Generalised gamma bidding model

A generalised gamma bidding model is presented, which incorporates many previous models. The log likelihood equations are provided. Using a new method of testing, variants of the model are fitted to some real data for construction contract auctions to find the best fitting models for groupings of bidders. The results are examined for simplifying assumptions, including all those in the main literature. These indicate no one model to be best for all datasets. However, some models do appear to perform significantly better than others and it is suggested that future research would benefit from a closer examination of these.

Keywords

Generalised gamma distribution, bidding models, grouping, least-squares cross-validation, goodness of fit.

1. Introduction

Much of the bidding literature is concerned with setting a mark-up, *m*, so that the probability, *Pr(m)*, of entering the winning bid reaches some desired level. Several composite density functions have been proposed for estimating this probability, the most frequent being the uniform (e.g., Vickrey, 1961; Fine and Hackemar, 1970; Cauwelaert and Heynig, 1978; Whittaker, 1970; and Grinyer and Whittaker, 1973), normal (e.g., Alexander, 1970; Emond, 1971; McCaffer, 1976; Mitchell, 1977; Cauwelaert and Heynig, 1978; Morrison and Stevens, 1980; Carr, 1983), lognormal (e.g., Arps, 1965; Brown, 1966; Crawford, 1970; Capen *et al*, 1971; Klein, 1976; Weverbergh, 1982; Skitmore and Pemberton, 1994); gamma (Friedman, 1956; Dougherty and Nozaki, 1975); and Weibull (Gates, 1967[1]; Oren and Rothkopf, 1975) - and with a variety of simplifying assumptions concerning parameter estimation (e.g., Friedman 1956, Carr 1982; Skitmore and Pemberton 1994).

Of the few empirical comparative studies made, Hossein (1977) claimed the gamma distribution to be the best fit, followed by lognormal and normal distributions, while McCaffer and Pettitt (1976) considered the normal distribution to be a better fit than uniform for their data. More recently, Skitmore (1991) reported a lognormal distribution to be better than uniform with Skitmore (2001, 2002b, 2004) and Skitmore and Lo (2002) finding the truncated lognormal to be the best fit followed by the truncated normal distribution, with the uniform distribution some way behind.

Fitting probability distributions to bidding data is not an easy task. The special features of construction contract[2] auctions, for example, are (1) each contract is different in size, (2) there are a small number of bidders in each auction and (3) different bidders bid in each auction (making the contract-bidder matrix usually well over 90% sparse). To date, only Skitmore (1991) has succeeded in using multivariable methods. As discussed by Skitmore (1991), the main difficulty resides in the extreme sparseness and asymmetry of the contract-bidder matrix, quoting Elman (1982) who, in considering the use of direct and iterative methods of solving large sparse non-symmetric systems of linear equations, found difficulties with direct methods due to the factoring process involved 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, prompting the general conclusion 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'. Faced with similar problems, Skitmore (1991) resorted to an iterative method that has since become established in the field (e.g., Skitmore and Pemberton, 1994; Skitmore *et al*, 2007). It should also be noted that the sparseness of the matrix increases as the dataset enlarges over time and new bidders are added.

A further difficulty until recently has been the lack of available methods for testing the suitability of the models and simplifying assumptions. This has been resolved to some extent by Skitmore (2001, 2002b) and Skitmore and Lo (2002), who developed both graphical and statistical methods. Skitmore (2002a, 2004) has also developed methods based on Dowe *et al*'s (1996) log score measure.

In this paper, an alternative approach is adopted in using a more general model in which the various distribution forms are represented as special cases depending on the value of certain parameters in the model. Of particular interest is Stacey's (1962) generalised gamma (GG) distribution, as this includes many of the common bidding distributions in the literature, such as the lognormal, Weibull and gamma, as special cases. Coincidentally, these are also the most commonly used in the statistical analysis of lifetimes (Lawless, 1982), where it is often convenient to move between these and their equivalent log forms – normal, extreme value and log-gamma – for parameter estimation and inference procedures in general. Skitmore's (1991) method of solving the log likelihood equations iteratively is followed to remove the contract size effect and a method is developed for grouping similar bidders involving a new method of testing the fit of potential grouping arrangements. An example application is provided by the analysis of some real data for construction contract auctions to find the best groupings of bidders, and the results are examined for a wide range of simplifying assumptions for parameter estimation.

2. GG model for bidding

Let $X_1, X_2, \ldots, X_k$ be independently distributed random variables, and suppose we generate one value, i.e., $x_1, x_2, \ldots, x_k$ from each variable. Letting $f(.)$ denote the probability density function (pdf) and $F(.)$ the distribution function (cdf) of X, then choosing a value, say x_i, the probability of this being less than one other value x_j $(j \neq i)$ we denote as P_{ij}. Assuming independence, the probability of x_i being the lowest value of all the x values is then

$$P_{i.} = \int_{-\infty}^{\infty} \omega_i f_i(x) S_i(x)^{\omega_i - 1} \prod_{j \neq i}^{k} S_j(x)^{\omega_j} dx \tag{1}$$

where $S_j(.)$ is the survival function, $1-F_j(.)$, of each of the remaining $X_{j\neq i}$ variables. This includes all the major bidding models as special cases when $\omega_i = \omega_j = 1$ (e.g., Friedman, 1956); Carr, 1982; Skitmore and Pemberton, 1994) and $\omega_j = \frac{P_{ji}}{P_{ij}}$ (Gates, 1967) model[3].

Stacey's (1962) GG distribution represents the sum of n-exponential distributed random variables and is defined in terms of its nonnegative scale and shape parameters (Shin *et al*, 2005). It has been used in many situations, with recent applications ranging from the analysis of drought data (Nadarajah and Gupta, 2007), speech recognition (Babu *et al*, 2012) and crab catches (Hvingel *et al*, 2012), to breast cancer survival analysis (Abadi *et al*, 2012) and estimating pregnancy times (Keiding *et al*, 2012**).**

For the GG family, the pdf and cdf involve the gamma function, $\Gamma(x)$, and incomplete gamma function, $I[\kappa, \lambda x]$, i.e.:

$$f(x) = \frac{\lambda\chi}{\Gamma(\kappa)}(\lambda x)^{\kappa\chi-1}\exp\left[-(\lambda x)^{\chi}\right] \qquad x>0 \tag{2a}$$

and

$$S(x) = 1 - I\left[\kappa, (\lambda x)^{\chi}\right] \qquad x>0 \tag{2b}$$

so that from (1)

$$P_{i.} = \int \frac{\lambda_i\chi_i}{\Gamma(\kappa_i)}(\lambda_i x)^{\kappa_i\chi_i-1}\exp\left[-(\lambda_i x)^{\chi_i}\right]\prod_{j\neq i}^{k}\left\{1 - I\left[\kappa_j, (\lambda_j x)^{\chi_j}\right]\right\}dx \tag{3}$$

where $\chi \rangle 0$, $\lambda \rangle 0$ and $\kappa \rangle 0$ are parameters, ω in (1) being subsumed within the λ parameter. This includes as special cases the exponential $(\chi = \kappa = 1)$, Weibull $(\kappa = 1)$, and gamma $(\chi = 1)$, with the lognormal distribution arising as a limiting form as $\kappa \to \infty$. Happily, therefore, the GG model includes many of the most commonly proposed bidding distributions as special cases.

3. Parameter estimation

3.1 Generalised gamma

To estimate the parameters, it is convenient to first consider Y=log$X$ instead of X and reparameterise (χ, λ, κ) to (α, σ, κ) in (3) by setting $\alpha = \log\lambda^{-1} + \chi\log\kappa$ and $\sigma = \chi^{-1}\kappa^{-1/2}$ after Prentice (1974). Then, by introducing subscripts *j* and *l* to denote bidder *j* and contract *l* (*j*=1,…,*r*; *l*=1,…,*c*), Skitmore's (1991) approach is followed where a log bid, y_{jl} is modelled by

$$y_{jl} = \alpha_j + \beta_l + \varepsilon_{jl}$$

where $\alpha_j = \log \lambda_j^{-1} + \chi_j \log \kappa_j$ is interpreted as a bidder location parameter, β_l a contract datum parameter and ε_{jl} is distributed according to p.d.f. $f_0\left(0, \sigma_j, \kappa_j\right)$. Although this is clearly a two way ANOVA regression problem, the major difficulty in applying direct methods of estimation, as discussed earlier, is that the bidder/auction matrix in the construction contract auction context is asymmetric and extremely sparse - usually over 90% so[4] - resulting in the need for a special maximum log likelihood procedure as follows.

Setting $z_{jl} = \dfrac{y_{jl} - \alpha_j - \beta_l}{\sigma}$, y_{jl} has the p.d.f.

$$\frac{\kappa_j^{\kappa - 1/2}}{\sigma_j \Gamma\left(\kappa_j\right)} \exp\left[\sqrt{\kappa_j}\, z_{jl} - \kappa_j \exp\left(\frac{z_{jl}}{\sqrt{\kappa_j}}\right)\right] \qquad -\infty < y_{jl} < \infty \tag{4}$$

So the log-likelihood is:

$$\log L = N\left(\kappa_j - 1/2\right)\log \kappa_j - N \log \Gamma\left(\kappa_j\right) - N \log \sigma_j + \sqrt{\kappa_j} \sum_j^r \sum_l^c \delta_{jl} z_{jl} - \kappa_j \sum_j^r \sum_l^c \delta_{jl} \frac{z_{jl}}{\sqrt{\kappa_j}}$$

where $\delta_{jl} = 1$ if bidder *j* bids for contract *l*

= 0 if bidder *j* does not bid for contract *l*

$N = \sum_l^c \sum_j^r \delta_{jl}$ = total number of bids

The maximum likelihood equations over the α's, β's and σ's are:

$$\frac{\partial LogL}{\partial \alpha_j} = -\frac{N\sqrt{\kappa_j}}{\sigma_j} + \frac{\sqrt{\kappa_j}}{\sigma_j} \sum_l^c \delta_{lj} \exp\left(\frac{z_{jl}}{\sqrt{\kappa_j}}\right)$$

$$\therefore e^{\hat{\alpha}_j} = \left[\frac{1}{n_j} \sum \delta_{lj} \exp\left(\frac{y_{lj} - \beta_l}{\hat{\sigma}_j \sqrt{\kappa_j}}\right)\right]^{\hat{\sigma}_j \sqrt{\kappa_j}} \tag{5}$$

$$\frac{\partial LogL}{\partial \beta_l} = -\frac{N}{\sigma_j} + \frac{1}{\sigma_j} \sum_j^r \delta_{lj} \exp\left(\frac{z_{jl}}{\sqrt{\kappa_j}}\right)$$

$$\therefore e^{\hat{\beta}_l} = \left[\frac{1}{m_l} \sum \delta_{lj} \exp\left(\frac{y_{lj} - \alpha_j}{\sigma_j \sqrt{\kappa_j}}\right)\right]^{\sigma_j \sqrt{\kappa_j}} \tag{6}$$

$$\frac{\partial LogL}{\partial \sigma_j} = -\frac{N}{\sqrt{\kappa_j}} - \sum\sum \delta_{lj} z_{lj} + \sum\sum \delta_{lj} z_{lj} \exp\left(\frac{z_{jl}}{\sqrt{\kappa_j}}\right)$$

$$\propto \sum\sum \delta_{lj} \left\{ z_{lj} \left[\exp\left(\frac{z_{jl}}{\sqrt{\kappa_j}}\right) - 1 \right] - \frac{1}{\sqrt{\kappa_j}} \right\} = 0 \quad (7)$$

where $n_j = \sum_l^c \delta_{jl}$, the number of bids made by bidder *j*, and $m_l = \sum_j^r \delta_{jl}$, the number of bids for contract *l*. Setting σ_j =1 and κ_j =1 and solving the α_j 's and β_l 's by iteration of (5) and (6) provides the required maximum likelihood estimates (m.l.e) for the exponential distribution. For the Weibull distribution, the parameter estimates can be obtained by setting κ_j =1 and solving the α_j 's and β_l 's by iteration of (5) and (6) for trial σ_j values - finding the best σ_j using the Newton-Raphson method[5] for (7). Similarly, the gamma parameters can be obtained by setting $\kappa_j = 1/\sigma_j^2$ and solving the α_j 's and β_j 's by iteration of (5) and (6) for trial σ_j values – again finding the best σ_j using the Newton-Raphson method for (7). On completion, the m.l.e's of the original parameters are then $\chi_j = \sigma_j^{-1} \kappa_j^{-1/2}$ and $\lambda_j = \exp(\alpha_j)^{-1}$.

3.2 Weighted lognormal

For the lognormal distribution, this method is impractical as setting κ to a very large value creates computational problems (e.g., for $\frac{1}{\sqrt{\kappa}}$). A better approach is therefore to work directly with the limiting form itself. Also, from (1), we include the weighting term ω.

Letting $Y = \log(X)$, $y_{jl} = \mu_j + \beta_l + \varepsilon_{jl}$ and ε_{jl} is distributed according to p.d.f. $f_0(0, \sigma_j)$. With $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ and $Q(z) = \int_z^\infty \phi(x) dx$, the log likelihood is

$$\log L(\mu_j, \sigma_j, \beta_l, \omega_j) = \sum_{l=1}^{c} \sum_{j=1}^{r} \delta_{lj} \left\{ \log \omega_j - \log \sigma_j - \frac{1}{2\sigma_j^2} (y_{lj} - \beta_{lj} - \mu_j)^2 + (\omega_j - 1) \log Q\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma_j} \right) \right\}$$

$$= \sum\sum \delta_{lj} \log \omega_j - \sum\sum \delta_{lj} \log \sigma_j - \sum\sum \delta_{lj} \frac{1}{2\sigma_j^2} (y_{lj} - \beta_{lj} - \mu_j)^2 + \sum\sum \delta_{lj} (\omega_j - 1) \log Q\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma_j} \right)$$

The first derivatives of Log L are

$$\frac{\partial LogL}{\partial \mu_j} = \sum_l \delta_{lj} \frac{1}{\sigma_j^2}\left(y_{lj} - \beta_l - \mu_j\right) + \sum_l \delta_{lj} \frac{1}{\sigma_j}\left(\omega_j - 1\right)\phi\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma_j}\right) \Big/ Q\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma_j}\right)$$

$$\frac{\partial LogL}{\partial \sigma_j} = -\sum_l \delta_{lj} \frac{1}{\sigma_j} + \sum_l \delta_{lj} \frac{1}{\sigma^3}\left(y_{lj} - \beta_l - \mu_j\right)^2 + \sum_l \delta_{lj} \frac{1}{\sigma}\left(\omega_j - 1\right)\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma_j}\right)\phi\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma}\right) \Big/ Q\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma}\right)$$

$$\frac{\partial LogL}{\partial \beta_l} = \sum_j \delta_{lj} \frac{1}{\sigma_j^2}\left(y_{lj} - \beta_l - \mu_j\right) + \sum_j \delta_{lj} \frac{1}{\sigma_j}\left(\omega_j - 1\right)\phi\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma_j}\right) \Big/ Q\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma_j}\right)$$

$$\frac{\partial \log L}{\partial \omega_j} = \sum_l \delta_{lj} \frac{1}{\omega_j} + \sum_l \delta_{lj} \log Q\left(\frac{y_{lj} - \beta_l - \mu_j}{\sigma_j}\right)$$

Putting $V(z) = \frac{\phi(z)}{Q(z)}$ and $z_{lj} = \left(y_{lj} - \beta_l - \mu_j\right)/\sigma_j$

Then the maximum likelihood equations for μ, σ, β and ω are

$$\frac{\partial LogL}{\partial \mu_j} = \sum_l \delta_{lj} z_{lj} + \sum_l \delta_{lj}\left(\omega_j - 1\right)V\left(z_{lj}\right) = 0 \tag{8}$$

$$\frac{\partial LogL}{\partial \sigma_j} = -\sum_l \delta_{lj} + \sum_l \delta_{lj} z_{lj}^2 + \sum_l \delta_{lj}\left(\omega_j - 1\right)z_{lj} V\left(z_{lj}\right) = 0 \tag{9}$$

$$\frac{\partial LogL}{\partial \beta_l} = \sum_j \delta_{lj} z_{lj} + \sum_j \delta_{lj}\left(\omega_j - 1\right)V\left(z_{lj}\right) = 0 \tag{10}$$

$$\frac{\partial LogL}{\partial \omega_j} = \sum_l \delta_{lj} \frac{1}{\omega_j} + \sum_l \delta_{lj} \log Q(z_{lj}) = 0 \Rightarrow \omega_j = \frac{n_j}{\sum_l^c \delta_{lj} \log Q\left(z_{lj}\right)} \tag{11}$$

One method of solution is to use the Newton-Raphson method to solve (8), (9) and (10) for μ_j, σ_j and β_l, using (11) to calculate ω_j from the trial μ_j, σ_j and β_l values on each iteration. Upon convergence, the estimates of σ_j^2 are biased, but can be adjusted by multiplying by the approximation (Skitmore 1991)

$$\left[\frac{n_j}{\left(n_j - 1\right)\left(1 - \frac{c-1}{N-r}\right)}\right] \tag{12}$$

4. Scoring function

4.1 Error term

To date, the only approach to quantifying the accuracy of bidding models has been to examine their predictions of the probability of a bidder winning a contact, by means of a modified form of the Dowe *et al* (1996) logscore (Skitmore, 2001). The method suffers from discontinuities, however, which makes it difficult to use as a grouping criterion. To overcome this problem, a new method of scoring is used. This is the mean square error (*MSE*) of predicting the lowest bid. The MSE is chosen as it is symmetrically very responsive to extreme errors, which is a big advantage for construction contract bidders, where underestimates of the actual low bid leads to bidding too low and hence to less than optimal profits or even losses, while overestimates leads to bidding too high and obtaining less work. Either way, both can be fatal financially to bidders. What is needed is a model that predicts competitors' bids equally close, whether over or under, and disproportionately heavily penalises those which produce extreme errors either way.

Let $x_{(1)i*l}$ denote the lowest bid for auction l *excluding* bidder i (that is, the lowest bid of bidder i's competitors) then the auction *MSE* is given by

$$MSE_l = \frac{1}{m_l} \sum_{i*=1}^{r} \delta_{il} \left(x_{(1)i*l} - \hat{x}_{(1)i*l} \right)^2 \tag{13}$$

where $\hat{x}_{(1)i*l}$ is the value of $x_{(1)i*l}$ estimated by the model.

7.2Estimation of $\hat{x}_{(1)i*l}$

The best estimate of $x_{(1)i*l}$ is provided by the expected value of $x_{(1)i*l}$. The p.d.f is

$$f\left(x_{(1)i*l}\right) = \sum_{i \neq i*}^{r} \delta_{il} \omega_i f_i(x) S_i(x)^{\omega_i - 1} \prod_{j \neq i, i*}^{r} \left[S_j(x)^{\omega_j} \right]^{\delta_{jl}} dx \tag{14}$$

so the expectation is

$$\hat{x}_{(1)i*l} = \int_{-\infty}^{\infty} x \sum_{i \neq i*}^{r} \delta_{il} \omega_i f_i(x) S_i(x)^{\omega_i - 1} \prod_{j \neq i, i*}^{r} \left[S_j(x)^{\omega_j} \right]^{\delta_{jl}} dx \tag{15}$$

Again using the transformation $y = \log(x)$, from (15) the expectation of the lowest remaining bid for the *GG distribution* is

$$\hat{y}_{(1)i*l} = \int_{-\infty}^{\infty} y \sum_{i \neq i*}^{r} \delta_{il} \frac{\kappa_i^{\kappa_i - 1/2}}{\sigma_i \Gamma(\kappa_i)} \exp\left(\sqrt{\kappa_i} \frac{y - \mu_i}{\sigma_i} - \kappa_i e^{(y-\mu_i)/\sigma_i \sqrt{\kappa_i}} \right) \prod_{j \neq i, i*}^{r} \left[Q\left(\kappa_j, \kappa_j^{\frac{y-\mu_j}{\sigma_j}} \right) \right]^{\delta_{jl}} dy \tag{16}$$

and for the *lognormal distribution*,

$$\hat{y}_{(1)i^*l} = \int_{-\infty}^{\infty} y \sum_{i \neq i^*}^{r} \delta_{il} \omega_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y-\mu_i)^2}{2\sigma_i^2} \right) S_i(y)^{\omega_i - 1} \prod_{j \neq i, i^*}^{r} \left[S_j \left(\frac{\sigma_i y + \mu_i - \mu_j}{\sigma_j} \right)^{\omega_j} \right]^{\delta_{jl}} dy \quad (17)$$

so that now

$$MSE_l = \frac{1}{m_l} \sum_{i^*=1}^{r} \delta_{il} \left(y_{(1)i^*l} - \hat{y}_{(1)i^*l} \right)^2 \quad (18)$$

5. Simplifying assumptions

5.1 Fixing parameters

As mentioned above, parameter estimation difficulties are expected. For a typical dataset comprising 400 auctions and 400 different bidders, there 1600 parameters to estimate. Of course, this problem can be alleviated by fixing some of the parameters to be equal. For example, we could assume that $\mu_1 = \mu_2 = \ldots =$ etc. This would reduce the number of parameters to be estimated by 400 in the above example. This is not a new approach in this field. Table 1 summarises the most obvious ways of doing this, highlighting the main simplifications made to date, the lognormal being shown separately for ease.

5.2 Pooling

An advancement on this is to pool the data also. This can be done empirically by means of an algorithmic selection procedure. As this is a forecasting situation, to do this meaningfully involves using the out sample error as the grouping criterion. The procedure used was as follows. First, set up a number of groups g=1,2, …, G and then to assign each bidder into a group. For pooling purposes, at this stage all the bidders in each group are assumed to be iid. Several operations are then involved. These are described below.

Model	Code	λ ($\lambda_0 = 1$)	χ	κ
Standard exponential	GG1	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 = \chi_2 = \ldots = 1$	$\kappa_1 = \kappa_2 = \ldots = 1$
Exponential	GG2	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 = \chi_2 = \ldots = 1$	$\kappa_1 = \kappa_2 = \ldots = 1$
1 param Weibull (1)	GG3	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 = \chi_2 = \ldots$	$\kappa_1 = \kappa_2 = \ldots = 1$
1 param Weibull (2)	GG4	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 \neq \chi_2 \neq \ldots$	$\kappa_1 = \kappa_2 = \ldots = 1$
2 param Weibull (1) (Gates 1967; Rothkopf 1969)	GG5	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 = \chi_2 = \ldots$	$\kappa_1 = \kappa_2 = \ldots = 1$
2 param Weibull (2)	GG6	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 \neq \chi_2 \neq \ldots$	$\kappa_1 = \kappa_2 = \ldots = 1$
1 param Gamma (1)	GG7	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 = \chi_2 = \ldots = 1$	$\kappa_1 = \kappa_2 = \ldots$
1 param Gamma (2)	GG8	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 = \chi_2 = \ldots = 1$	$\kappa_1 \neq \kappa_2 \neq \ldots$
2 param Gamma (1)	GG9	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 = \chi_2 = \ldots = 1$	$\kappa_1 = \kappa_2 = \ldots$
2 param Gamma (2) (Weverbergh 1982)	GG10	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 = \chi_2 = \ldots = 1$	$\kappa_1 \neq \kappa_2 \neq \ldots$
2 param GG (1)	GG11	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 = \chi_2 = \ldots$	$\kappa_1 = \kappa_2 = \ldots$
2 param GG (2)	GG12	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 = \chi_2 = \ldots$	$\kappa_1 \neq \kappa_2 \neq \ldots$
2 param GG (3)	GG13	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 \neq \chi_2 \neq \ldots$	$\kappa_1 = \kappa_2 = \ldots$
2 param GG (4)	GG14	$\lambda_1 = \lambda_2 = \ldots = 1$	$\chi_1 \neq \chi_2 \neq \ldots$	$\kappa_1 \neq \kappa_2 \neq \ldots$
3 param GG (1)	GG15	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 = \chi_2 = \ldots$	$\kappa_1 = \kappa_2 = \ldots$
3 param GG (2)	GG16	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 = \chi_2 = \ldots$	$\kappa_1 \neq \kappa_2 \neq \ldots$
3 param GG (3)	GG17	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 \neq \chi_2 \neq \ldots$	$\kappa_1 = \kappa_2 = \ldots$
3 param GG (4)	GG18	$\lambda_1 \neq \lambda_2 \neq \ldots$	$\chi_1 \neq \chi_2 \neq \ldots$	$\kappa_1 \neq \kappa_2 \neq \ldots$
Log Normal		μ ($\mu_0 = 0$)	σ	ω ($\omega_0 = 1$)
Standard	LN1	$\mu_1 = \mu_2 = \ldots = 0$	$\sigma_1 = \sigma_2 = \ldots = 1$	$\omega_1 = \omega_2 = \ldots = 1$
	LN2	$\mu_1 \neq \mu_2 \neq \ldots$	$\sigma_1 = \sigma_2 = \ldots = 1$	$\omega_1 = \omega_2 = \ldots = 1$
	LN3	$\mu_1 = \mu_2 = \ldots = 0$	$\sigma_1 = \sigma_2 = \ldots$	$\omega_1 = \omega_2 = \ldots = 1$
	LN4	$\mu_1 = \mu_2 = \ldots = 0$	$\sigma_1 \neq \sigma_2 \neq \ldots$	$\omega_1 = \omega_2 = \ldots = 1$
Carr (1983)	LN5	$\mu_1 \neq \mu_2 \neq \ldots$	$\sigma_1 = \sigma_2 = \ldots$	$\omega_1 = \omega_2 = \ldots = 1$
2 param lognormal (Weverbergh, 1982; Skitmore, 1991)	LN6	$\mu_1 \neq \mu_2 \neq \ldots$	$\sigma_1 \neq \sigma_2 \neq \ldots$	$\omega_1 = \omega_2 = \ldots = 1$
	LN8	$\mu_1 = \mu_2 = \ldots = 0$	$\sigma_1 = \sigma_2 = \ldots = 1$	$\omega_1 \neq \omega_2 \neq \ldots$
	LN10	$\mu_1 \neq \mu_2 \neq \ldots$	$\sigma_1 = \sigma_2 = \ldots = 1$	$\omega_1 \neq \omega_2 \neq \ldots$
	LN12	$\mu_1 = \mu_2 = \ldots = 0$	$\sigma_1 = \sigma_2 = \ldots$	$\omega_1 \neq \omega_2 \neq \ldots$
	LN14	$\mu_1 = \mu_2 = \ldots = 0$	$\sigma_1 \neq \sigma_2 \neq \ldots$	$\omega_1 \neq \omega_2 \neq \ldots$
	LN16	$\mu_1 \neq \mu_2 \neq \ldots$	$\sigma_1 = \sigma_2 = \ldots$	$\omega_1 \neq \omega_2 \neq \ldots$

Table 1: Summary of models

1. Calculate group MSE_g

Step 1. Set contract number *l*=1

Step 2. Remove contract *l*, estimate the group *g*'s bidders' parameters using the above maximum likelihood formula in conjunction with the remaining contracts and bidders. Being iid, this means that all the bidders in group *g* will have the same parameters

Step 3. Insert the parameters obtained in step 2 into (18) to calculate MSE_l

Step 4. For $l = 2,\ldots,c$, reinstate contract *l*-1, leave out contract *l* and repeat steps 2 to 3

Step 5. Set the group $MSE_g = \sum_l^c MSE_l$

The total MSE is $MSE_{total} = \sum_{g=1}^{G} MSE_g$. The aim now is to find an allocation of bidders to groups that produces the lowest MSE_{total}. First, we start with all the bidders placed in group 1 (*G*=1) and then place bidders from group 1 into group 2 (G=2) until there is no further improvement in MSE, at which point bidders are placed in group 3 (G=3), etc until no further improvement in MSE is possible. The procedures for G=1 and *G*>1 follow.

2. Allocation of bidders, G=1

Set *g*=1 and calculate $MSE_{g=1}$ and hence MSE_{total} as above. This is the *baseline* MSE_{total}. Set $\min MSE_{total}$ = *baseline* MSE_{total}

3. *Allocation of bidders, G>1*

Step 1. Set *g*=1 and bidder *b*=1

Step 2. Remove bidder *b* from group *g* and place in group *G*

Step 3. Calculate $MSE_1, MSE_2, \ldots$ as above and hence $MSE_{total(b)} = \sum_{g=1}^{G} MSE_g$ and reinstate *b* into group *g*

Step 4. Repeat steps 2 and 3 for *b*=2, …

Step 5. Repeat steps 2 and 3 for *g*=2,…,*G* and *b*=1, …

Step 6. Find the lowest $MSE_{total(b)}$. If $MSE_{total(b)} < \min MSE_{total}$ remove the bidder *b*, for which $MSE_{total(b)}$ is minimum, place in group *G* and reset $\min MSE_{total} = MSE_{total(b)}$

Step 7. Repeat Steps 1 to 6 immediately above

Step 8. Repeat Step 7 until no more $MSE_{total(b)} < \min MSE_{total}$ can be found.

Upon reaching the point where no more $MSE_{total(b)} < \min MSE_{total}$ can be found for a new *G* value, the procedure then checks if any more single transfers between groups will improve $\min MSE_{total}$ before terminating.

Of course, the method described here is only the one used in the research and is modelled on the forward stepwise method of variable selection in regression analysis. Many other grouping algorithms are possible.

	Case 1			Case 2			Case 3			Case 4			Rank	
Model	**Baseline MSE**	**Final MSE**	**Rank**	**Baseline MSE**	**Final MSE**	**Rank**	**Baseline MSE**	**Final MSE**	**Rank**	**Baseline MSE**	**Final MSE**	**Rank**	**Sum**	**Over all**
GG1	4.632203	4.632203	28	3.788572	3.788572	29	3.684643	3.684643	29	7.971837	7.971837	29	115	29
GG2	4.632203	4.632203	29	3.788572	3.782243	28	3.684643	3.683275	28	7.971837	7.971675	28	113	28
GG3	0.009378	0.009378	20	0.042495	0.042495	17	0.011708	0.011708	15	0.069402	0.069402	22	74	19
GG4	0.009378	0.009378	21	0.042495	0.042495	18	0.011708	0.011708	16	0.069402	0.069402	23	78	21
GG5	0.009378	0.003773	3=	0.042495	0.02581	3	0.011708	0.007665	3	0.069402	0.051474	20	29	6
GG6	0.009378	0.003773	3=	0.042495	0.034574	10	0.011708	0.007623	2	0.069402	0.053313	21	36	10
GG7	0.009889	0.009889	22	0.048841	0.048841	23	0.013091	0.013091	23	0.048954	0.048954	19	87	23
GG8	0.009889	0.008499	17	0.048841	0.04749	21	0.013091	0.012962	21	0.048954	0.047565	17	76	20
GG9	0.009889	0.003763	2	0.048841	0.027391	6	0.013091	0.007755	4	0.048954	0.021697	3	15	2
GG10	0.009889	0.003573	1	0.048841	0.027119	5	0.013091	0.009111	8	0.048954	0.021639	2	16	3
GG11	0.009889	0.009889	23	0.048828	0.048828	22	0.013062	0.013062	22	0.048399	0.048399	18	85	22
GG12	0.009889	0.008409	15	0.048828	0.039107	14	0.013062	0.012086	17	0.048399	0.032257	12	58	14
GG13	0.009889	0.009182	19	0.048828	0.046383	20	0.009889	0.009182	10	0.048399	0.042961	15	64	16
GG14	0.009889	0.008409	16	0.048828	0.039107	15	0.013062	0.012086	18	0.048399	0.032257	13	62	15
GG15	0.009889	0.007833	14	0.048828	0.043848	19	0.013062	0.012685	20	0.048399	0.046788	16	69	18
GG16	0.009889	0.007748	12	0.048828	0.03839	12	0.013062	0.011688	13	0.048399	0.030983	10	47	12
GG17	0.009889	0.009131	18	0.048828	0.042074	16	0.013062	0.012424	19	0.048399	0.040489	14	67	17
GG18	0.009889	0.007748	13	0.048828	0.03839	13	0.013062	0.011688	14	0.048399	0.030983	11	51	13
LN1	1.244456	1.244456	24	0.959893	0.959893	26	0.959372	0.959372	25	2.020992	2.020992	27	102	26
LN2	1.244456	1.244456	25	0.959893	0.951989	25	0.959372	0.957704	24	2.020992	2.020802	26	100	24
LN3	0.006748	0.006748	11	0.034872	0.034872	11	0.009647	0.009647	12	0.026523	0.026523	9	43	11
LN4	0.006748	0.004563	6	0.034872	0.03279	9	0.009647	0.009249	11	0.026523	0.024542	6	32	8
LN5	0.006748	0.005992	9	0.034872	0.020875	1	0.009647	0.008385	5	0.026523	0.021913	4	19	4=
LN6	0.006748	0.004935	8	0.034872	0.026769	4	0.009647	0.008798	6	0.026523	0.020756	1	19	4=
LN8	2.60814	1.516058	26	3.483638	2.645717	27	3.683598	3.316535	27	0.974126	0.936129	25	105	27
LN10	2.60814	1.574765	27	3.483638	0.750217	24	3.683598	0.963543	26	0.974126	0.44544	24	101	25
LN12	0.00632	0.006244	10	0.034873	0.032083	8	0.009644	0.008864	7	0.026618	0.026121	8	33	9
LN14	0.00632	0.004742	7	0.034873	0.030526	7	0.009644	0.009174	9	0.026618	0.025803	7	30	7
LN16	0.00632	0.004561	5	0.034873	0.022425	2	0.009644	0.007553	1	0.026618	0.022202	5	13	1

Table 2: Results of analyses

6. Analysis

Four sets of bidding data, termed here **Cases 1-4**, were analysed:

The **Case 1** data comprised a donated set from a construction company operating in the London area. They cover much of the company's building contract bidding activities during a twelve-month period in the early 1980's and comprise 51 auctions for which a full set of bids, together with the identity of each bidder, were available.

The **Case 2** data comprise a donated set from a north of England County Council for building contract auction bids over an approximately four year period prior to July 1982. The resulting number of auctions for which a full set of bids, together with the identity of the bidder, was available for analysis totals 218.

The **Case 3** data comprise the **Case 1** data supplemented by a similar set of 373 auctions obtained from the records of a bidding information agency in the London area for the period November 1976 to February 1977.

The **Case 4** data were obtained from the Hong Kong Architectural Services Department for their building contract auction bids for the period November 1990 to November 1996. The resulting number of auctions for which a full set of bids, together with the identity of the bidder, was available for analysis total 267.

The above procedure was applied for each model for each dataset. Table 2 summarises the results of this analysis, showing the baseline and the final MSE upon termination, with the rank order of each model for each case. Therefore, the best models for **Case 1** are GG5, GG6 GG9 and GG10 all of which have very similar MSE termination values of 0.003573, 0.003673, 0.003773 and 0.003773 respectively. For **Case 2** the lognormal LN5 (0.020875) appears to be the clear favourite followed by LN16 (0.022425) while for **Case 3** the leading contenders are LN16 (0.007552), GG5 (0.007623) and GG6 (0.007665), with LN6 (0.020756) being the standout for **Case 4**, followed by GG10 (0.021639), GG9 (0.021697) and LN5 (0.021913). There is therefore no single model that performs best for all datasets. However, some models do appear to perform significantly better[6] than other models. Based on the overall rankings these are LN16, GG9, GG10, LN5, LN6, GG5, LN14, LN4, GG6 and LN3. It is also of interest to note that the London datasets (**Case 1** and **Case 3**) have a much lower terminal MSE and hence better fit in general, which may possibly be due to (1) the close proximity, intense competition, good market intelligence and (2) with the predominant use of selective (short-listed) tendering in that location. In contrast, **Case 2** experiences (2) but not (1), while **Case 4** has (1) but not (2).

7. Discussion

Table 3 summarises these top 11 models for each **Case** in terms of their percentage above the lowest final MSE and Table 4 details the resulting number of bidder groupings under each parameter for each model. From this, some observations can be made.

Firstly, Table 4 shows that some of the models overlap. For example, the GG6 model for **Case 1**, resulted in all the bidders having the same second parameter despite having an unconstrained second parameter. This therefore reduces GG6 to the GG5 model, which deliberately allows only one group of bidders for that parameter. Hence, the results are the same for each model, although GG5 should take preference as it has less parameters to estimate. The same applies also to **Case 3**.

	Case 1			Case 2			Case 3			Case 4			Rank	
Model	**Final MSE**	**Rank**	**% above lowest**	**Final MSE**	**Rank**	**% above lowest**	**Final MSE**	**Rank**	**% above lowest**	**Final MSE**	**Rank**	**% above lowest**	**Avg % above lowest**	**Overall**
LN16	0.004561	5	27.651833	0.022425	2	7.4251497	0.007553	1	0	0.022202	5	6.9666602	10.510911	1
GG9	0.003763	2	5.3176602	0.027391	6	31.214371	0.007755	4	2.6047711	0.021697	3	4.5336288	10.917608	2
GG10	0.003573	1	0	0.027119	5	29.911377	0.009111	8	17.100209	0.021639	2	4.2541916	12.816444	3
LN6	0.004935	8	38.119228	0.026769	4	28.234731	0.008798	6	14.150943	0.020756	1	0	20.126226	4
LN5	0.005992	9	67.702211	0.020875	1	0	0.008385	5	9.9224806	0.021913	4	5.5742918	20.799746	5
LN14	0.004742	7	32.717604	0.030526	7	46.232335	0.009174	9	17.669501	0.025803	7	24.315861	30.233825	6
LN4	0.004563	6	27.707809	0.03279	9	57.077844	0.009249	11	18.337118	0.024542	6	18.240509	30.34082	7
LN12	0.006244	10	74.755108	0.032083	8	53.691018	0.008864	7	14.790162	0.026121	8	25.847948	42.271059	8
GG5	0.003773	3	5.5975371	0.02581	3	23.640719	0.007665	3	1.4611872	0.051474	20	147.99576	44.673801	9
LN3	0.006748	11	88.860901	0.034872	11	67.051497	0.009647	12	21.70623	0.026523	9	27.784737	51.350841	10
GG6	0.003773	4	5.5975371	0.034574	10	65.623952	0.007623	2	0.9182736	0.053313	21	156.85585	57.248903	11

Table 3: Top 11 models compared

Case		LN16	GG9	GG10	LN5	LN6	GG5	LN14	LN4	LN12	GG6	LN3
1	Baseline MSE	0.006320	0.009889	0.009889	0.006748	0.006748	0.009378	0.006320	0.006748	0.006320	0.009378	0.006748
	Final MSE	0.004561	0.003763	0.003573	0.005992	0.004935	0.003773	0.004742	0.004563	0.006244	0.003773	0.006748
	Rank	5	2	1	9	8	3	7	6	10	4	11
	1st param	5	10	9	4	3	26	1	1	1	26	1
	2nd param	1	1	1	1	3	1 (GG9)	5	4	1	1 (GG9)	1
	3rd param	7	1	2	1	1	1	4	1	2	1	1
2	Baseline MSE	0.034873	0.048841	0.048841	0.034873	0.034873	0.042495	0.034873	0.034872	0.034873	0.042495	0.034872
	Final MSE	0.022425	0.027391	0.027119	0.020875	0.026769	0.025810	0.030526	0.032790	0.032083	0.034574	0.034872
	Rank	2	6	5	1	4	3	7	9	8	10	11
	1st param	6	14	14	17	5	19	1	1	1	4	1
	2nd param	1	1	1	1	4	1 (GG9)	3	4	1	2	1
	3rd param	7	1	2	1	1	1	3	1	20	1	1
3	Baseline MSE	0.009644	0.013091	0.013091	0.009647	0.009647	0.011708	0.009644	0.009647	0.009644	0.011708	0.009647
	Final MSE	0.007553	0.007755	0.009111	0.008385	0.008798	0.007665	0.009174	0.009249	0.008864	0.007623	0.009647
	Rank	1	4	8	5	6	3	9	11	7	2	12
	1st param	3	22	2	7	3	20	1	1	1	28	1
	2nd param	1	1	1	1	1 (LN5)	1 (GG9)	2	6	1	1 (GG9)	1
	3rd param	3	1	2	1	1	1	4	1	28	1	1
4	Baseline MSE	0.026618	0.048954	0.048954	0.026523	0.026523	0.069402	0.026618	0.026523	0.026618	0.069402	0.026523
	Final MSE	0.022202	0.021697	0.021639	0.021913	0.020756	0.051474	0.025803	0.024542	0.026121	0.053313	0.026523
	Rank	5	3	2	4	1	20	7	6	8	21	9
	1st param	4	10	11	16	6	30	1	1	1	15	1
	2nd param	1	1	1	1	7	1 (GG9)	3	12	1	2	1
	3rd param	4	1	3	1	1	1	1 (LN4)	1	11	1	1

Table 4: Top 11 models – number of groupings

Similar considerations for LN6 and LN14, and inspection of Table 3 leads to the conclusion that the three highest ranked models of LN16, GG9 and GG10 are likely to be the best models, with LN5 and LN6 being good reserves.

8. Conclusions

Construction contract auctions are characterised by (1) each contract being different in size, (2) a small number of bidders for each auction, (3) different bidders bidding for each contract (making the contract-bidder matrix usually well over 90% sparse). For the purpose of modelling, several probability distribution forms have been proposed in the past but with little testing for suitability to date. Stacey's GG model includes many of the forms as special cases, depending on the value of the parameters involved, and the maximum likelihood equations have been presented here as a means of estimating these parameters. This involved the provision of extra parameters to accommodate (1) and the use of iteration procedures for (2). For (3), bidders with similar parameters were placed into groups empirically by a least-squares cross-validation procedure according to a new scoring measure. This was applied to four sets of real bidding data under a variety of simplifying assumptions, including all those used previously in the main literature. Examination of the results for the 29 models involved indicates that no one single model performs best for all datasets. However, some models do appear to perform significantly better than other models with the data used. Based on the overall rankings, these are LN16 (weighted lognormal with constant variance), GG9 (gamma with unitary second parameter and constant power term) and GG10 (Weverbergh's (1982) gamma with just unitary second parameter), with LN5 (Carr's (1982) unweighted lognormal and constant variance) and LN6 (Weverbergh (1982) and Skitmore's (1991) unweighted lognormal) being good reserves, and suggest it may be beneficial in future to concentrate on this subset of five models for further development.

It should be noted that all five models take a very large amount of computing time to build. The estimation for LN16, for example, took over 5 years continual processing on the university supercomputer. A mitigating factor however, is that, once the analysis is completed, adding a further set of contract auction bids to an existing set should take a relatively small time to analyse. Also, the work described in this paper is aimed at identifying some of the most suitable statistical models to use with bidding data of this kind. Before commencing, there was no extant empirical knowledge at all of the most likely candidates by multivariable analysis – with very few exceptions, all the models in the literature having been assumed on *apriori* reasoning. Having now reduced the possible models to a small number of serious contenders, future research in this area would benefit well from the development of more economical approaches (such as the preliminary allocation of bidders to groups by a non- least-squares cross-validation over fit method perhaps). Also, in view of the seemingly contingent nature of the situation, it would be useful to investigate the exogenous factors involved, such as local bidding procedures or intensity of market conditions. In addition, as suggested by one reviewer, an investigation is warranted of alternative measures to MSE of goodness of fit, such as those based the chi square or log likelihood statistic.

Acknowledgements

Thanks go to Ross McVinish for pointing out some errors in the original formulae. Also, to the QUT High Performance Computing Lab for sanctioning my continual use of their Cray Supercomputer over the last eight years."

References

Abadi A Amanpour F Bajdik C and Varai P (2012). Breast cancer survival analysis: Applying the generalized gamma distribution under different conditions of the proportional hazards and accelerated failure time assumptions. *International Journal of Preventive Medicine* **3**(9) 644-651.

Alexander AB (1970). What price estimating accuracy? Paper 534, Metal Fabricating Institute, Rockford, Illinois.

Arps JJ (1965). A strategy for sealed bidding. *Journal of Petroleum Technology*, **17**(9): 1033-1039.

Babu KS Yarramalle S and Penumatsa SV (2012). Text-independent speaker recognition using emotional features and generalized gamma distribution. *International Journal of Computer Applications* **46**(2): 16-18.

Brown KC (1966). A theoretical and statistical study of decision making under uncertainty - a competitive bidding for leases of offshore petroleum tracts, PhD dissertation, Southern Methodist University, Dallas.

Capen EC Clapp RV and Campbell WM (1971). Competitive bidding in high risk situations. *Journal of Petroleum Technology* **23**(6): 641-653.

Carr RI (1982). General bidding model. *Journal of the Construction Division*. Proceedings of the American Society of Civil Engineers, **108** (CO4): 639-650.

Carr RI (1983). Impact of number of bidders on competition. *Journal of Construction Engineering and Management, Proceedings of the American Society of Civil Engineers* **109**(1): 61-73.

Cauwelaert FV and Heynig E (1978). Correction of bidding errors: The Belgian solution. *Journal of the Construction Division*, American Society of Civil Engineers **105**(CO1): 13-23.

Crawford PB (1970). Texas offshore bidding patterns. *Journal of Petroleum Technology* **22**(3): 283-289.

Dougherty EL and Nozaki M (1975). Determining optimum bid fraction. *Journal of Petroleum Technology* **27**(3): 349-356.

Dowe DL Farr GE Hurst AJ and Lentin KL (1996). Information-theoretic football tipping. Technical Report no 96/297, School of Computer Science and Software Engineering, Monash University, Clayton, Melbourne, Victoria 3168, Australia.

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

Emond LJ (1971). Analytical strategy for the competitive price setter. *Cost and Management*. Sep-Oct: 6-11.

Fine B and Hackemar G (1970). Estimating and bidding strategy. *Building Technology and Management* **8**(9): 8-9.

Friedman L (1956). A competitive bidding strategy. *Operations Research* **1**(4): 104-112.

Gates M (1967). Bidding strategies and probabilities. *Journal of the Construction Division, Proceedings of the American Society of Civil Engineers* **93** (CO1): 75-107.

Grinyer PH and Whittaker JD (1973). Managerial judgement in a competitive bidding model. *Operational Research Quarterly* **24**(2): 181-191.

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

Hvingel C Kingsley MCS and Sundet JH (2012). Survey estimates of king crab (*Paralithodes camtschaticus*) abundance off Northern Norway using GLMs within a mixed generalized gamma-binomial model and Bayesian inference. *ICES Journal of Material Science* **69**(8): 1416-1426.

Keiding N Hansen OKH Sørensen DN and Slama R (2012). The current duration approach to estimating time to pregnancy. *Scandinavian Journal of Statistics: Theory and Applications* **39**(2):185-204.

Klein JD (1976). Joint ventures and bidding for offshore oil. In Amihud Y (ed.), *Bidding and Auctioning for Procurement and Allocation*. New York University Press: New York University

Lawless JF (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons ISBN 0 471 08544 8.

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

McCaffer R and Pettitt AN (1976). Distribution of bids for buildings and roads contracts. *Operational Research Quarterly* **27**(4i): 835-843.

Mitchell MS (1977). The probability of being the lowest bidder. *Applied Statistics* **2**(2): 191-194.

Morrison N and Stevens S (1980). *Construction cost database.* 2nd annual report of research project by Dept of Construction Management, University of Reading, for Property Services Agency, Directorate of Quantity Surveying Services, DOE, UK.

Nadarajah S and Gupta AKA (2007). Generalized gamma distribution with application to drought data. *Mathematics and Computer Simulation* **74**(1): 1-7.

Oren SS and Rothkopf MH (1975). Optimal bidding in sequential auctions. *Operations Research* **23**(6): 1080-1090.

Prentice RL (1974). A log gamma model and its maximum likelihood estimation. *Biometrika* **61**(3): 539-544.

Rothkopf MH (1969). A model of rational competitive bidding. *Management Science* **15**(7): 362-373.

Shin JW Chang JH and Kim NS (2005). Statistical modeling of speech based on generalized gamma distribution. *IEEE Signal Processing Letters* **12**(1): 258-261.

Skitmore RM (1991). The contract bidder homogeneity assumption: An empirical analysis. *Construction Management and Economics* **9**(5): 403-429.

Skitmore RM (2001). Graphical method for identifying high outliers in construction contract auctions. *Journal of the Operational Research Society* **52**(7): 800-809.

Skitmore RM (2002a). Predicting the probability of winning sealed bid auctions: a comparison of models. *Journal of the Operational Research Society* **53**: 47-56.

Skitmore RM (2002b). Identifying non-competitive bids in construction contract auctions. *OMEGA: International Journal of Management Science* **30**(6): 443-449.

Skitmore RM (2004). Predicting the probability of winning sealed bid auctions: the effects of outliers on bidding models. *Construction Management and Economics* **22**(1): 101-109.

Skitmore RM and Lo HP (2002). A method for identifying high outliers in construction contract auctions. *Engineering, Construction and Architectural Management* **9**(2): 90-130.

Skitmore RM and Pemberton J (1994). A multivariate approach to contract bidding mark-up strategies. *Journal of the Operational Research Society* **45**(11): 1263-1272.
Skitmore RM Pettitt AN and McVinish R (2007). Gates' bidding model. *ASCE Journal of Construction Engineering and Management* **133**(11): 855-863.
Stacey E.W (1962). A generalization of the gamma distribution. *Ann Math Stat* **33** (3): 1187-1192.
Thomas DJ and Smith JM (1990). Joseph Raphson, FRS. *Notes and Records of the Royal Society of London* **44**(2): 151-167.
Vickrey W (1961). Counterspeculation auctions and competitive sealed transfers. *Journal of Finance* **16** (March) 8-37.
Weverbergh M (1982). Competitive bidding: estimating the joint distribution of bids. Working paper 82-79, Dec, Centrum voor Bedrijfseconomie en Bedrijfseconometrie Universiteit Antwerpen - USFIA.
Whittaker JD (1970). A study of competitive bidding with particular reference to the construction industry. PhD thesis. City University, London, UK.
Ypma JT (1995). Historical development of the Newton–Raphson method. *SIAM Review* **37**(4): 531-551.

[1] See Skitmore *et al* (2007).
[2] For construction contract auctions, the term "contract" is equivalent to "item" in the more general bidding literature.
[3] Letting σ^2 denote the second moment in $f\left(.,\sigma^2\right)$ we can further associate $\sigma_i = 0$ with Friedman (1956) and $\sigma_1 = \sigma_2 = \ldots = \sigma_k$ with Carr (1982).
[4] The opportunities for merging cells are limited, as even with the bidder groupings described later, there is no obvious way in which contracts can be merged as well.
[5] See Ypma (1995) for example, on the background to this method, said to "lurk inside millions of modern computer programs" (Thomas and Smith, 1990).
[6]Being an essentially nonparametric analysis, the term significance is used here in an observational rather than statistical sense.