Predicting the probability of winning sealed bid auctions: The effects of outliers on bidding models

Skitmore, Martin

Published in:
Construction Management and Economics

DOI:
10.1080/0144619042000186103

PREDICTING THE PROBABILITY OF WINNING SEALED BID AUCTIONS: THE EFFECTS OF OUTLIERS ON BIDDING MODELS

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

Martin Skitmore
School of Construction Management and Property
Queensland University of Technology
Gardens Point
Brisbane Q4001
Australia

PREDICTING THE PROBABILITY OF WINNING SEALED BID AUCTIONS: THE EFFECTS OF OUTLIERS ON BIDDING MODELS

ABSTRACT

This paper is concerned with the effect of outliers on predictions of the probability of tendering the lowest bid in sealed bid auctions. Four of the leading models are tested relative to the equal probability model by an empirical analysis of three large samples of real construction contract bidding data via *all-in* (in-sample), *one-out* and *one-on* (out-of-sample) frames. Outliers are removed in a sequence of cut-off values proportional to the standard deviation of bids for each auction. A form of logscore is used to measure the ability to predict the probability of each bidder being the lowest. The results show that, although statistically significant in some conditions, all the models produce rather poor predictions in both *one-out* and *one-on* mode, with the effects of outliers being generally small.

Keywords: Bidding models, bidding theory, construction contracts, empirical tests, predicted probability, probability of lowest bid, sealed bid auctions, tendering theory, logscore test, outliers.

INTRODUCTION

The probability of individual contestants winning a bidding auction can be a useful piece of information for many people, not least the contestants themselves. Potential bidders can utilise this information to decide in which auctions to participate, when to try to obtain an invitation to participate, whether to enter a bid and, if so, the dollar value of the bid. Similarly, the auctioneer can also utilise the information in deciding when and how to hold the auction, how many and which bidders to invite, and the criterion for determining the winner.

Most of the literature on the subject is concerned with setting a price, x, so that the probability, $Pr(x)$, of winning the auction reaches some desired level. Several models have been proposed for predicting $Pr(x)$, and these have been subject to quite lengthy, but as yet inconclusive, discussion based on the theoretical merits of each model. Most of the empirical studies that have taken place have concentrated on fitting a single underlying probability density function to all bids, including the uniform (Fine and Hackemar, 1970; Whittaker, 1970; Grinyer and Whittaker, 1973), normal (McCaffer, 1976; Skitmore, 1986) and gamma (Friedman, 1956; Hossein, 1977) while recent work (Skitmore, 2001; Skitmore and Lo, 2002) uses the single underlying density approach in examining the effects of removing potential outliers - finding the truncated lognormal to be the best fit followed by the truncated normal distribution, with the uniform distribution way behind.

A great deal of attention has centred at the theoretical level on the actions of individual

This is an Accepted Manuscript of an article published in 2004 by Taylor & Francis in Construction Management and Economics, available online: https://doi.org/10.1080/0144619042000186103

competitors, information about which has been regarded as "critical" (Griesmer *et al*, 1967). Most models implicitly require the analyst to develop probability distributions for each competitor's bid (de Neufville *et al*, 1977), which means that the model builder is faced with the problem of explicating probability laws for opposing bids (Weverbergh, 1981), and often with a palpable lack of empirical supporting evidence. Skitmore's (1991) multivariate model provides a good solution to this, although at the cost of a parametric assumption in the form of a composite density. Friedman's (1956) model, on the other hand, is particularly susceptible demanding, as it does, the collection of bid/cost estimate ratios against **each** competitor in order to construct frequency distributions of sufficient dimension to enable probability density functions to be fitted. This difficulty can be partly overcome by recourse to the single underlying density assumption - only certain bidders for which a good quantity of data is available being treated individually (Capen *et al*, 1971; Curtis and Maines, 1973; Fuerst, 1977; Morin and Clough, 1969; for instance). Assumptions of this kind can be misleading, however, as it has been shown that differences in assumptions of the spreads of opposing bids can have significant effects on results (Carr, 1982; Weverbergh, 1982, p.26).

All the work mentioned so far has been concerned with model *fitting*. The main issue here, however, is the degree to which the models are able to predict events. For bidding models, one event of importance is the probability of entering the lowest bid in a future auction. The main interest, therefore, is in the goodness of the *forecast* of this probability. Until recently, there have been no empirical tests of this in construction contract bidding, presumably due to the lack of development of appropriate tests. This situation has now changed, with the application of a form of logarithmic scoring function to assess the main models proposed in the literature (Skitmore, 2002). However, work on the effects of outliers (Skitmore, 2001; Skitmore & Lo, 2002) suggests that significant improvements might be possible through their removal and this has yet to be examined.

This paper, therefore, further develops the Skitmore (2002) approach to enable the examination of effects of potential outliers on the main models found in the literature.

MODELS, DATA, TESTING FRAMES AND OUTLIER REMOVAL

- The five models investigated comprise: (1) Friedman's (1956) model; (2) Gates' (1967) model; (3) Carr's (1982) model; and (4) two versions of Skitmore's (1991) model. These were tested for superiority over Pim's (1974) equal probability model.
- Three datasets were used, termed Case A, B and C. Case A and B are the same as Case 2 and 4 respectively in Skitmore (2002). Case C is a set of data for a series of Hong Kong construction contracts.
- The three testing frames are termed *all-in* (in-sample), *one-out* and *one-on* (out-of-sample).

The models, testing frames and Cases A and B data are described in detail in Skitmore (2002) including the necessary modifications that had to be made due to the limitations of the data and nature of the analysis. The method of outlier removal is described in Skitmore (2001). A summary is provided in the Appendix.

TESTS

Background

The only method available at the present time for assessing the accuracy of probability predictions is Dowe *et al*'s (1996) scoring method, devised to test the accuracy of Australian football tipsters. It is therefore designed for a two-person/team zero-sum game in which team A either wins (in which case team B loses) or team A loses (in which case team B wins). The Dowe *et al* scoring method is:

Score (if A wins) = $1+\log_2 p$, or (if A loses) = $1+\log_2(1-p)$

where p is the tipsters' predicted probability of A winning. Thus, if the tipsters predict the probability of A winning as 1, and A wins, then they are assigned a score of $1+\log_2 = 1$. If A loses, on the other hand, they receive a score of $1+\log2(0) = \infty$. Similarly, if they predict the probability of A winning as 0, and A wins, then they are assigned a score of $1+\log_2(0)= \infty$. If A loses, on the other hand, they receive a score of $1+\log2 = 1$. So, if we average the tipsters' scores over a series of trials, the one with the significantly lowest average score can be adjudged the winner (Dowe *et al* make no reference to significance).

For construction contractors' predicted probabilities of winning, it is necessary to extend this method to more than 2 teams. A few alternatives are available. One is, following the spirit of the Dowe *et al* game, just to test the prediction of the winner and ignore the other k-1 predictions in a k-bidder auction. A better way though, and one that is more suitable to testing bidding models where all the bidders results are of interest, is to test the probability prediction of each bidder.

To do this, we first write Dowe *et al*'s formula as:

$$1+\pi_i \log_2 p_i + (1-\pi_i)\log_2(1-p_i)$$

where π_i is the actual result (π_i=1 if bidder i wins and π_i=0 if bidder i doesn't win). So, for all the bidders in a k-bidder auction we get the auction score:

$$S = \sum_{i=1}^{k} \{1+\pi_i \log_2 p_i + (1-\pi_i)\log_2(1-p_i)\} \qquad (1)$$

Significance

Significance is, by definition, the probability of a chance result occurring. A chance result in this case, occurs when all chances of winning bidders are equiprobable, ie., $p_1=p_2=\ldots p_n = 1/k$. Thus, the expected value of the logscore for a chance result is, from (1):

$$S_e = \sum_{i=1}^{k} \left\{1+\pi_i \log_2 \frac{1}{k} + (1-\pi_i)\log_2\left(1-\frac{1}{k}\right)\right\} \qquad (2)$$

This is an Accepted Manuscript of an article published in 2004 by Taylor & Francis in Construction Management and Economics, available online: https://doi.org/10.1080/0144619042000186103

which is, expanded:

$$S_e = k + \log_2\left(\frac{1}{k}\right) + (k-1)\log_2\left(\frac{k-1}{k}\right) \tag{3}$$

Significance can, therefore, be assessed by the departure of S from S_e and will vary according to k. To make S independent of k, we can standardise by dividing (3) into (1), ie.,

$$S_s = \frac{\sum_{i=1}^{k}\{1 + \pi_i \log_2 p_i + (1-\pi_i)\log_2(1-p_i)\}}{k + \log_2\left(\frac{1}{k}\right) + (k-1)\log_2\left(\frac{k-1}{k}\right)} \tag{4}$$

Characteristics

Eqn (4) has some nice characteristics for assessing bidding probability predictions:

Prediction	**model**	**S_s**
Best possible prediction	$p_i = \pi_i$	0
Equiprobable prediction	$p_1 = p_2 = ... = p_k = 1/k$	1
Worst possible prediction	$p_i = 1 - \pi_i$	∞
Zero's prediction	$p_1 = p_2 = ... = p_k = 0$	∞
Unitary prediction	$p_1 = p_2 = ... = p_k = 1$	∞

Note: with most of the other alternatives tried, predicting a zero probability for all bidders scores unreasonably well due to being incorrect for only one bidder.

Simplification

Eqn (4) simplifies, seemingly without loss, by removing the 1+ and freeing the log base (Dowe *et al* describes the rationale for the 1+ as to avoid psychological problems in playing a game where the winner is closest to zero!), ie.,

$$S_s = \frac{\sum_{i=1}^{k}\{\pi_i \log p_i + (1-\pi_i)\log(1-p_i)\}}{\log\left(\frac{1}{k}\right) + (k-1)\log\left(\frac{k-1}{k}\right)} \tag{5}$$

Comparing datasets

Eqn (5) gives the standardised Dowe *et al* score of a single auction. To compare datasets just involves averaging S_s over all auctions for each dataset, ie.,

$$S_s^* = \frac{1}{c}\sum_{j}^{c} S_{sj} \qquad (6)$$

where there are j=1,2,…,c auctions in the dataset, each comprising k_j bids.

To compare different models, S_s^* is calculated for each Case and *sample-frame* - lower values indicating better predictions than high values. As there are no statistical tests that indicate the significance of these differences, values were generated by computer simulation to obtain the approximate 5% percentage points for each Case, *sample frame* and outlier cut-off value.

RESULTS

Fig 1a-c shows the best *all-in frame* results for the five models (**F**, **G**, **C**, **S1** and **S2** representing the Friedman, Gates, Carr and the two Skitmore models respectively) in comparison with the Pim equal probability model (**E**) for q=1,2, ..., 30 with outlier removal cut-off values of –0.5, -0.4, …, 3.0 for Cases A-C. This shows all the models, especially Friedman and Gates, performing well (ie., below unity) for Cases A and B with the best results occurring without the removal of outliers. The results for the Case C dataset, however, with a much greater average number of bidders per auction, are much worse, with Friedman's model being particularly poor.

Fig 2a-c shows the equivalent *one-out frame* results and Fig2 3a-c the equivalent *one-on frame* results. These show the models offer little improvement on equal probability in *ex-post* forecasting- with the Friedman's model performing particularly poorly.

Whilst improvement in *ex-post* forecasting on the equal probability model is clearly at best marginal, it is of interest to see the extent to which these marginal improvements are significant. By generating random values for each auction and repeating the analysis many times, it is possible to approximate the 5% significance values for each of the five sets of ex-post results. Comparing these with the values shown in Figs 2 and 3 enables the significant results to be identified (Table 1). One point of interest in these results is that the majority were obtained for $q \geq 10$, ie., the best results occurred when grouping together bidders with more than 10 bids in the database. Of course, setting q to a very high value would result in all bidders being treated as equal and thus equiprobable. With the sparse bidding matrices that occur with these datasets (and, typically, construction contract auction datasets in general), $q \geq 10$ involves an amount of grouping that is getting quite close to this and hence explains the closeness of the results to the equal probability model.

This is an Accepted Manuscript of an article published in 2004 by Taylor & Francis in Construction Management and Economics, available online: https://doi.org/10.1080/0144619042000186103

CONCLUSIONS

Previous work in auction bidding has largely been carried out without any real supporting data. In the context of construction contract auction bidding, it has been doubted that sufficient data can be mustered for each bidder for any effective predictions to be made. In analysing some real and typical sets of construction contract auction bid data it has been possible to compare the major models here against pure chance, showing that at best only a marginal improvement on chance seems possible.

Of particular interest, however, is the extent to which grouping provides the best results in the analysis, with most involving $q \geq 10$ where q is a function of the number of bids made by each competitor. This is essentially the Friedman/Gates grouping criterion. Other *a priori* grouping criteria are possible, such as size and workload of the bidder or the size and type of project, and which may produce better results. Alternatively, it should be possible to group bidders empirically in such a way as to produce the best *expost* logscores - to enable bidder groupings to be *revealed* rather than *predetermined*. Some preliminary work suggests that very substantial improvements may be made by using this latter approach.

REFERENCES

Capen, E.C., Clapp, R.V., Campbell, W.M., 1971, Competitive bidding in high risk situations, *Journal of Petroleum Technology*, Jun, 641-53.

Carr RI (1982). General bidding model. *Journal of the Construction Division*. Proceedings of the American Society of Civil Engineers, **108** (CO4), 639-50.

Curtis, F.J., Maines, P.H., 1970, Selecting price quotations for an industrial firm's sale of individual contract projects, *Operations Research* **18** 1220-4.

De Neufville, R.D., Hani, E.N., Lesage, Y., 1977, Bidding model: effects of bidder's risk aversion, *Journal of the Construction Division, Proceeding of the American Society of Civil Engineers*, **103** (CO1) 57-70.

Dowe D L, Farr G E, Hurst A J, and Lentin K L (1996). Information-theoretic football tipping. Technical Report no 96/297, School of Computer Science and Software Engineering, Monash University, Clayton, Melbourne, Victoria 3168, Australia.

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

Friedman L (1956). A competitive bidding strategy. *Operations Research* **1**(4): 104-12.

Fuerst, M., 1977, Bidding models: truths and comments, *Journal of the Construction Division, Proceeding of the American Society of Civil Engineers*, **103** (CO1) 139-52.

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

Griesmer, J.H., Levitan, R.E., Subik, M., 1967, Towards a study of bidding processes, part 4: games with unknown costs, *Naval Research Logistics Quarterly* **14** 415-33.

Grinyer, P.H., Whittaker, J.D., 1973, Managerial judgement in a competitive bidding model, *Operational Research Quarterly* **24**(2): 181-191

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

McCaffer, R., 1976, Contractor's bidding behaviour and tender price prediction, PhD thesis,

Loughborough University of Technology.
Morin, T.L., Clough, R.H., 1969, OPBID: competitive bidding strategy model, *Journal of the Construction Division, Proceedings of the American Society of Civil Engineers*, **95**(CO1) 85-106.
Pim J C (1974). Competitive tendering and bidding strategy. *National Builder*, **55** (11) pp. 541-545, **56** (2) pp. 56-57, **56** (10) pp.361-365, **57** (3) pp. 68-70.
Skitmore, R.M., 1986, A model for the construction project selection and bidding decision, PhD. thesis, Department of Civil Engineering, University of Salford.
Skitmore R M (1991). The contract bidder homogeneity assumption: an empirical analysis. *Construction Management and Economics*, **9**(5) 403-29.
Skitmore R M (2001) Graphical method for identifying high outliers in construction contract auctions. *Journal of the Operational Research Society* **52** 800-9.
Skitmore, R.M., Lo, H.P. 2002. A method for identifying high outliers in construction contract auctions. *Engineering, Construction and Architectural Management* **9**(2) 90-130.
Skitmore R M and Pemberton J (1994). A multivariate approach to contract bidding mark-up strategies. *J Opl Res Soc*, **45**(11) 1263-72.
Skitmore R M and Pettitt A N (in press) On Gates' bidding model *Journal of Construction Engineering and Management*.
Weverbergh, M., 1981, Competitive bidding models: an overview, Working paper 81-72, Aug, Centrum voor Bedrijfseconomie en Bedrijfseconometrie Universiteit Antwerpen - USFIA.
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 J D (1970). A study of competitive bidding with particular reference to the construction industry. PhD thesis, City University, London.

APPENDIX: MODELS, DATA AND TESTING FRAMES

Models

Letting $X_1, X_2, \ldots, X_k$ be independently distributed random variables then, if we generate one value, ie., $x_1, x_2, \ldots, x_k$ from each variable, the probability, $P_{i.}$, of x_i being the lowest is given by

$$P_{i.} = \int_{-\infty}^{\infty} f_i(x) \prod_{\substack{j=1 \\ j \neq i}}^{k} S_j(x)\, dx \qquad \textit{(A1)}$$

where $S_j(x)$ is the well-known survivor function $\int_x^{\infty} f_j(y)dy$. For the Friedman model, $f_i(x)$ has zero variance; for the Gates model, a proportional hazard function is used (Skitmore & Pettitt, in press), enabling $S_j(x)$ to be replaced with $\left[S_i(x)\right]^{P_{ji}/P_{ij}}$ where P_{ij} is the probability that

$x_i < x_j$, so that ***(A1)*** reduces to $P_{i.} = \dfrac{1}{1 + \sum_{\substack{j=1 \\ j \neq i}}^{k} \dfrac{1 - P_{ij}}{P_{ij}}}$; for the Carr model, $f(\,)$ is the normal density with equal variances; and for the Skitmore model, $f(\,)$ is lognormal.

(1) Equal probability model

The equal probability model, as implied by Pim, directly predicts each bidder's probability of being the lowest bidder as *1/k*. Of course, the same result is obtained when all bids are identically and independently distributed. This is essentially the control model representing chance. A good model should, of course, outperform the equal probability model by definition.

(2) Friedman's model

Friedman's approach is to transform x_i by dividing by the reference bidder's cost estimates, c_1, i.e.,

$$x_i^F = x_i / c_1 \qquad \textbf{\textit{(A2)}}$$

the shape and other distribution parameters for x_i^F being estimated from the frequency distribution of the x_i^F ratios for each competitor. x_1^F is however assigned the arbitrary parameter values $\mu_1 = x^*/c^*$ and $\sigma_1^2 = 0$, where x^* and c^* are the reference bidder's bid and cost estimate for the next auction.

Friedman's approach relies heavily on the availability of data, the theoretical density functions being fitted provided there are data for "enough previous contracts". A range of criterion values was used to determine this, ie., q=1,2,...,30, where q denotes the minimum number of previous bidding encounters between the reference bidder and a specific competitor. Where the actual number of previous bidding encounters between the reference bidder and a specific competitor was less than q, the probability was estimated as the mean 'success', $\bar{p}$, of the reference bidder against all other bidders, ie.,

$$\bar{p} = \frac{1}{n'} \sum_{j=1}^{n'} P_{ij} \qquad \textbf{\textit{(A3)}}$$

P_{ij} being estimated by the ratio of the number of previous auctions where bidder i's cost estimates were less than bidder j's bids to the total number of auctions where bidder i bid against bidder j. Where no previous meetings of a pair of bidders had taken place, each bidder in the pair was assigned a 0.5 probability of underbidding the other.

This is an Accepted Manuscript of an article published in 2004 by Taylor & Francis in Construction Management and Economics, available online: https://doi.org/10.1080/0144619042000186103

To enable $P_{i.}$ to be estimated for all bidders, the bidder i's cost estimate values were substituted with the bid value in Friedman's parameter estimation procedure and in computing P_{ij}. Similarly, the constant x^*/c^* was modified to $x^*/x^* = 1$.

(3) Gates' model

Gates' approach is to estimate $P_{i.}$ directly, by the formula $P_{i.} = \dfrac{1}{1+\sum_{\substack{j=1 \\ j\neq i}}^{k} \dfrac{1-P_{ij}}{P_{ij}}}$

Gates' model is recommended for use in situations where there is "sufficient bidding data relating to every competitor bidder on the particular job". As with Friedman's model, a range of q values was tried and the same procedure used where less than q meetings occurred. To enable $P_{i.}$ to be estimated for all bidders, cost estimates were again substituted by bid values.

(4) Carr's model

Carr uses Friedman's transformation for x_i and x_1 where, in addition, x_1 is also substituted by c_1, resulting in the transformation

$$x_1^C = c_1/\overline{c}_l \qquad \textbf{\textit{(A4)}}$$

The arbitrary assumption that the x_i^F and x_1^C are normally and homogeneously (equal variances) distributed then allows the straightforward estimation of the required parameters from the frequency distribution of the pooled x_i^F ratios.

Carr's approach presents no difficulty in handling sparse data as normality and homogeneity are assumed. The only change made was to again substitute the cost estimates c_1 with the bids x_1 to enable all bidder's $P_{i.}$ estimates to be made. Bidders with less than q data points were assigned equal probabilities ($1/k$).

(5) Skitmore's model

Skitmore's approach uses approximate maximum likelihood estimates to fit the model $y_{ij} \sim N(\alpha_i+\beta_j,\ s_i^2)$ to the transformed values $y_{ij} = \mathbf{ln}(x_{ij}-mx_{(1)j})$ where m is a constant with $0.5<m<0.9$ ($x_{(1)j}$ being the value of the lowest bid for the jth auction) and $N(\,)$ is the Normal probability density function. The probability prediction for y_1

$$\Pr(y_1) = \int_{-\infty}^{\infty} \frac{e^{-\frac{y_1^2}{2}}}{\sqrt{2\pi}} . \prod_{i=2} \left[\int_{y_i=\frac{\sigma_1 y_1+\mu_1-\mu_i}{\sigma_i}}^{\infty} \frac{e^{-\frac{y_i^2}{2}}}{\sqrt{2\pi}} dy_i \right] dy_1 \qquad \textbf{\textit{(A5)}}$$

is then obtained by substituting the estimates α_i and s_i into μ_i and σ_i respectively .

The model is difficult to apply in this form however as it involves the additional estimation of $x_{(1)}$. More recently, Skitmore and Pemberton (1994) circumvented this problem by recourse to a simple natural log transformation, ie. $x_i^S = \ln x_i$, and this was also used here. From this, the probability of each bidder winning is computed. In cases where there was less than q data points for a bidder, that bidder was assigned an average α and s value.

The Skitmore model treats bidders with a single data point (one recorded previous bid) as having an alpha value based on that bid. A modification was also used which assigns these bidders with the average alpha value of all other bidders. The simple natural log transformation was again used.

Data

Three sets of data, termed here Cases A-C, were analysed.

The Case A data comprised a donated set from a north of England County Council for building contract bids over approximately four years prior to July 1982. The resulting number of contracts for which a full set of bids, together with the identity of the bidder, was available for analysis totalled 218.

The Case B data comprised a donated set from a construction company operating in the London area. They covered much of the company's building contract bidding activities during a twelve month period in the early 1980's and comprised 51 auctions for which a full set of bids, together with the identity of each bidder. These were 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 C data were obtained from the Hong Kong Architectural Services Department for their building contract bids for the period November 1990 to November 1996. The resulting number of contracts for which a full set of bids, together with the identity of the bidder, was available for analysis totalled 267.

Testing frames

As the sole purpose of bidding models is for use in forecasting future outcomes, it is necessary to apply the models to out-of-sample data in addition to the in-sample data. Three forecast sample-testing frames are applied: (1) the *all-in* frame, (2) the *one-out* frame and (3) the *one-on* frame

The all-in frame

The *all-in* frame comprises all the data used to build the model. The testing procedure then simply tests the model against the data from which the model was built. In a similar way to

This is an Accepted Manuscript of an article published in 2004 by Taylor & Francis in Construction Management and Economics, available online: https://doi.org/10.1080/0144619042000186103

the regression coefficient of determination, the error rate for *all-in* frame analysis is an unrealistically low measure of forecasting ability due to the self-fulfilling nature of the procedure.

The one-out frame

The *one-out* frame involves the use of the cross-validation procedure, by which the first auction is omitted from the model building process - the resulting model being applied to the omitted auction - this procedure being repeated, with replacement, for all the auctions in the Case set. Cross validation provides a reasonably realistic simulated out-of-sample test provided no significant time, or sequencing, effects are involved. The method is equivalent to the regression deleted residual analysis

The one-on frame

The *one-on* frame provides a simulation that is perhaps the closest to forecasting reality. A small sample of, say 13 auctions is used to build a model, which is then applied to the 14th auction - this procedure being repeated with 14 auctions to build the model which is applied to the 15th auction, etc. Of course, if the model performs better than chance, the results tend to improve as the number of auctions used to build the model increases. The final model, which incorporates all except the last auction, coincides with the final *one-out* model. The *one-out* frame results are therefore indicative of the final stages of the *one-on* results.

Outlier removal

Skitmore (2001) examined six methods of outlier removal comprising: (1) highest *k'* bids, (2) highest *k-m* bids, (3) bids higher than the average bid plus x_1 times the standard deviation, (4) bids x_2 times higher than the mean bid, (5) bids higher than x_3 times the lowest bid, and (6) highest x_4% bids. These were applied systematically to all the auctions in six datasets, with the resulting reduced number of bids being tested graphically for conformity with the uniform, log uniform, normal and censored and uncensored lognormal distributions. This showed the uncensored lognormal to be the most appropriate distribution and method (3) to be the best censoring method, in terms of minimum number of outliers removed – a result confirmed statistically in a later analysis by Skitmore and Lo (2002), which showed method (3) to also be the best for producing the normal distribution. For the work described here, where homogeneity of bidders is not assumed, the amount of computations involved was such as to prohibit a full examination of all six methods. Instead, just method (3) was used on the grounds of its previous success – the cut-off being applied in successive multiples of the original standard deviation of the auction bids.

Model	Case A	Case B	Case C
One-out			
E	-	-	-
F	-	-	-
G		0.7 to 1.0	0.6 to 3.0
C		0.1 to 3.0	0.1 0.3 0.5
S1		-0.1 to 3.0	-0.5
S2		-0.1 to 0.1 0.4 0.6 to 3.0	-
One-on			
E	-	-	-
F	-	-	-
G	-	-	0.1 to 0.2 0.4 to 0.5 0.9 to 1.4
C	-	0.6 to 1.8	-
S1	-0.4 to –0.3 0.0 to 0.7 1.1 to 1.2 1.3 to 1.7	0.7 to 3.0	-0.4 to 0.0
S2	-0.4 to –0.1 0.1 to 0.6 1.4 to 1.6	-0.5 to 3.	-0.4 0.0 to 0.1

Table 1: Significant results

Fig 1a: Case A all-in results

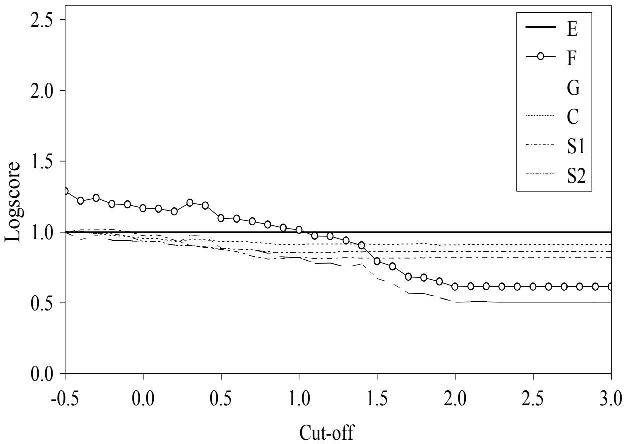


Fig 1b: Case B all-in results

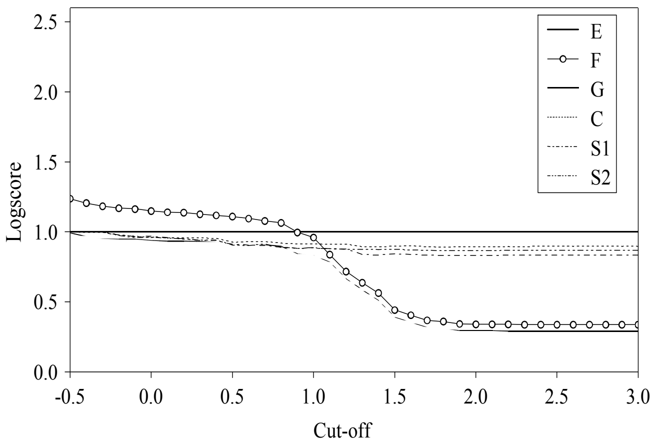


Fig 1c: Case C all-in results

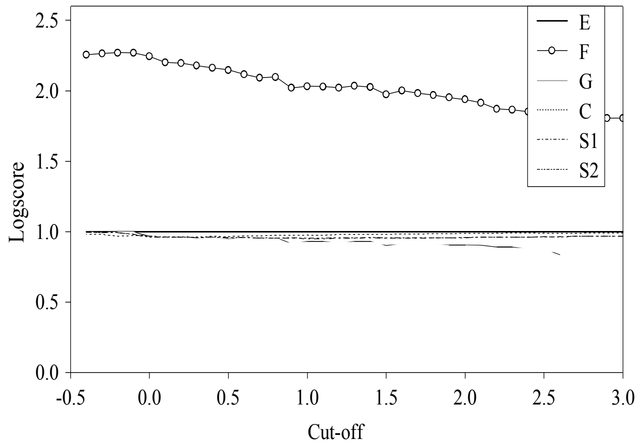


Fig 2a: Case A one-out results

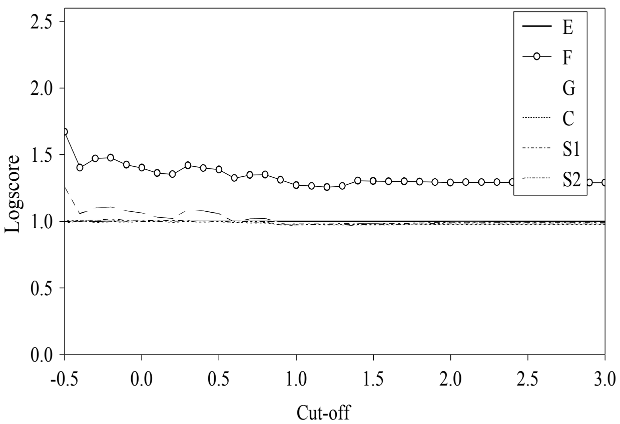


Fig 2b: Case B one-out results

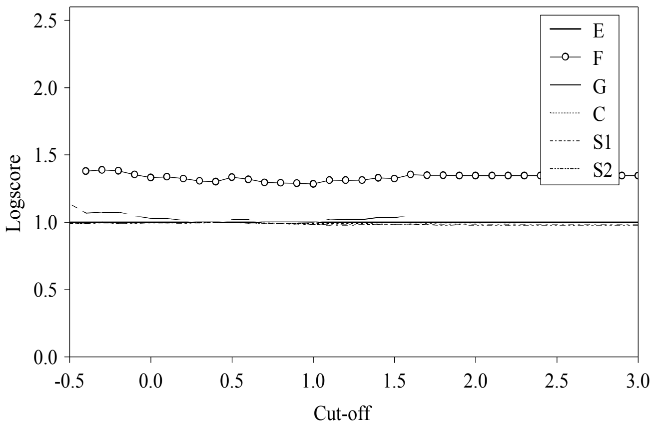


Fig 2c: Case C one-out results

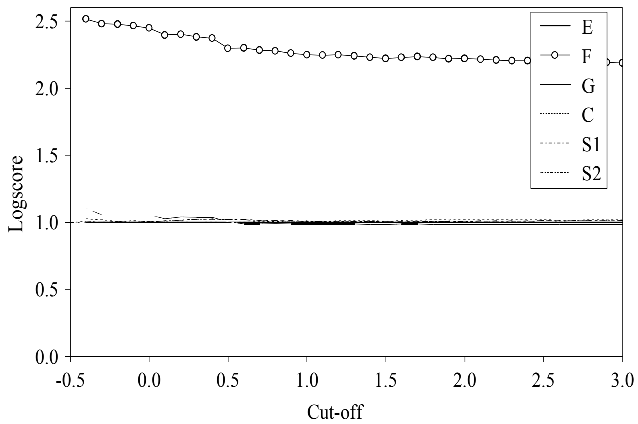


Fig3a: Case A one-on results

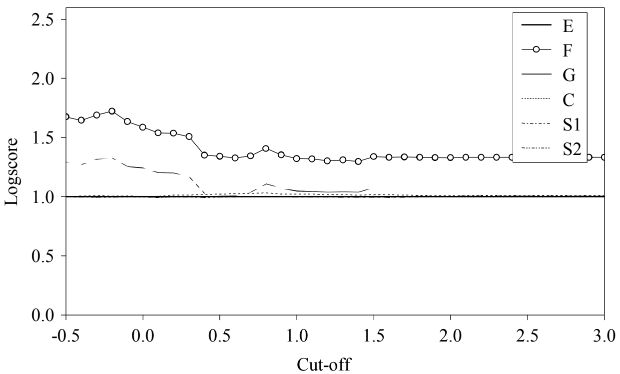


Fig 3b: Case B one-on results

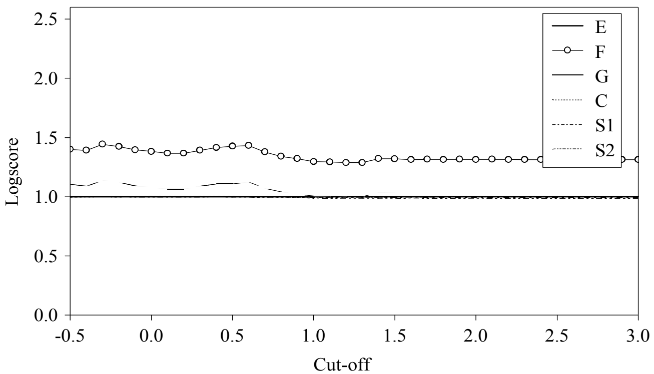
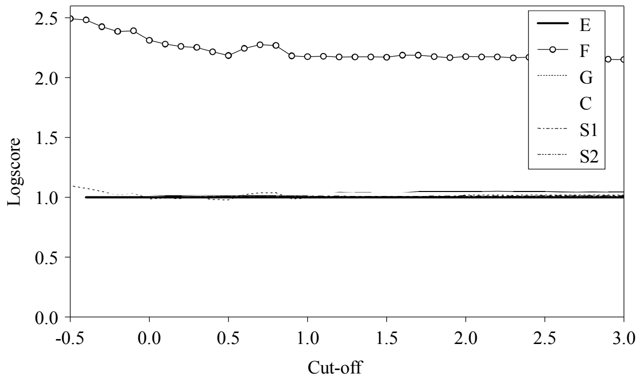


Table 3c: Case C one-on results



This is an Accepted Manuscript of an article published in 2004 by Taylor & Francis in Construction Management and Economics,
available online: https://doi.org/10.1080/0144619042000186103