Standard deviation of bids for construction contract auctions

Pablo Ballesteros-Pérez, Ph.D.[1*] ; Martin Skitmore, Ph.D.[2] ; Alberto Cerezo-Narváez, Ph.D.[3] ;

Mª Carmen González-Cruz, Ph.D.[4] ; Andrés Pastor-Fernández, Ph.D.[5] ; Manuel Otero-Mateo, Ph.D.[6]

Abstract

Previous research has confirmed that the *distribution* of bids for construction auctions can be reasonably modelled with the Lognormal distribution. The location parameter of this distribution (the mean μ) has been found to have a good linear correlation with the bidders' *cost estimates*. However, the scale parameter (standard deviation of the bids, σ) remains noticeably difficult to anticipate.

By analyzing 13 construction auction datasets, hard evidence is provided that the high variability of σ observed in construction *auctions* is mostly due to *sample* size (number of bids per auction). Moreover, we show that the coefficient of variation (σ/μ) of log-transformed bids follows the same χ^2 distribution in *uncapped auctions*. This means the σ's *population* value in similar auctions is nearly *proportional to* μ provided the bid price is not upper limited. Other findings are that more frequent bidders do not tend to bid lower, but their dispersion is narrower than sporadic bidders. These findings allow the introduction of important simplifications in construction bidding models, especially when access to historical data is limited.

Keywords

Construction contracts; auctions; bids; tendering; dispersion; bid modelling; bid forecasting.

[1*] **Corresponding author**: Universitat Politècnica de València. Camino de Vera s/n, 46022 Valencia (Spain) pablo.ballesteros.perez@gmail.com

[2] Queensland University of Technology, Brisbane city QLD 4000 (Australia), rm.skitmore@qut.edu.au

[3] Universidad de Cádiz, Avda. Universidad de Cádiz nº10, Puerto Real, 11519 Cádiz (Spain) alberto.cerezo@uca.es

[4] Universitat Politècnica de València. Camino de Vera s/n, 46022 Valencia (Spain) mcgonzal@dpi.upv.es

[5] Universidad de Cádiz, Avda. Universidad de Cádiz nº10, Puerto Real, 11519 Cádiz (Spain) andres.pastor@uca.es

[6] Universidad de Cádiz, Avda. Universidad de Cádiz nº10, Puerto Real, 11519 Cádiz (Spain) manuel.otero@uca.es

Introduction

Public tendering and procurement are essential in most economies as they give competitive bidders the opportunity to secure public contracts (Bergman and Lundberg 2013). In the construction domain, tendering takes the form of a *reverse auction* (Ahmed et al. 2016) in which bids are offers for contracts made by interested contractors to carry out some construction-related work. Irrespective of the inclusion of other technical or *non-price features*, these bids always involve an economic offer (Ballesteros-Pérez et al. 2015d).

Due to its implications in competitive markets, the study of economic bids has been subject to extensive research (Runeson and Skitmore 1999). Statistical bidding models have traditionally been among the most popular, as these are capable of handling risk and uncertainty. They also enable a potentially substantial amount of theoretical knowledge to be applied to real-life bidding problems (Skitmore 1986). For instance, they can be used by contractors to increase their competitiveness and/or profits [e.g. (Carr 1982; Friedman 1956; Skitmore 1991)]; applied in avoiding collusion by law enforcement agencies [e.g. (Bajari and Ye 2003; Ballesteros-Pérez et al. 2013b, 2015c; Signor et al. 2019, 2020a)]; and in designing improved awarding criteria by contracting authorities [e.g. (Ballesteros-Pérez et al. 2015d, 2016b; Bergman and Lundberg 2013)].

All these applications share the need to model bids as statistical distributions and imply some conditions of stability across auctions and/or bidders' behavior (Yuan 2011). The complete specification of many such distributions involves three parameters: usually referred to as shape, location, and spread (Skitmore 1986). Regarding *shape*, extant studies have proposed or assumed many distributions for modelling bids (e.g. Uniform, Normal, Weibull, Lognormal, and Gamma) (Skitmore 2014). However, Ballesteros and Skitmore (2017), in performing an extensive empirical study, demonstrated that the Lognormal distribution offers the best fit in most situations. Hence, this distribution should be the first choice when modelling the set of bids submitted by different bidders to a single auction. The set of bids submitted by a *single bidder to different auctions* has also been demonstrated empirically to be well modelled by Lognormal distributions (Ballesteros-Pérez and

Skitmore 2017) and by two-parameter Gamma distributions (Skitmore 2014), although each bidder's distribution parameter values may differ (Skitmore 1991, 2014).

Regarding the *location* of the distribution of bids, we generally refer to it as the *expected value*, mean bid, or average bid. Some studies also refer to it as the market fair price (Benton and McHenry 2010). Due to differences across countries, tender specifications, markets conditions, or economic periods, construction contract bids usually include a varying amount of *mark-up* for profit (Skitmore and Pemberton 1994). However, we are not interested here in inferring the profits from the submitted bids, but mostly in predicting the expected value of the bids. In this respect, there have been several empirical studies identifying a *strong mathematical correlation between the bidders' cost estimates and the average bid of a past auction* (Ballesteros-Pérez et al. 2012a; Ballesteros-Pérez and Skitmore 2017). Figure 1 shows an example of this mathematical relationship in grey-colored dots and grey dashed regression lines. The data from Figure 1 are from one of the auction datasets described later, but this strong regression relationship holds in other datasets too. Moreover, the relationship is usually stronger − the datapoints are located nearer the regression line − when bid values are transformed into their logarithmic values (see bottom graph of Figure 1).

<Insert Figure 1 here>

However, the bid *spread* (also named the *standard deviation* or *dispersion* of the bids) often behaves rather erratically in varying significantly even between similar auctions (Skitmore 2001). Skitmore's (1986) first attempts to predict the spread with three datasets found that such estimates were not readily available. Having attempted several variance stabilizing transformations with little success, the main problem was that the tests for homoscedasticity (i.e. the assumption of a constant variance of bids across auctions) were strongly dependent on the bid values being Normally distributed. Similarly, subsequent work by Ballesteros-Pérez et al. (2013a, 2016b) concluded that no obvious mathematical expression or regression relationship could anticipate bid spread values from other auction parameters with sufficient accuracy to be of any practical use.

The green dots in Figure 1 provide an illustrative example of this problem. In the top graph (in natural scale), it is difficult to appreciate the different orders of magnitude this parameter can take. But in the bottom graph (in log scale), it is easy to appreciate the strong level of variation of the green dots even for auctions with a similar mean bid. This high variability has also been found in multiple accounts of the classical construction bidding literature. In them, akin studies have reported bid coefficients of variation (ratio of the standard deviation of the bids to their mean) from around 2% (Morin and Clough 1969) to 15% (Fine and Hackemar 1970), including many values in between: 4% (Beeston 1982), 7.5% (Gates 1967), 10% (Rubey and Milner 1966), etc.

In view of this, no studies have considered that the standard deviation of bids could actually *be the same for similar auctions*. However, it is worth noting that the confidence intervals (CI) of the bids *sample* standard deviation (SD) are very sensitive to low sample sizes (Gurland and Tripathi 1971). For example, for an auction with just two bidders (N=2), the chi-square distribution (χ^2) that models the variability of the SD has 1 degree of freedom (df=N-1) (Lancaster 1971). The result is that a 95% CI of the SD ranges from 0.45×SD to 31.9×SD! In an auction with more bidders, for example N=10 bidders, there are 9 degrees of freedom for estimating the SD. In this case a 95% CI ranges from 0.69×SD to 1.83×SD. Hence, even with a sample (auction) of 10 bidders, the standard deviation of the *population* bids can still be almost 85% higher or 30% lower than the SD of the *sample* bids.

Therefore, the first objective of this study is to check whether the high variability observed in the standard deviation of the sample bids is due to a low sample size (low number of bids per auction). That is, whether a common constant or maybe proportional standard deviation of the *population* bids exists for similar auctions. The second objective involves analyzing if individual bidders' bids also share the same distribution parameters with each other, especially regarding their dispersion, which has been much less studied in the literature. For achieving both objectives, we will analyze the bid dispersion (scale) parameters of 13 representative construction datasets from four continents with various characteristics (auction types, countries, time periods, nature of works, etc.)

Literature review

Bidding studies since Friedman (1956) assume that "by keeping a record of the competitors' past bids, it is possible to evaluate their bidding habits". In the same vein, McCaffer and Pettitt (1976) pointed out that there is substantial evidence that bidding processes are much more than purely random. Hence, a bidder should be able to model the bidding behavior of competitors by tracking their bids and, in turn, use historical data for analyzing past bids and/or predicting future bids. As a result, most classical bidding models are built from the archival information of the bids of competitors [e.g. (Carr 1982; Friedman 1956; Gates 1967; Mercer and Russell 1969; Pim 1974; Wade and Harris 1976)]. These bids are generally stored as ratios (each bid divided by the cost estimate of the bidder holding the data – often referred to as "the reference" bidder) (Stark and Rothkopf 1979). By analyzing these ratios, the reference bidder is theoretically capable of calculating the probability of underbidding its competitors in a future auction and, hence, being awarded the contract.

However, this approach has important limitations. First, as Friedman (1956) also observes, the reference bidder needs a sufficient number of previous bids of a bidder to provide an accurate representation of its behaviors (Friedman suggested at least 30 bids). For construction contract auctions, where the auction-bidder matrix is invariably over 90% sparse (Skitmore 2014), it is difficult, if not impossible, to gather such an amount of information for every competing bidder. Second, it is usually difficult to anticipate which bidders will submit (or not) a bid for an upcoming auction (Ballesteros-Pérez et al. 2016a). Third, as the reference bidder needs to calculate its cost estimates for all (or nearly all) previous auctions to calculate the bid to cost ratios, these are generally only available when the reference bidder participated in those auctions. Overall, these limitations pose a significant challenge regarding the amount of information that can be realistically gathered and converted into *actionable* information.

With the intention of removing some of these barriers, other bidding-related models have resorted to alternative strategies when dealing with the variability of bids. Skitmore (1991), for example, proposes a bidding model comprising a location and scale parameter for each bidder and a location parameter for each auction to empirically disavow the general applicability of the homogeneity assumption (that the two bidder parameter values are not significantly different between bidders). Ballesteros-Pérez et al. (2013a) propose a bid forecasting model where statistical distributions represented the lowest, average, and maximum bids, instead of individual bids – although this simplification ignores the influence of the number of bidders, which leads to insufficiently accurate results. Similarly, many multivariate regression models have also been proposed to anticipate the likely range of bidders' bids [e.g. (Brocas et al. 2015; Lan Oo et al. 2007; Williams 2003)] − these generally resort to multiple parameters (e.g. project size, location, client, nature of works, etc.) as independent variables. However, the latter only provide deterministic estimates, which do not allow an exhaustive analysis that considers uncertainty and risk factors. More recently, other researchers have started to apply machine learning and artificial algorithms to model winning bids from incomplete auction datasets (for instance, datasets where only the lowest bid and pre-tender estimates are available) (García-Rodríguez et al. 2019; 2020). However, these algorithms are also extremely data intensive.

Alternatively, other models have tried to break down the bid modelling problem into smaller chunks with more manageable scopes. In this regard, some attempts have been made to anticipate the total number of potential bidders who might submit a bid in an upcoming auction (Ballesteros-Pérez et al. 2015b). Some models focus on anticipating the identities of specific participating bidders (Ballesteros-Pérez et al. 2016a). Others anticipate only *the number of new bidders* (bidders from which there is as yet no previous information), as well as the size of the bidders' population (market size) (Ballesteros-Pérez et al. 2019; Ballesteros-Pérez and Skitmore 2016). Finally, models measuring the performance (effectiveness) of some bidders from past auctions have also been proposed (Ballesteros-Pérez et al. 2014, 2015a). However, all these models are fragmented and empirical by nature. That is, they suffer from the usual problems associated with the lack of theoretical development, in being simply practical tools incapable of producing generalizable results.

Consequently, there is significant room for improvement in theory-based bidding models. But these improvements must also overcome some of the three limitations stated earlier (information demand, need to anticipate the identities of future likely bidders, and/or dependence on the availability of bidders' cost estimates). One approach to this is to anticipate the standard deviation of bids, as an improved understanding of this parameter could enable most bidding models to be significantly simplified. The first and most obvious response is to model auction bids as being randomly generated from a (lognormal) distribution whose parameters (location and scale) are known. Another option is to simplify bidding models by expressing the bid ratios as a function of their respective auction's mean bid (instead of another bidder's cost estimate). In fact, since there is usually a strong regression relationship between the bidders' cost estimates and the mean bid, only a few past datapoints of previous auctions' cost estimates would suffice to provide a good estimate of a future auction's mean bid. Then, we could study separately the bidders' bids dispersion around that mean bid, thanks to the bid-to-mean-bid ratios. These ratios are much easier to calculate from previous auctions, no matter if there is no cost estimate available. Hence, the amount of actionable information would be much less limiting, and the bidding models that handle it much simpler.

However, before resorting to these alternative bidding analysis strategies, it is necessary to ensure that our estimates of the standard deviation are sufficiently accurate. To do this, it is important to understand the level of variation of bids and anticipate its (population) value.

Research methods

This section first describes the auction datasets used in the analysis. Then, the mathematical transformations are described that are performed on the bid standard deviations to enable them to be compared irrespective of the contract size.

Datasets of auctions

To draw valid conclusions, 13 extensive and representative auctions datasets are used. The characteristics of these datasets are summarized in Table 1. Access to the raw data of all datasets is possible via the *supplemental online material* and from the original sources stated in the second column of Table 1 (Ballesteros-Pérez et al. 2012a, 2015a; Ballesteros-Pérez and Skitmore 2017; Brown 1986; Drew 1995; Fu 2004; Runeson 1987; Shaffer and Micheau 1971; Skitmore 1991, 1981, 1986).

<Insert Table 1 here>

The datasets are deemed representative as they contain bidding data from five countries (United Kingdom, United States of America, Hong Kong, Australia, and Spain) and four continents (Europe, America, Asia, and Oceania). The number of contracts (auctions) of each dataset is never fewer than 45, even after removing auctions with less than two bidders (the minimum needed to calculate the bid standard deviation). The sources of these datasets are published papers and/or dissertations. This allows for replicability by other researchers.

As shown in the 'Description' column, the nature of the works is quite varied (different types of buildings and different types of civil engineering works). Their time span is also quite wide, with the earliest being from 1965 and the latest from 2014. Similarly, some datasets span 2 years while others range up to 10 years. The latter feature allows for some longitudinal comparisons within the same dataset (e.g. considering different market periods, even the potential impact of economic crises).

Regarding the number of bidders (see column 'Avg. Nº bids/auction'), the datasets have auctions with small and large numbers of bidders (from 5 to 30 bids per auction on average). This is convenient for identifying the possible impact of the auctions' sample size on the bid standard deviation. The first eight datasets also include information concerning which bidder submitted each bid, that is, the bidders' identities (see column "Bidders' ID"). This information enables an analysis to be made on the differences between the bidding outcomes of more frequent *vs* sporadic bidders. Additionally, some datasets include the reference bidder's or project designer's cost estimates of most (sometimes all) auctions.

Finally, the last column of Table 1 indicates which auction datasets correspond to uncapped auctions (those where bidders have no price ceiling when submitting their bids) or capped auctions (those where bidders must necessarily underbid a given price generally made public beforehand by the contracting authority). As shown later, the difference between capped and uncapped tenders is relevant to the bid standard deviation similarity across auctions. The HK159 and HK259 datasets are classified as "mixed": they involve Class A contractors that are only eligible to bid up to HKD 3 million, Class B who can bid up to HKD 15 million, and Class C who can bid any value. This means that the Class A and B contractors' bids can be considered to be capped to some extent, while Class C bids are uncapped. That is, although most auctions in these two datasets are uncapped, some of their bids can be considered to be capped.

Analysis

As introduced earlier in Table 1, each dataset contains contracts (auctions) with different types of works. For example, the first dataset (UK51) encompasses building-related auctions. Some of these contracts involve the construction of a building, some the building design, and some maintenance and/or repair activities. Yet, for the purpose of this study, it is assumed that the contracts within each dataset are relatively homogeneous, i.e. their scope, project client, and geographical area are relatively similar. Even though this may not be true in some datasets, especially in datasets spanning long time periods, it is noted that, should the findings hold under these restrictive conditions, they will also hold in most real-life auction settings.

Each dataset is also analyzed separately: this is clearly necessary, as the data are from different countries, types of work, and time periods. Since contracts from each dataset have different economic sizes, and akin to previous research on bid dispersion (Skitmore 1981), the Coefficient of Variation (CV) − a standardized measure of dispersion calculated as the standard deviation of the bids divided by their mean − is used as a substitute for the bid standard deviation. However, a CV needs a stable base of comparison. In this case, the denominator of the CV ratio (the mean of the bids) is not reliable as the statistical distribution of the bids is generally not symmetrical. As discussed earlier, previous research has proven that the distribution of bids in construction auctions can be reasonably represented with Lognormal distributions (Ballesteros-Pérez and Skitmore 2017). Hence, the *logarithm* of the bids are analyzed instead of their *natural* (monetary) value. With this approach, the distribution the log bids then becomes approximately symmetrical and the CV is a better (dimensionless) representation of the bids dispersion.

Before continuing, we introduce some basic notation to understand the upcoming calculations:

μ, σ are the (unknown) *log* bids *population* mean and standard deviation, respectively.

m_j, s_j are the *log* bids *sample* mean and standard deviation, respectively, for auction j. That is, these are the values we observe of μ and σ in each auction j.

CV_j is the *sample* coefficient of variation of the log-transformed bids in auction j, i.e. $CV_j=s_j/m_j$.

$b_{(i)j}$ is the i^{th} lowest *log* bid in auction j. (e.g. $b_{(2)3}$ is the 2nd lowest log bid in the 3rd auction in the dataset).

b_{kj} is bidder k's *log* bid in auction j. Here, k refers to the *identity* of bidders, not their *position*.

N_{ij} is the number of bids in auction j.

N_j is the number of auctions in the dataset.

N_k is the number of different bidders (identities) in the dataset.

To keep the notation as simple as possible, additional subscripts are not used to refer to each of the 13 datasets, nor to refer to natural (instead of log) bids.

As anticipated and shown earlier in Figure 1, the standard deviations of both the *natural* and *log* bids are highly variable, and this variability remains after calculating the coefficient of variation CV_j of each auction. The question now is: (1) is this variability the result of each auction having unique characteristics and, hence, each auction having a different *population* standard deviation σ? Or, alternatively, (2) are the characteristics of each auction sufficiently similar that the variability of the observed s_j values are simply due to *sampling errors* (but they all share the same σ value)?

To provide an answer, we need to check whether the s_j values of each dataset (now expressed as CV_j) are from the same statistical distribution with the same parameter values. If this is the case, then question (2) can be regarded as correct. This would also imply that the *population* standard deviation of each auction is approximately proportional to its mean. This verification is simple, but not evident. Indeed, this had never been tested in the construction bidding domain, nor in other industries.

If auction *j*'s *log* bids *sample* standard deviation is given by

$$s_j = \sqrt{\frac{\sum_1^{N_{ij}} \left(b_{(i)j} - m_j\right)^2}{N_{ij} - 1}} \tag{1}$$

Then, the CV of the log bids of auction *j* is:

$$CV_j = \frac{s_j}{m_j} = \sqrt{\frac{\sum_1^{N_{ij}} \left(b_{(i)j} - m_j\right)^2}{m_j^2 \cdot (N_{ij} - 1)}} = \sqrt{\frac{\sum_1^{N_{ij}} \left(\frac{b_{(i)j}}{m_j} - 1\right)^2}{N_{ij} - 1}} \tag{2}$$

This is the best estimate of the *population* CV for auction *j*. It is known that a Chi-square (χ^2) distribution represents the distribution of the sum of squares of *n* independent standard normal random variables (Normal distribution with mean=0 and st. dev.=1) (Bartlett and Kendall 1946). It can also be demonstrated that the sum of squares of *n* independent standard normal random variables X_i minus their mean $\bar{X}$ follow a Chi-square distribution with *n*-1 degrees of freedom (Lancaster 1971):

$$\sum_1^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2 \tag{3}$$

In our case, X_i corresponds to the auction *j*'s log bids, that is, $b_{(i)j}$; whereas *n* corresponds to the number of bids in auction *j*, that is, N_{ij}. Now, if a unique *population* standard deviation σ exists and is common to all auctions in a dataset, so should be its coefficient of variation CV= σ/μ . However, by working with the CVs of a symmetrical distribution, we know that μ=1. Hence, the best *estimate* of the *population coefficient of variation* ($\widehat{CV}$) consists of applying expressions (1) and (2) to *all bids in the dataset* instead of to a single auction, i.e.:

$$\widehat{CV} = \sqrt{\frac{\sum_1^{N_j} \sum_1^{N_{ij}} \left(b_{(i)j} - \mu\right)^2}{\mu^2 \left(\left(\sum_1^{N_j} N_{ij}\right) - 1\right)}} \approx \sqrt{\frac{\sum_1^{N_j} \sum_1^{N_{ij}} \left(\frac{b_{(i)j}}{m_j} - 1\right)^2}{\left(\sum_1^{N_j} N_{ij}\right) - 1}} \tag{4}$$

However, construction contract auction datasets usually contain outliers (e.g. Skitmore 2004) – in this case, abnormally high or low bids. These can be from transcription errors, but also from excessively aggressive or conservative bidders (Signor et al. 2020b). In most cases, these bids are not representative of a truly competitive market and must be removed before expression (4) is applied for calculating the $\widehat{CV}$ value of each dataset. Skitmore (2001, 2004) and Skitmore and Lo (2002) suggested several approaches to remove outliers in seeking to find the best distributional shape for construction contract bids. However, Tukey's fences are used here for removing outliers as the distribution of the (log) bids is approximately Normal (Tukey 1977). Namely, bids that fall outside the following range are excluded:

$$[Q_1 - 1.5(Q_3 - Q_1),\ Q_3 + 1.5(Q_3 - Q_1)\,] \tag{5}$$

where Q_1 and Q_3 are the lower and upper quartiles, respectively, of all $b_{(i)j}/m_j$ values in each dataset.

Now, from expressions (3) and (4) it is easily inferred that the probability of obtaining each auction j's bids standard deviation s_j in the same dataset is given by

$$Prob(s_j') = \ CDF\ \chi^2_{N_{ij}-1}\left[x = (N_{ij} - 1)\left(\frac{CV_j}{\widehat{CV}}\right)^2\right] \tag{6}$$

where $Prob(s_j')$ is the quantile (probability) of obtaining each auction j's log bids sample standard deviation s_j; and $CDF\ \chi^2_{N_{ij}-1}$ is the cumulative distribution function of a Chi-square distribution with N_{ij} -1 degrees of freedom evaluated at $x = (N_{ij} - 1)\left(\frac{CV_j}{\widehat{CV}}\right)^2$. The square term appears because the χ^2 distribution actually models the *variance*, not the *standard deviation*.

Hence, expression (6) is applied to all auctions in each dataset to obtain the quantiles of all their s_j values. If they indeed follow the same chi-squared distribution, then they will adhere to a bisector line in a QQ plot (as Figure 2 in the next section shows). Also, it must be noted that only quantiles (probabilities) can be compared here, as each auction follows a χ^2 distribution with a different number of bidders (degrees of freedom).

After taking logs of all the bids, expression (4) is applied to obtain the best estimate of the *population* coefficient of variation ($\widehat{CV}$) of each dataset. Three calculation approaches are used: approach (a) implements expression (4) directly from all log bids without excluding any outliers, and approach (b) excludes outlying log bids according to expression (5) [the number of outliers (bids) removed in each dataset can be inferred by the difference between the 'Nº valid bids' for (a) and (b)]. Approach (c) is used when a dataset only contains the auctions' (sample) mean and standard deviation values without no information on the individual bids. In this case, the *natural* mean (m_j*) and standard deviation (s_j*) values can be converted to their log-equivalent m_j and s_j by:

$$m_j = LN\left(\frac{\left(m_j^*\right)^2}{\sqrt{\left(m_j^*\right)^2+\left(s_j^*\right)^2}}\right) \tag{7}$$

$$s_j = \sqrt{LN\left(1+\frac{\left(s_j^*\right)^2}{\left(m_j^*\right)^2}\right)} \tag{8}$$

where LN(·) is the natural logarithm, and the population estimate of the CV be calculated as:

$$\widehat{CV} = median\{CV_j\} \qquad j=1, 2\ldots N_j \tag{9}$$

Results

The results for three different calculation approaches for population estimate of the CV (noted as $\widehat{CV}$) are shown in Table 2.

<Insert Table 2 here>

Figure 2 shows the QQ plots of the χ^2 distribution quantiles of all auctions' CV_j values in the 13 datasets for the three calculation approaches.

<Insert Figure 2 here>

In graph (a), with no outliers removed, the quantile lines obtained in all datasets depart significantly from the bisector line, which means that the series of auction CV_j values do not follow the same χ^2 distribution, and therefore the auctions in each dataset do not share the same $\widehat{CV}$. Graph (b), with outliers removed, shows much better fitting results: with the exception of the two capped auction datasets (dashed lines). In this case, most curves have a significant adherence to the bisector line. The probability values are lower than in graph (a), indicating that not removing outliers resulted in $\widehat{CV}$ values being overestimated.

Finally, calculation approach (c) using the auctions' CV_j median also shows a good goodness of fit to the bisector line (same χ^2 distribution with varying degrees of freedom). The exceptions in this case are the same two capped auction datasets (SP51 and SP110 in dashed lines) and the two Asian datasets (HK199 and HK266 in dotted lines). However, it is worth remembering that the latter datasets are mixed (contained both capped and uncapped bidders). As a result, both calculation approaches (b) and (c) seem quite satisfactory. However, whenever possible, approach (b) is more appropriate as it seems a little more precise.

However, perhaps it could be argued that this goodness of fit is not remarkable. It must be borne in mind, though, that the auctions of each dataset encompass a wide variety of types of works, economic sizes, and bidders' identities; even auctions up to 10 years apart in many cases. Considering all these sources of variability, the resemblance of the CV_j values to the same χ^2 distribution is indeed quite high.

However, Kolmogorov-Smirnov (K-S) tests have been implemented to provide a numerical assessment of the Chi-square distribution fit to the bids dispersion. K-S fit tests measure the maximum deviation between the actual and theoretical cumulative probabilities (D_{max}) of each dataset. If the probability of occurrence (p-value) of such D_{max} is too high (generally a p-value>95%), then, the null hypothesis is rejected. In our case, the null hypothesis is that a single population CV value exists for all auctions in the (same) dataset. Table 3 shows a summary of the K-S fit tests in the 13 datasets.

<Insert Table 3 here>

As can be seen, in six out of the 12 datasets, the null hypothesis is rejected (p-values>5%). Two of these six cases correspond to the *capped* datasets (SP51 and SP116) which, as expected and shown in Figure 2, deviate substantially from the Chi-square model. However, the four datasets that reject the null hypothesis (UK218, HK266, UK272 and UK537), when analyzed in shorter time spans, also pass the test. In this regard, Table 4 presents another round of K-S tests, but this time with these datasets split in two halves. Assuming shorter time spans (compare the 'Period' column in Tables 3 and 4), it is expected that each sub dataset was subject to lower market volatility. Hence, the auctions contained that in each sub dataset apparently become more homogeneous and eventually pass the test.

<Insert Table 4 here>

Discussion

It can be concluded, then, that the population coefficient of variation of the log bids in *uncapped* auctions is nearly constant whenever the auctions are relatively homogeneous. An estimate of this CV value is what we have called $\widehat{CV}$ and can be approximated by expressions (4) or (9). Hence, once the value of $\widehat{CV}$ is calculated, it is straightforward to anticipate a future auction's log bids standard deviation (σ) by multiplying $\widehat{CV}$ by the forecasted mean (μ) of the log bids. This can be achieved by resorting, for instance, to the usually strong regression relationship between the auction's cost estimate and the mean of the log bids, as exemplified in the grey dots of Figure 1.

However, regarding *capped* auctions, i.e. those in which bidders can only underbid a pre-set maximum price, the same does not hold. This is to be expected as, with this type of auction, if the upper bid price (sometimes called the Pre-Tender Estimate, PTE) is too close (or even below) to what most bidders deem as a competitive bid, then they bid very near the PTE. In these cases, very low bid dispersions are to be expected. The opposite happens when the PTE is much higher than the competitive market price of a contract. Hence, it seems reasonable that the constant CV assumption will *not* hold in capped auctions.

An alternative is to introduce in the analysis of capped auctions another variable that takes into account the (positive or negative) difference between the auction's cost estimate and the PTE to make the calculation of the population $\widehat{CV}$ more accurate. For this type of analysis, though, more construction capped auctions datasets with information of bidders' cost estimates would be necessary, which are extremely difficult to obtain for competitive reasons (bidders seldom share their cost estimates).

Practical relevance

The implications of these findings are plentiful in the construction bidding domain. In bid forecasting, for example, the assumption of a *constant* population coefficient of variation of bids for each auction can lead to simpler bidding models. Specifically, in these simplified models, the coefficient of variation could be treated as a random variable with a fixed mean subject only to random disturbances in its estimation. Such models can be used by contractors to increase their profit margins and/or the probability of being awarded a contract. To date, many bidding models have been too complex for most practical settings (Ballesteros-Pérez et al. 2012a). This had been mostly the result of some of their basic parameters being very difficult to anticipate − σ being eminently among them. With this paper's contributions, these models can be reformulated in simpler mathematical terms and, most importantly, require much less historical data to be operational.

Other applications of the assumption of a constant population coefficient of variation of bids also encompass the potential simplification of current collusion-detection models. Collusion is a widespread phenomenon in which some bidders condition the award of a contract to a previously (and secretly) agreed bidder. This is obviously an unethical and illegal practice, as it undermines the benefits of a competitive market, awards contracts with abnormally high mark-ups, and consumes an excessive amount of resources in its policing and detection. Most law enforcement agencies and contracting authorities usually have difficulty in finding evidence of collusion from simply analyzing auctions results. However, by understanding the reduced variation to be expected in the bids standard deviation, collusion-detection models will be able to provide more reliable *reference scenarios* (Signor et al. 2020a) that describe what a truly competitive set of bids must look like. By establishing comparisons against this reference scenario, it may be easier to identify non-competitive bids and pursue further evidence of criminal activity – at least until such bidders develop their own counter measures (Skitmore and Cattell 2013).

Another application of the assumption of a constant population of the coefficient of variation bids will allow a better design of tender specifications and economic scoring formulae (ESF). ESF are mathematical expressions governing the allocation of the bidders' scores as a function of their economic bids (Ballesteros-Pérez et al. 2012b, 2015d). For example, being able to predict the range in which competitive bids will vary will allow contracting authorities to set more realistic criteria for determining abnormally low bids (e.g. disqualify bids which are 2 or 3 standard deviations below the pre-tender estimate). This is still an ongoing problem when trying to set a cut-off limit that separates truly competitive from reckless bids (Ballesteros-Pérez et al. 2013b, 2015c). Additionally, better ESF should also be able to better distribute the whole range of the economic scoring among all possible bids in an auction while avoiding the phony (economic) bid weighting (Ballesteros-Pérez et al. 2015d). This is a pervasive phenomenon of multi-attribute auctions where both economic and technical aspects are evaluated.

Bidding patterns of individual bidders

Finally, there is the question of whether there are significant bidding behavior differences between bidders. That is, since the coefficient of variation of the log bids is nearly constant in homogeneous uncapped auctions, is this the consequence of individual bidders' bids also having the same dispersion?

<Insert Figure 3 here>

Figure 3 helps answer this question with two graphs taken from the first 8 datasets, as those are the only ones with the bidders' identities known. The top graph describes the evolution of the *average* log bids as we add more bidders' bids. However, to make these bids comparable, each of these bidders' bids have been divided beforehand by their respective auction's log bid mean (that is, we work now with b_{kj}/m_j values). Namely, values around 1 are obtained (which would equal the auctions' log bid mean) irrespective of the size of auctions involved in the calculation.

Additionally, each dataset (represented in one curve each) contains N_k bidders. These are ordered from those who bid most frequently to those who bid less so. This means that, for example, in the X-value N_k=5, the average of the top 5 most frequent bidders' bids ($m_{k=5}$) is being taken.

Analogously, in the bottom graph, the same dimensionless bids are taken but calculating their standard deviation ($s_{k=5}$). However, in this graph, the s_k values are divided by each dataset *population* log bid standard deviation (σ). The value of σ is calculated as $\sigma=\widehat{CV}\cdot\mu$, but, in this case, μ=1, as all bid values are already divided by their respective m_j value. Again, using this ratio allows all s_k/σ values to be compared under the same scale. Moreover, all end in 1 when all bids have been introduced into the calculation. This happens because at $x=N_k$, $s_k = \sigma$.

As the top graph (describing the relative bidders' bids with respect to the auction mean bid) shows, more frequent bidders do not necessarily bid more aggressively. In our analysis, this outcome can be inferred by observing that the Y-values of the curves for the first X-values are sometimes above and sometimes below 1 for different datasets. If more frequent bidders were indeed more aggressive (submitted lower bids), then all curves would depart from a value <1. They would also approach the horizontal X=1 line always from below as we incorporated less frequent bidders' bids in the computation of the average bid. This, as can be easily seen, does not happen in several datasets.

However, in the bottom graph, the bids dispersion of more frequent bidders is lower than the average population dispersion in all datasets (curves). Analogously, this is inferred from all curves remaining below 1 until almost all N_k bidders have been included in the analysis. The only exception, but very succinctly, may be dataset HK199 (green dotted line), but as noted earlier, this dataset contains mixed (capped and uncapped) bidders.

Therefore, Figure 3 prompts the conclusion that bidders who bid more frequently do not necessarily submit lower bids, but instead, their bid dispersion is lower. These results are in line with the results of De Silva et al. (2003). Through a series of first-price sealed bid road auctions, they found that *entrants* (those who compete for a contract) generally submit lower (more aggressive) bids than the incumbents (bidders who are already performing the contract). However, this phenomenon does not happen because entrant firms are more efficient, but because their costs evidence a higher dispersion than the incumbents'. Hence, it is likely that one of the entrants (the one with the most relevant cost items being incidentally lower than the incumbent's) eventually wins the auction. More recently, Camboni and Valbonesi (2020) found that bid prices offered by incumbents are also frequently higher than the entrants' lowest bid. Paradoxically, this outcome could not be predicted neither from the contract, nor the auction characteristics.

Then, a lower bids dispersion seems to be the consequence of more frequent bidders knowing their market segment and clients better and/or producing more accurate contract cost estimates. This is what some researchers have coined as superior market-price alignment (Skitmore 1987). This lower bid dispersion may also be the result of a more consistent bidding strategy in the form of ranges or bidding mark ups more focused in the medium/long term rather than in the short term. In this vein, De Silva et al. (2003) also showed that bidders with more backlog usually bid less aggressively.

Hence, the coefficient of variation of log bids is nearly constant in homogeneous uncapped auctions, but it is not for individual bidders' bids; that is, each bidder has its own bid distribution. How is it possible, then, that the sets of bids from different auctions have the same coefficient of variation? The only possible explanation is that the proportion of veteran versus novice bidders across auctions is approximately constant. Veteran (frequent) bidders have lower bidding dispersions, whereas novice (sporadic or new) bidders have higher dispersions. As less frequent bidders continue to submit more bids, they keep narrowing their bids dispersion. But new bidders will also keep arriving and counteract the overall auction bids dispersion. This is a dynamic process in which, as the data evidences, bids dispersion maintains an approximately constant balance.

Yet, we can observe that the bidding dispersion from the most veteran to more sporadic bidders is not that big (between 5-20 % lower in a log scale). This means that, in bid forecasting and analysis models, the error derived from assuming that all bidders have the same bids dispersion (equal to the auction bids dispersion) will be relatively small.

Conclusions

Previous research has confirmed that the distribution of bid values in construction auctions can be reasonably approximated with Lognormal distributions. Common Lognormal distributions have two parameters: the mean (μ) and the standard deviation (σ). μ is known to have a good log linear correlation with the bidders' cost estimates. Hence, even counting only on a limited dataset of previous auctions, it should be easy to infer a good μ estimate from the future auction's cost estimate.

However, no studies to date have proposed a mathematical expression to anticipate the standard deviation (σ) of log bids from other auction variables. In particular, the *sample* standard deviation values of a set of homogeneous auctions seemed very erratic and not to follow any predictable pattern.

In the present study, we provide hard empirical evidence that the population coefficient of variation (the σ/μ ratio) of log bids for each auction is approximately constant for homogeneous *uncapped* auctions − those in which bidders can submit their bids without an upper price limitation. Homogeneous auctions refer here to those that share a similar nature of works, project client, and geographical proximity. With this type of auction, the high variation observed in the auctions' sample standard deviations, even in very similar auctions, is the consequence of low sample sizes (number of bidders). Namely, most construction contract auctions usually have a low number of bidders (<15). This number of datapoints (bids) is frequently insufficient to produce a good estimate of the (population) standard deviation from a single or few auctions.

In analyzing a wide and representative set of 13 auction datasets from four continents and different time periods, we have proposed two calculation approaches of this nearly constant coefficient of variation (noted here as $\widehat{CV}$). One of them – the more accurate – requires all the bidders' bids, whereas the second can produce a reasonable estimate of the coefficient of variation whenever only the mean and standard deviation values of the auction bids are available. In implementing both approaches, it is concluded that all auctions' bid standard deviation values follow the same chi-square (χ^2) distribution with varying degrees of freedom − implying that the population coefficient of variation of the log bids for each auction across homogeneous auctions can be regarded as nearly constant, the recorded variability being accounted for as random sampling error.

Additionally, in comparing the performance of more versus less frequent bidders through an analysis of the mean and dispersion values of their bids, it is concluded that more frequent bidders do not necessarily bid more aggressively (submit lower bids) than sporadic bidders. Instead, they usually evidence a lower bids dispersion (their bids variation around the bid average is narrower). Yet, it has been shown that this dispersion is not usually lower than 80% to 95% of the population bids standard deviation. This means that most bidding models that differentiate by bidders' identities when forecasting the lowest bid may not incur in great inaccuracies by assuming that all bidders (frequent and new alike) follow the same μ and σ parameters.

Finally, it is acknowledged that the number of capped auction datasets has not been sufficient to delve into the additional complexities of capped tendering. For example, there might be a way of adding a correction coefficient of the population estimate of the coefficient of variation of log bids ($\widehat{CV}$), which takes into account the relative distance between the pre-tender estimate and the (forecasted) mean log bid. However, for such analysis, a larger number of capped auction datasets with information of the contracts' bidders' cost estimates would be necessary, but, as information relating to cost estimates are usually difficult to obtain from bidders for competitive reasons, this analysis remains pending for future research.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

The first author acknowledges the Spanish Ministry of Science and Innovation for his Ramon y Cajal contract (RYC-2017-22222) co-funded by the European Social Fund.

Supplemental Data

The 13 construction auction datasets are available online in the ASCE Library (www.ascelibrary.org).

References

Ahmed, M. O., El-adaway, I. H., Coatney, K. T., and Eid, M. S. (2016). "Construction Bidding and the Winner's Curse: Game Theory Approach." *Journal of Construction Engineering and Management*, American Society of Civil Engineers (ASCE), 142(2), 04015076.

Bajari, P., and Ye, L. (2003). "Deciding between competition and collusion." *Review of Economics and Statistics*, 85(4), 971–989.

Ballesteros-Pérez, P., del Campo-Hitschfeld, M. L., Mora-Melia, D., and Domínguez, D. (2015a).

"Modeling bidding competitiveness and position performance in multi-attribute construction auctions." *Operations Research Perspectives*, 2, 24–35.

Ballesteros-Pérez, P., González-Cruz, M. C., and Cañavate-Grimal, A. (2012a). "Mathematical relationships between scoring parameters in capped tendering." *International Journal of Project Management*, 30(7), 850–862.

Ballesteros-Pérez, P., González-Cruz, M. C., and Cañavate-Grimal, A. (2013a). "On competitive bidding: Scoring and position probability graphs." *International Journal of Project Management*, 31(3), 434–448.

Ballesteros-Pérez, P., González-Cruz, M. C., Cañavate-Grimal, A., and Pellicer, E. (2013b). "Detecting abnormal and collusive bids in capped tendering." *Automation in Construction*, 31, 215–229.

Ballesteros-Pérez, P., González-Cruz, M. C., Fernández-Diego, M., and Pellicer, E. (2014). "Estimating future bidding performance of competitor bidders in capped tenders." *Journal of Civil Engineering and Management*, 20(5), 702–713.

Ballesteros-Pérez, P., González-Cruz, M. C., Fuentes-Bargues, J. L., and Skitmore, M. (2015b). "Analysis of the distribution of the number of bidders in construction contract auctions." *Construction Management and Economics*, 33(9), 752–770.

Ballesteros-Pérez, P., González-Cruz, M. C., Pastor-Ferrando, J. P., and Fernández-Diego, M. (2012b). "The iso-Score Curve Graph. A new tool for competitive bidding." *Automation in Construction*, 22.

Ballesteros-Pérez, P., and Skitmore, M. (2016). "Estimating the number of new and repeated bidders in construction auctions." *Construction Management and Economics2*, 34(12), 919–934.

Ballesteros-Pérez, P., and Skitmore, M. (2017). "On the distribution of bids for construction contract auctions." *Construction Management and Economics*, 35(3), 106–121.

Ballesteros-Pérez, P., Skitmore, M., Das, R., and del Campo-Hitschfeld, M. L. (2015c). "Quick Abnormal-Bid–Detection Method for Construction Contract Auctions." *Journal of Construction Engineering and Management*, 141(7), 04015010.

Ballesteros-Pérez, P., Skitmore, M., Pellicer, E., and González-Cruz, M. C. (2015d). "Scoring rules and abnormally low bids criteria in construction tenders: a taxonomic review." *Construction Management and Economics*, 33(4), 259–278.

Ballesteros-Pérez, P., Skitmore, M., Pellicer, E., and Gutiérrez-Bahamondes, J. H. (2016a). "Predicting the identity of bidders in procurement auctions." *International Journal of Project Management*, 34(2), 158–172.

Ballesteros-Pérez, P., Skitmore, M., Pellicer, E., and Zhang, X. (2016b). "Scoring rules and competitive behavior in best-value construction auctions." *Journal of Construction Engineering and Management*, 142(9), 04016035.

Ballesteros-Pérez, P., Skitmore, M., Sanz-Ablanedo, E., and Verhoeven, P. (2019). "Forecasting the Number and Distribution of New Bidders for an Upcoming Construction Auction." *Journal of Construction Engineering and Management*, 145(10), 1–10.

Bartlett, M. S., and Kendall, D. G. (1946). "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation." *Supplement to the Journal of the Royal Statistical Society*, 8(1), 128–138.

Beeston, D. T. (1982). *Estimating market variance in building cost techniques: new directions*. P.S. Brandon. E & F.N. Spon. ISBN 0-419-12940-5 p2.

Benton, W. C., and McHenry, L. F. (2010). *Construction Purchasing and Supply Chain Management*. *ISBN 978-0-07-154886-1*, McGraw-Hill Education.

Bergman, M. a., and Lundberg, S. (2013). "Tender evaluation and supplier selection methods in public procurement." *Journal of Purchasing and Supply Management*, Elsevier, 19(2), 73–83.

Brocas, I., Carrillo, J. D., and Castro, M. (2015). "The nature of information and its effect on bidding behavior: Laboratory evidence in a first price common value auction." *Journal of Economic Behavior & Organization*, Elsevier B.V., 109, 26–40.

Brown, K. C. (1986). *A theoretical and statistical study of decision making under uncertainty – competitive bidding for leases of offshore petroleum tracts. PhD dissertation, Southern Methodist University*.

Camboni, R., and Valbonesi, P. (2020). "Incumbent and Entrant Bidding in Scoring Rule Auctions: A Study on Italian Canteen Services." *B.E. Journal of Economic Analysis and Policy*, (ahead of print). doi: 10.1515/bejeap-2019-0146.

Carr, R. I. (1982). "General Bidding Model." *Journal of the Construction Division*, ASCE, 108(4), 639–650.

Drew, D. S. (1995). *The effect of contract type and size on competitiveness in construction contract bidding. PhD thesis, Department of Surveying, The University of Salford*.

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), 104–12.

Fu, W. (2004). *The effect of experience in construction contract bidding. PhD thesis, Department of Building and Real Estate, Hong Kong Polytechnic University*.

García-Rodríguez, M. J., Rodríguez-Montequín, V., Ortega-Fernández, F., and Villanueva-Balsera, J. M. (2019). "Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning." *Complexity*, 2360610.

García Rodríguez, M. J., Rodríguez Montequín, V., Ortega Fernández, F., and Villanueva Balsera, J. M. (2020). "Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain." *Complexity*, (T. C. Silva, ed.), 2020, 1–20.

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

Gurland, J., and Tripathi, R. C. (1971). "A Simple Approximation for Unbiased Estimation of the Standard Deviation." *The American Statistician*, 25(4), 30–32.

Lan Oo, B., Drew, D., and Lo, H.-P. (2007). "Modelling contractors' mark-up behaviour in different construction markets." *Engineering, Construction and Architectural Management*, 14(5), 447–462.

Lancaster, H. O. (1971). "The Chi-squared Distribution." *Biometrische Zeitschrift*, Wiley, 13(5), 363–364.

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

Mercer, A., and Russell, J. I. T. (1969). "Recurrent Competitive Bidding." *Operational Research Quarterly*, 20(2), 209–221.

Morin, T. L., and Clough, L. H. (1969). "OPBID: Competitive bidding strategy model." *Journal of the Construction division ASCE*, 95, 85–106.

Pim, J. C. (1974). "Competitive tendering and bidding strategy." *National Builder*, 55, 541–545.

Rubey, H., and Milner, W. W. (1966). *A Statistical approach to bidding. Ed. McMillan*. Pages 229-265.

Runeson, G., and Skitmore, M. (1999). "Tendering theory revisited." *Construction Management and*

Economics, Taylor & Francis Group, 17(3), 285–296.

Runeson, K. G. (1987). *Analysis of building price estimates. MSc thesis, School of Building, University of New South Wales*.

Shaffer, L. R., and Micheau, T. W. (1971). “Bidding with competitive strategy models.” *Journal of the Construction division ASCE*, 97, 113–126.

Signor, R., Love, P. E. D., Belarmino, A. T. N., and Alfred Olatunji, O. (2020a). “Detection of Collusive Tenders in Infrastructure Projects: Learning from Operation Car Wash.” *Journal of Construction Engineering and Management*, 146(1), 05019015.

Signor, R., Love, P. E. D., Marchiori, F. F., and Felisberto, A. D. (2020b). “ Underpricing in Social Infrastructure Projects: Combating the Institutionalization of the Winner’s Curse .” *Journal of Construction Engineering and Management*, American Society of Civil Engineers (ASCE), 146(12), 05020018.

Signor, R., Love, P. E. D., Vallim, J. J. C. B., Raupp, A. B., and Olatunji, O. (2019). “It is not collusion unless you get caught: The case of ‘Operation Car Wash’ and unearthing of a Cartel.” *Journal of Antitrust Enforcement*, 7(2), 177–202.

De Silva, D. G., Dunne, T., and Kosmopoulou, G. (2003). “An empirical analysis of entrant and incumbent bidding in road construction auctions.” *Journal of Industrial Economics*, 51(3), 295–316.

Skitmore, M. (1981). “Why do tenders vary?” *The Chartered Quantity Surveyor*, 4, 128–129.

Skitmore, M. (1986). *A model for the construction project selection and bidding decision. PhD thesis, Department of Civil Engineering, University of Salford*.

Skitmore, M. (1991). “The contract bidder homogeneity assumption: an empirical analysis.” *Construction Management and Economics*, 9(5), 403–429.

Skitmore, M. (2001). “Graphical Method for Identifying High Outliers in Construction Contract Auctions.” *Journal of the Operational Research Society*, 52(7), 800–809.

Skitmore, M. (2004). “Predicting the probability of winning sealed bid auctions: the effect of outliers on bidding models.” *Construction Management and Economics*, 22(1), 101–109.

Skitmore, M. (2014). “Generalised gamma bidding model.” *Journal of the Operational Research Society*, Palgrave Macmillan, 65(1), 97–107.

Skitmore, M., and Cattell, D. (2013). “On being balanced in an unbalanced world.” *Journal of the Operational Research Society*, 64, 138–146.

Skitmore, M., and Lo, H. P. (2002). “A method for identifying high outliers in construction contract auctions.” *Engineering Construction and Architectural Management*, 9(2), 90–130.

Skitmore, M., and Pemberton, J. (1994). “A Multivariate Approach to Construction Contract Bidding Mark-up Strategies.” *Journal of the Operational Research Society*, 45(11), 1263–1272.

Stark, R. M., and Rothkopf, M. H. (1979). “Discussion of ‘Bidding strategies and probabilities’ by Marvin Gates.” *Operations Research*, 27, 364–390.

Tukey, J. (1977). *Exploratory data analysis*. Addison-Wesley. ISBN: 978-0-201-07616-5.

Wade, R. L., and Harris, B. (1976). “LOMARK: a bidding strategy.” *Journal of the Construction Division ASCE*, 102, 197–211.

Williams, T. P. (2003). “Predicting final cost for competitively bid construction projects using regression models.” *International Journal of Project Management*, 21(8), 593–599.

Yuan, X.-X. (2011). “A correlated bidding model for markup size decisions.” *Construction Management and Economics*, Routledge, 29(11), 1101–1119.

Dataset	Source	Description	Period	Nº bids	Nº auctions	Avg Nº bids/auction	Avg mean bid	Avg st. dev.	Avg skew.	Avg ex. kurt.	Cost estimates provider	Bidders' IDs	Auction type
UK51	(Skitmore, 1991)	London building contracts	1981-82	318	51	6.24	1,770,351	86,753	0.417	0.049	Single bidder	Yes	Uncapped
UK218	(Skitmore, 1986)	North of England public works contracts	1979-82	1,235	218	5.67	143,382	10,238	0.142	0.208	-	Yes	Uncapped
UK373	(Skitmore, 1986)	London building contracts	1976-77	1,915	373	5.13	828,705	45,438	0.253	0.339	-	Yes	Uncapped
US62	(Brown, 1986)	USA Government agency building contracts	1976-84	417	62	6.73	914,019	71,863	0.686	1.142	Project designer	Yes	Uncapped
US50	(Shaffer and Micheau, 1971)	USA building contracts	1965-69	235	50	4.7	921,970	62,652	0.185	-0.054	Single bidder	Yes	Uncapped
HK199	(Drew, 1995)	Primary, secondary schools, police, fire stations & hostels in Hong Kong	1981-90	2,531	199	12.72	1,122,132	118,579	0.85	1.164	-	Yes	Unknown
HK266	(Fu, 2004)	Hong Kong Administrative Services Department contracts	1991-96	3,566	266	13.3	5,500,889	566,769	0.793	1.15	Project designer	Yes	Unknown
AU152	(Runeson, 1987)	General contractors' bids for New South Wales Public Works & Housing	1972-82	1,316	152	8.66	1,605,075	101,345	0.656	0.933	Project designer	No	Uncapped
AU160	(Runeson, 1987)	Specialist contractors' bids for New South Wales Public Works & Housing	1972-82	1,010	160	6.27	230,346	27,515	0.432	0.579	Project designer	No	Uncapped
UK272	(Skitmore, 1981)	BCIS detailed analyses of UK contracts	1969-79	1,670	272	6.14	835,921	46,048	0.134	0.223	-	No	Uncapped
UK537	(Ballesteros-Pérez and Skitmore, 2017)	BCIS detailed analyses of UK building contracts	1979-90	3,392	537	6.32	1,397,781	128,051	0.217	0.169	-	No	Uncapped
SP51	(Ballesteros-Pérez et al., 2012)	Spanish waste water treatment plants and sewer systems	2007-08	761	51	14.93	3,236,328	226,291	-0.084	-0.377	Single bidder	No	Capped
SP116	(Ballesteros-Pérez et al., 2015b)	Spanish High speed railway contracts	2008-14	3,300	110	30.0	37,610,797	2,222,114	0.447	0.835	-	No	Capped

Table 1. Auction datasets summary.

Dataset	All bids all auctions (a)		All bids all auctions no outliers (b)		Median auction values (c)	
	Nº valid bids	$\widehat{CV}$	Nº valid bids	$\widehat{CV}$	Nº valid auctions	$\widehat{CV}$
UK51	318	**0.0045**	301	**0,0031**	50	**0.0031**
UK218	1,235	**0.0117**	1,154	**0,0086**	210	**0.0083**
UK373	1,915	**0.0051**	1,827	**0,0041**	360	**0.0040**
US62	387	**0.0157**	343	**0,0070**	58	**0.0070**
US50	228	**0.0064**	216	**0,0048**	48	**0.0048**
HK199	2,531	**0.0080**	2,405	**0,0060**	198	**0.0058**
HK266	3,566	**0.0067**	3,427	**0,0053**	266	**0.0052**
AU152	1,307	**0.0052**	1,244	**0,0039**	149	**0.0040**
AU160	1,002	**0.0172**	932	**0,0117**	159	**0.0112**
UK272	1,660	**0.0051**	1,602	**0,0045**	270	**0.0043**
UK537	3,392	**0.0145**	3,114	**0,0035**	505	**0.0034**
SP51	643	**0.0048**	638	**0,0041**	46	**0.0046**
SP116	3,300	**0.0046**	3,054	**0,0029**	109	**0.0030**

Table 2. Population coefficients of variation estimates ($\widehat{CV}$) of each dataset with three calculation approaches.

Dataset	Period	$\widehat{CV}$	D_{max}	Nº valid auctions	p-value
UK51	1981-82	0,0031	0,0837	50	0,267
UK218	**1979-82**	0,0086	0,1240	210	**0,998**
UK373	1976-77	0,0041	0,0572	360	0,833
US62	1976-84	0,0070	0,1298	58	0,799
US50	1965-69	0,0048	0,1046	48	0,490
HK199	1981-90	0,0060	0,0877	198	0,925
HK266	**1991-96**	0,0053	0,1097	266	**0,997**
AU152	1972-82	0,0039	0,0970	149	0,911
AU160	1972-82	0,0117	0,0955	159	0,910
UK272	**1969-79**	0,0045	0,0907	270	**0,972**
UK537	**1979-90**	0,0035	0,0760	505	**0,996**
SP51	**2007-08**	0,0041	0,1261	46	**1,000**
SP116	**2008-14**	0,0029	0,2489	109	**1,000**

Table 3. Kolmogorov-Smirnov test results of a single Chi-Squared (χ^2) distribution fitting each dataset (p-values rejecting the null hypothesis for α>95% highlighted in bold)

Dataset	**Period** (approx.)	$\widehat{CV}$	D_{max}	Nº valid auctions	**p-value**
UK218	1979-80	0,0085	0,1053	105	0,848
	1981-82	0,0088	0,0979	105	0,799
HK266	1991-93	0,0050	0,0925	133	0,841
	1994-96	0,0057	0,0915	133	0,798
UK272	1969-74	0,0044	0,0785	134	0,830
	1975-79	0,0042	0,0880	136	0,799
UK537	1979-85	0,0032	0,0753	254	0,913
	1986-90	0,0038	0,0716	251	0,877
SP51	2007	0,0040	0,1395	25	**0,978**
	2008	0,0045	0,1635	16	**0,987**
SP116	2008-09	0,0027	0,2185	55	**0,988**
	2010-14	0,0032	0,2661	54	**0,999**

Table 4. Kolmogorov-Smirnov test results of a single Chi-Squared (χ^2) distribution fit in those datasets rejecting the null hypothesis in Table 3 (p-values still rejecting the null hypothesis for α>95% highlighted in bold)

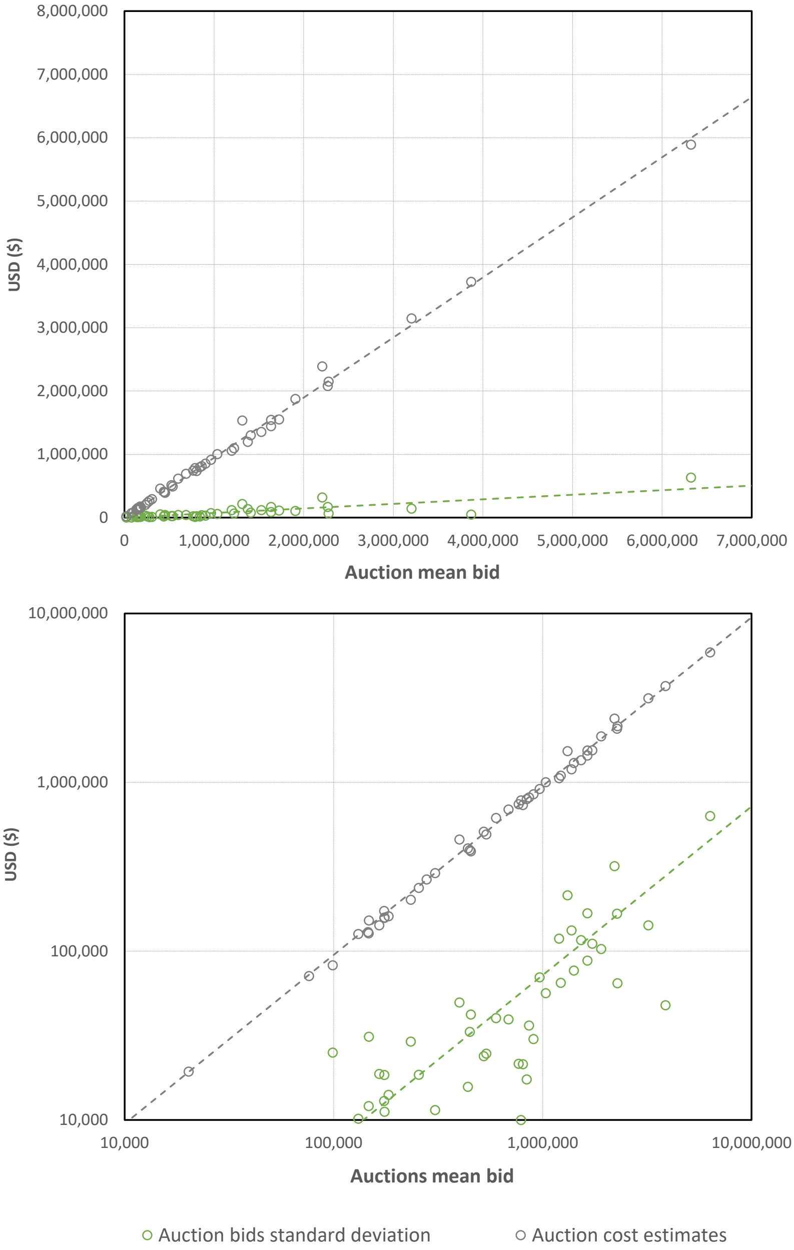

Fig. 1. Example of relationships between the auction's mean bid (X-axis), cost estimate and standard deviation (Y-axis) (auction dataset US50).

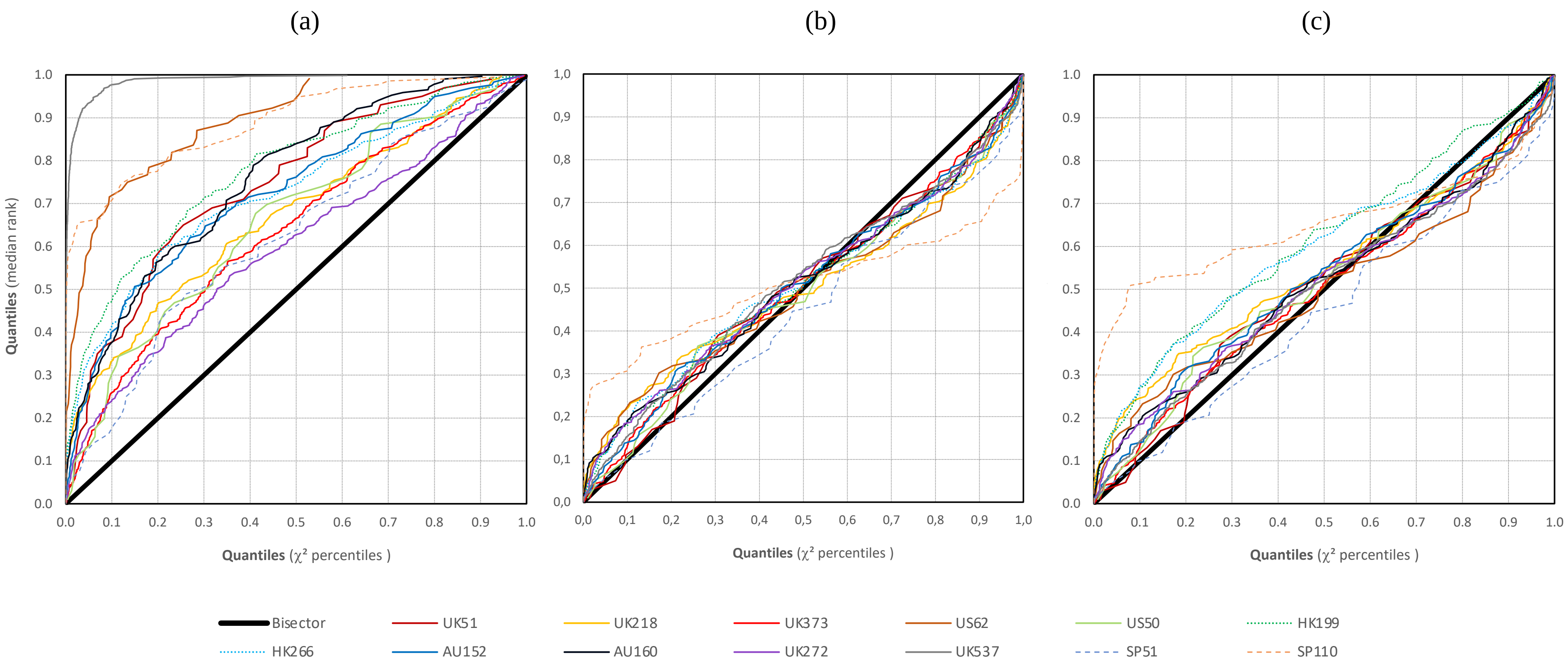

Fig. 2. χ^2 distribution QQ plots of all auctions' CV_j values in the 13 datasets with three calculation approaches: (a) all bids all auctions, (b) all bids all auctions without outliers, and (c) median auction values.

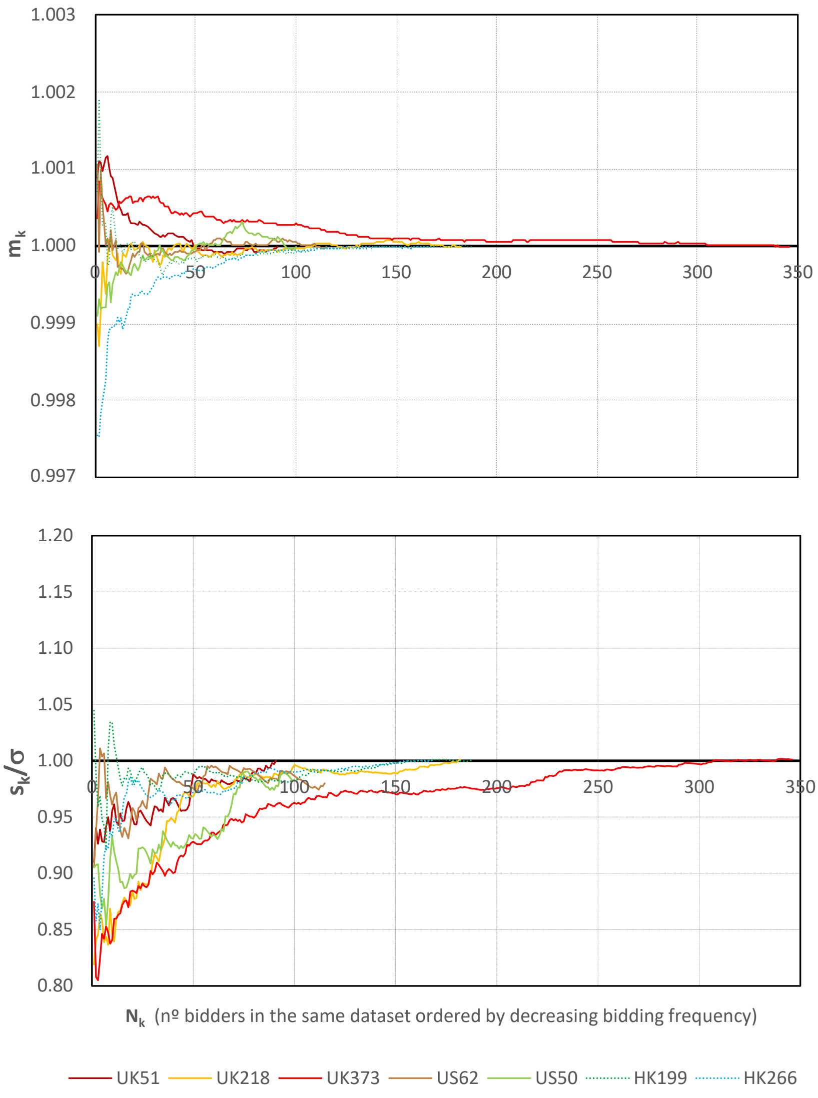

Fig. 3. Variation of bidding competitiveness (expressed in log bids location and dispersion) of the N_k bidders in the 13 auction datasets.