8	0.000	474.3	0.000
$p=5$	58.8	0.160	183.4	0.000	449.4	0.000
$p=6$	58.5	0.166	187.7	0.000	439.1	0.000
$p=7$	58.5	0.167	192.9	0.000	434.4	0.000
$p=8$	58.5	0.166	197.8	0.000	432.2	0.000

Table 7. ANOVA example for case 1 $\{y_{ij}=\ln(x_{ij})\}$

Source	SSQ	d.f.	Mean square	F	P
Between bidders	1.1312	91	0.0124	5.263	0.000
Within bidders	0.4134	175	0.0024		
Total	1.5446	266	0.0058		

and, therefore, inappropriately tested by the F test), these results clearly indicate that the bidders cannot be modelled as bidding from a distribution with the same location parameter.

Table 8. ANOVA results for all cases

	Cases					
	1		2		3	
Transformation	F	P	F	P	F	P
$y_{ij}=\ln(x_{ij}-m\bar{x}_j)$						
$m=0.0$	5.26	0.000	4.54	0.000	2.59	0.000
$m=0.1$	5.29	0.000	4.54	0.000	2.59	0.000
$m=0.2$	5.35	0.000	4.53	0.000	2.58	0.000
$m=0.3$	5.42	0.000	4.51	0.000	2.58	0.000
$m=0.4$	5.57	0.000	4.48	0.000	2.58	0.000
$m=0.5$	5.67	0.000	4.40	0.000	2.08	0.000
$m=0.6$	6.30	0.000	—	—	—	—
$y_{ij}=\ln(x_{ij}-mx_j)$						
$m=0.0$	5.26	0.000	4.54	0.000	2.59	0.000
$m=0.1$	5.20	0.000	4.55	0.000	2.60	0.000
$m=0.2$	5.13	0.000	4.57	0.000	2.61	0.000
$m=0.3$	5.04	0.000	4.48	0.000	2.62	0.000
$m=0.4$	4.93	0.000	4.59	0.000	2.63	0.000
$m=0.5$	4.80	0.000	4.61	0.000	2.65	0.000
$m=0.6$	4.62	0.000	4.62	0.000	2.67	0.000
$m=0.7$	4.40	0.000	4.63	0.000	2.69	0.000
$m=0.8$	4.08	0.000	4.60	0.000	2.71	0.000
$m=0.9$	3.58	0.000	4.48	0.000	2.71	0.000
$m=0.95$	3.18	0.000	4.27	0.000	2.70	0.000
$m=0.99$	2.57	0.000	3.68	0.000	2.57	0.000
$m=0.999$	2.58	0.000	3.12	0.000	2.33	0.000
$y_{ij}=x_{ij}^{1/p}$						
$p=1$	2.58	0.000	1.93	0.000	1.11	0.104
$p=2$	3.70	0.000	3.32	0.000	1.87	0.000
$p=3$	4.23	0.000	3.87	0.000	2.16	0.000
$p=4$	4.50	0.000	4.10	0.000	2.29	0.000
$p=5$	4.66	0.000	4.22	0.000	2.37	0.000
$p=6$	4.77	0.000	4.30	0.000	2.41	0.000
$p=7$	4.84	0.000	4.34	0.000	2.44	0.000
$p=8$	4.90	0.000	4.38	0.000	2.46	0.000

Summary and conclusions

This paper describes an empirical analysis of three sets of data aimed at deriving a suitable basic statistical model to represent the behaviour of construction contract bidders. The assumption that, once contract size is eliminated, all bidders draw bids from an identical probability distribution, offers a convenient starting point for model building.

In designing a procedure to test this assumption, the first problem encountered is that the popular method of eliminating contract size, by taking the ratio of bids to some measure of contract size (mean bid, lowest bid, designer's estimate, etc.), presents rather extreme and generally unrecognized difficulties in the analysis and interpretation of results. In proposing a new model to overcome this problem, it is shown that parameter estimation essentially reduces to the standard regression procedure, but that the sparsity of the resulting matrix system produces results severely distorted by computational rounding errors. Having solved this second problem by including a contract size datum parameter and substituting an iterative procedure for the regression method, the required scale and location parameters for individual bidders were successfully estimated on the assumption of a common Normal distribution.

Following previous contract-based analyses, the pooled residuals were then examined for compliance with the Normal assumption. This assumption was not found to be appropriate, the distributions for all three data sets each being similarly heavy-tailed corresponding to the Gram-Charlier Type A distribution $\mu_3=0.16$ and $\mu_4=4.5$. Heavy-tailed empirical distributions of these kind are, however, known to occur when samples of common shape but different scale parameters are pooled. In overcoming this final problem, the analysis of the unpooled residuals, a method similar to McCaffer and Pettitt's (1976) analysis of the distribution of contract bids, was used by testing the uniformity of the distribution of the approximate probability of each of five shape test statistics obtained for each bidder. By recording the test fails over a variety of log and power transformations, it was established that a three-parameter Log-normal model (equivalent to the transformation $y_{ij}=\ln(x_{ij}-mx_j)$ for $0.5<m<0.9$ to Normality, where x_j is the lowest bid for the contract) was appropriate for all three data sets.

Finally, the application of Bartlett's test and analysis of variance to the data indicated that both the common variance and common means assumption are generally invalid.

The general conclusions of this research are that the same three-parameter Log-normal distribution within the range specified above may be reasonably used to model the behaviour of all the bidders for these data. Estimates of the scale and location parameters, however, must be made separately for each bidder, the iterative method described in this paper providing the necessary technique for the task.

Acknowledgements

The author is indebted to Mr Roy Thomas of the Department of Civil Engineering and Drs Ernest Wilde and Mike Patefield of the Department of Mathematics, University of Salford, for their kind assistance in this research.

References

Afifi, A.A. and Elashoff, R.M. (1966). Missing observations in multivariate statistics: 1. Review of the literature. *Journal of the American Statistical Association*, **61**, 595–606.

AICBOR, Associated Industrial Consultants Limited and Business Operations Research Limited (1967). *Report of the Joint Consulting Team on Serial Contracting for Road Construction*. Ministry of Transport, London.

Aitchison, J. and Brown, J.A.C. (1957). *The Log-normal Distribution*. Cambridge University Press, Cambridge.

Alexander, A.B. (1970). *What Price Estimating Accuracy?* Paper No. 534, Metal Fabricating Institute Inc., Rockford, Illinois.

Ali, M.A. and Giacotto, C. (1982). The identical distribution hypothesis for stock market prices – location and shift alternatives. *Journal of the American Statistical Association*, **77**, 19–28.

Barnes, N.M.L. (1971). The design and use of experimental bills of quantities for civil engineering contracts. PhD thesis, University of Manchester Institute of Science and Technology.

Beeston, D.T. (1974). *One Statistician's View of Estimating*. Cost Study 3, July, RICS Building Cost Information Service, London.

Beeston, D.T. (1982). Estimating market variance. In *Building Cost Techniques: New Directions* (edited by P.S. Brandon), pp. 265–77. E. and F.N. Spon, London.

Brown, D.M. (1974). A comparative study of four building organisations in relation to their environment. MSc thesis, University of Bath.

Capen, E.C., Clapp, R.V. and Campbell, W.M. (1971). Competitive bidding in high-risk situations. *Journal of Petroleum Technology*, June, pp. 641–53.

Cauwelaert, F.V. and Heynig, E. (1978). Correction of bidding errors: The Belgian solution. *Journal of the Construction Division, ASCE*, **105**(CO1), 13–23.

Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, **41**, 135–56.

Crawford, P.B. (1970). Texas offshore bidding patterns. *Journal of Petroleum Technology*, **22**, 283–9.

Elman, H.C. (1982). Iterative methods for large sparse nonsymmetric systems of linear equations. PhD thesis, Yale University.

Emond, L.J. (1971). Analytical strategy for the competitive price setter. *Cost and Management*, September–October, pp. 6–11.

Fine, B. and Hackemar, G. (1970). Estimating and bidding strategy. *Building Technology and Management*, **8** (9), pp. 8–9.

Flanagan, R. and Norman, G. (1985). Sealed bid auctions: An application to the building industry. *Construction Management and Economics*, **3**, 145–61.

Friedman, L. (1956). A competitive bidding strategy. *Operations Research*, **4**, 104–112.

Geary, R.C. (1935). The ratio of the mean deviation as a test of normality. *Biometrika*, **27**, 310–32.

Geary, R. C. (1936). Moments of the ratio of the mean deviation to the standard deviation for normal samples. *Biometrika*, **28**, 298–307.

Grinyer, P.H. and Whittaker, J.D. (1973). Managerial judgement in a competitive bidding model. *Operational Research Quarterly*, **24**(2), 181–91.

Harvey, J.R. (1979). Competitive bidding on Canadian public construction contracts, stochastic analysis for optimization. PhD thesis, School of Business Administration, University of Western Ontario.

Hocking, R.R. and Smith, W.B. (1968). Estimation of parameters in the multivariate normal distribution with missing observations. *Journal of the American Statistical Association*, **63**, 159–73.

Hossein, B.R. (1977). Risk analysis of tendering policies for capital projects. PhD thesis, University of Bradford.

Hsu, D.A., Miller, R.B. and Wichern, D.W. (1974). On the stable Paretian behaviour of stock market prices. *Journal of the American Statistical Association*, **69**, 108–113.

Johnston, R.H. (1978). *Optimisation of the Selective Competitive Tendering System by the Construction Client*. Transport and Road Laboratory Report No. 855 DOG.

Kendall, M.G. and Stuart, A. (1963). *The Advanced Theory of Statistics, Vol. 1: Distribution Theory*. Charles Griffin, London.

Klein, J.D. (1976). Joint ventures and bidding for offshore oil. In *Bidding and Auctioning for Procurement and Allocation* (edited by Y. Amihud), pp. 38–42. New York University Press, New York.

McCaffer, R. (1976). Contractors' bidding behaviour and tender price prediction. PhD thesis, Loughborough University of Technology.

McCaffer, R. and Pettitt, A.N. (1976). Distribution of bids for buildings and road contracts. *Operational Research Quarterly*, **27**(4), 835–43.

Mitchell, M.S. (1977). The probability of being the lowest bidder. *Applied Statistics*, **2**(2), 191–4.

Morrison, N. (1984). The accuracy of quantity surveyors' cost estimating. *Construction Management and Economics*, **2**, 57–75.

Morrison, N. and Stevens, S. (1980). *Construction Cost Data Base*. Second Annual Report of Research Project by Department of Construction Management, University of Reading, for Property Services Agency, Directorate of Quantity Surveying Services, Department of the Environment.

Oren, S.S. and Rothkopf, M.H. (1975). Optimal bidding in sequential auctions. *Operations Research*, **23**(6), 1080–90.

Park, W.R. (1966). *The Strategy of Contracting for Profit*. Prentice-Hall, Englewood Cliffs, N.J.

Pearson, E.S. (1963). Some problems arising in approximating to probability distributions, using moments. *Biometrika*, **50**(42), 95–111.

Pelto, C.R. (1971). The statistical structure of bidding for oil and mineral rights. *Journal of the American Statistical Association*, **66**, 456–60.

Pettitt, A.N. (1975). *Small Sample Percentage Points of the Anderson-Darling Statistic for Testing Normality*. Math. Res. No. 71, Department of Mathematics, Loughborough University of Technology.

Runeson, G. (1988). An analysis of the accuracy of estimating and the distribution of tenders. *Construction Management and Economics*, **6**, 357–70.

Shaffer, L.R. and Micheau, T.W. (1971). Bidding with competitive strategy models. *Journal of the Construction Division, ASCE*, **97**(CO1), 113–26.

Shapiro, S.S., Wilk, B.B. and Chen, H.J. (1968). A comparative study of various tests of normality. *Journal of the American Statistical Association*, **63**, 1343–72.

Skitmore, R.M. (1981). Bidding dispersion – an investigation into a method of measuring the accuracy of building cost estimates. MSc thesis, University of Salford.

Skitmore, R.M. (1982). A bidding model. In *Building Cost Techniques: New Directions* (edited by P.S. Brandon), pp. 278–89. E. and F.N. Spon, London.

Skitmore, R.M. (1989). *Contract Bidding in Construction: Strategic Management and Modelling*. Longman, London.

Spiegelhalter, D.J. (1983). Diagnostic tests of distributional shape. *Biometrika*, **70**(2), 401–409.

Weverbergh, M. (1982). *Competitive Bidding: Estimating the Joint Distribution of Bids*. Working Paper No. 82–79, Centrum voor Bedrijfseconomie en Bedrijfseconometrie, University of Antwerp.

Whittaker, J.D. (1970). A study of competitive bidding with particular reference to the construction industry. PhD thesis, City University, London.

Wilson, O.D. and Sharpe, K. (1988). Tenders and estimates: A probabilistic model. *Construction Management and Economics*, **6**, 225–45.