Ballesteros-Perez, Pablo, Skitmore, Martin, Sanz-Ablanedo, Enoc, & Verhoeven, Peter
(2019)
Forecasting the number and distribution of new bidders for an upcoming construction auction.
Journal of Construction Engineering and Management - ASCE, *145*(10), Article number: 040190561 1-10.

https://doi.org/10.1061/(ASCE)CO.1943-7862.0001694

Forecasting the number and distribution of new bidders for an upcoming construction auction

Pablo Ballesteros-Pérez, Ph.D.[1] ; Martin Skitmore, Ph.D.[2] ;
Enoc Sanz-Ablanedo Ph.D.[3*] ; Peter Verhoeven, Ph.D.[4]

Abstract

Estimating the number of new bidders in construction auctions is relevant for both private companies and contracting authorities. For private companies, it allows the total number of competing bidders to be estimated which may lead to better adjustments of future bids. For contracting authorities, it allows the population size of all potential bidders' to be estimated and thus to implement better awarding criteria. Mathematical models for forecasting the number of new bidders and the population size of all potential bidders are, however, very scarce in the construction management literature.

In this paper, we propose an Exponential model for predicting the average number of new bidders based on an urn analogy. The model allows the number of new bidders to be estimated as a function of new versus total participating bidders observed in previous auctions. The parameter estimates obtained from the model also allow the statistical distribution of the number of potential new bidders to be modelled using a sum of Binomial distributions. We validate the Exponential model on three published construction auction datasets, showing that the proposed model significantly outperforms the most advanced model for performing similar tasks – the Multinomial model proposed by Ballesteros-Pérez & Skitmore (2016).

Keywords: bidders, auction, forecasting, exponential model, binomial, competitiveness.

---

[1] Senior Researcher, School of Engineering, Universidad de Cádiz, Spain, pablo.ballesteros @uca.es
[2] Professor, School of Civil Eng. and Built Env., Queensland University of Technology, Australia, rm.skitmore@qut.edu.au
[3*] Associate Prof., Dept. of Mining, Topography and Structure Technology, Universidad de León, Spain, esana@unileon.es
[4] Associate Prof., QUT Business School, Queensland University of Technology, Australia, peter.verhoeven@qut.edu.au

Introduction

A ubiquitous feature of the construction industry worldwide is the use of tendering for the simultaneous selection of contractors and soliciting a competitive price. This generally involves a sealed bid auction, in which the number of bidders is taken to be a proxy of the level of competition in the market. It is not surprising therefore that both the number and the identity of potential competitors are among the most relevant factors when a bidder is deciding whether to submit a bid or not (Ahmad and Minkarah 1988; Shash 1993).

The number of bidders and their dominant competitive profiles greatly condition several auction outcomes (e.g. Dyer, Kagel, & Levin 1989; Hu, 2011; Levin & Ozdenoren 2004; Takano, Ishii, & Muraki 2014). The winner's curse, in which an unaware bidder submits an abnormally low bid that is eventually awarded, is one of the most celebrated (Capen et al. 1971). Indeed, research confirms that while experienced bidders are more successful on average (Fu, Drew, & Lo 2002, 2003), inexperienced bidders are more prone to submit abnormally high or low bids (Ballesteros-Pérez et al. 2015b).

All these facts are clearly relevant to competing bidders in assessing a more realistic level of competition and their competitors' pricing strategies. But they are also relevant to owners in designing their awarding criteria (Liu et al. 2015), as those that restrict market entry make markets less efficient (the opportunity cost is greater, with prices higher than in perfect competition). On the other hand, awarding contracts to bidders who cannot cover their costs may create serious problems down the production line, and may even lead to project failure.

Bidding models have been developed to aid decision making by both owners and bidders. They are usually statistical in nature, making full use of random variables as a simplification mechanism by means of agglomerating the considerable uncertainties involved in construction tendering. A fundamental simplification applied to many theoretical treatments is to avoid the complexities involved in having to deal with different bidders' profiles by assuming that bidders' bids are independently and identically distributed (*iid*) (see Klemperer (2004) for a review of the main contributions). However, empirical studies of construction bids by Oo, Drew, & Lo (2010) and Skitmore (1991) have demonstrated the untenability of *iid* in real construction auctions.

Other, more serious, Bid Tender Forecasting Models (BTFMs) analyze bidders' competitive profiles separately (e.g. Pablo Ballesteros-Pérez, González-Cruz, & Cañavate-Grimal 2013; Carr 1982; Friedman 1956; Gates 1967; Skitmore & Pemberton 1994). This is done from historical auction data, while bidders with no past record are modeled as "average" bidders. However, quantifying the *number* of these new bidders *in advance* of the auction has other problems, such as inferring their potential participation when there is no previous available information about them (Runeson and Skitmore 1999).

The result is that BTFMs normally work well in auction datasets with a high proportion of regular bidders. Unfortunately, this is again not the case in the construction industry, where there is usually a large population of bidders with varied and partially overlapping areas of expertise, and where an irregular few bid or are selected to bid (Skitmore 2013b). BTFMs that take into consideration different bidding profiles, despite being more information-demanding, are generally the most accurate (Ballesteros-Pérez et al. 2016a). These BTFMs can vary in complexity, from considering just two competitive bidding profiles (e.g. regular vs new bidders) to treating every single bidder's identity separately (Ballesteros-Pérez et al. 2016c). Irrespective of how many and which bidding profiles are considered, however, they all need to quantify how many bidders of each bidding category (profile) will submit a bid. Without this information and considering the number of new bidders is a significant proportion of the total number of bidders as discussed earlier, these BTFMs cannot produce reliable estimates.

However, apart from Mercer & Russell's (1969) (unsuccessful) attempt to infer the appearance of new bidders from the periodicity of bid submissions of frequent bidders, and Ballesteros-Pérez & Skitmore's (2016) recent Multinomial model, no other methods have been published for predicting the number of new bidders.

In this paper, we propose an Exponential model for anticipating the number of new participating bidders in an upcoming auction based around the Binomial distribution. We provide not only the statistical distribution of the number of new bidders but also an improved estimate of the size of the population of potential bidders. We estimate the model on three sets of construction bidding data and find the proposed model to be twice as accurate as the Multinomial model, the only previous known model available. The approach suggested in this paper is relevant to practice for both open and selective tendering schemes.

This paper is structured as follows. In the *literature review* section, we briefly summarize the major works on anticipating the number of bidders and the population size of potential bidders. Due to its relevance, the Multinomial model, the only existing model for anticipating the number of new bidders, is described in detail. In the *materials and methods* section, we describe the proposed Exponential model from an urn analogy, its constituent equations, and how to implement the model for forecasting purposes. In the *model validation* section, we apply the Exponential model to three datasets of construction auctions, and compare the model's performance with that of the Multinomial. A dedicated section on the *distribution of the number of new bidders* follows where we show how the Binomial distribution can be used to estimate the number of new bidders in an upcoming auction. In the *population of bidders* section, we illustrate how the population of bidders grows in the three auction datasets. In the *discussion* section, we consider the practical relevance of the proposed model, as well as how the model may be improved. Finally, the *conclusions* section summarizes the major contributions of the paper, including limitations and avenues for future research.

Literature review

The number of *new* bidders participating in an upcoming auction can be inferred from the proportion of past *new* bidders divided by the *total* number of participating bidders, as well as from the population size of potential bidders. However, neither obtaining, nor interpreting these variables turns out to be a trivial task.

Friedman (1956) was among the first researchers to suggest that the Poisson distribution may provide a suitable model to estimate the total number of participating bidders. The Poisson distribution depends on a single parameter λ whose best unbiased estimate corresponds to the average number of participating bidders in previous auctions. Much empirical work, however, has found both supportive and unsupportive evidence of such a fit. A further assertion by Friedman is that the Poisson distribution may model the errors (deviations) of the number of bidders instead – both claims being rejected by Skitmore's (1986) empirical analysis of three sets of real UK construction data.

Since Friedman's work, a plethora of statistical distributions (e.g. Normal, Log-Normal, Uniform, Weibull, Gamma, and Laplace) have been proposed for modelling the total number of participating bidders (Ballesteros-Pérez, Skitmore, et al. 2015; Engelbrecht-Wiggans, Dougherty, & Lohrenz 1986; Skitmore 2013a; Stark & Rothkopf 1979). In this regard, Ballesteros-Pérez et al. (2015) performed an extensive fit analysis spanning 12 datasets of construction tenders from four continents. They conclude that, on average, the lognormal distribution performs best, closely followed by the Normal, Logistic, and Log-Logistic distributions.

Despite the generally poor performance in terms of modelling the number of construction bids, the Poisson model has endured in practice. It is still the most popular model, not just in construction auctions, but also in other settings such as online auctions (Mohlin et al. 2015) and numerical simulations (Takano et al. 2014).

Alternative proposals to treat the number of bidders participating in an auction stochastically have been made by Rubey & Milner (1966), who suggested resorting to experience and observation to anticipate the average value of the participants in upcoming auctions. This approach has been refined by many other researchers who confirmed that, indeed, the number of bidders is different depending on other aspects such as project type and size (Azman 2014; Drew and Skitmore 2006), client (Ballesteros-Pérez, González-Cruz, Pastor-Ferrando, & Fernández-Diego 2012), geographical location (Al-Arjani 2002; Benjamin 1969) and market conditions (Ngai et al. 2002; Skitmore 1981).

However, most of these aspects are quite difficult to standardize (Lan Oo et al. 2007; Oo et al. 2010a; b) and/or are strongly context-specific (their regression relationships remain valid providing the region, client, economic context, and/or the type of projects are similar) (Ballesteros-Pérez et al. 2015a). The only refinement (to having purely random variables) that works in most contexts has been to resort to the contract size (project budget) as a proxy for the total number of participating bidders. In particular, by classifying past auctions into homogeneous categories (same or similar project type, client, and location), this involves exploiting the generally moderate correlation between contract size and the number of bidders to make better predictions of λ (Rickwood 1972; Wade and Harris 1976). However, the proposed model focuses on forecasting the number of new bidders in an upcoming auction, not the total number of participating bidders. The latter has been the subject of other recent analyses by Ballesteros-Pérez et al (2015a). Furthermore, we express the number of new bidders as a proportion of the total participating bidders. This will allow the forecasting of both variables to be treated as separate, independent, problems.

Estimating the population size of all potential bidders has proven even more elusive than the number of participating bidders per auction. So far, only Ballesteros-Pérez & Skitmore (2016) have successfully attempted this in an approach proposed in parallel with their Multinomial model. This basically resorts to dividing the total amount of different bidders identified so far (the size of the bidders' identities database) by the proportion of new versus total bidders in the last auction. This estimate provides a reasonably close approximation, but has the disadvantage of suffering high variability. Therefore, a large number of auctions is generally required to obtain accurate estimates.

Finally, concerning the number of new bidders, researchers to date have only produced one model: the Multinomial model described in the next subsection. The advanced reader, though, may also think of other alternative routes to derive a relatively good estimate of the number of new bidders in an upcoming auction. For example, subtracting the number of frequent bidders (those who have already been identified in the database) from the population size of all potential bidders, and then estimating how likely it is that a proportion of those will submit a bid again in upcoming auctions. However, this approach has several problems. First, it requires classifying the identities of bidders in relatively homogeneous groups (Ballesteros-Pérez, Skitmore, Pellicer, & Gutiérrez-Bahamondes 2016; Shaffer & Micheau 1971; Wade & Harris 1976). This can be very misleading as bidders may bid for different types of work (multi-market scheme) (Morin and Clough 1969) or stop bidding altogether when all their resources are busy (Lan Oo et al. 2012; Skitmore 1988). As both act in opposite directions, analyzing bidders in homogeneous groups can be very unreliable. Second, it still requires an estimate of the total number of participating bidders in an upcoming auction. So far, there are no reliable models to accomplish this task. There are no other alternatives so far.

The multinomial model

Currently, only one model has been proposed to estimate the number of new bidders. This model is an implementation of the multinomial distribution for construction auctions proposed by Ballesteros-Pérez & Skitmore (2016). It contains two variants, a trinomial model for forecasting the total number of different bidders for auction i+1 (the current auction is indexed as auction i), and a Binomial model for forecasting the current number of once bidders for auction i+1. The difference between the Trinomial and the Binomial models is that when bidders become twice bidders, they are promoted to a different category and are no longer counted the Binomial model, only by the Trinomial model. This is the reason why the Trinomial model will be the one compared later.

The Multinomial model works by applying random walks to sets of nonce-, once-, twice-, thrice-bidders, etc. up to the maximum possible number of bids submitted by each bidder up to and including auction i. All subgroups of bidders who have been bidding more frequently in the past are also assumed proportionally more likely to submit another bid in auction i+1. Therefore, the Multinomial model basically counts how many once-, twice-, thrice-bidders and so on, have been currently identified in the tender dataset. Next, it estimates how likely it is that each of these bidders will submit another bid, and sums these estimates to provide an overall estimate of the total (probabilistic) number of (new and frequent) bidders. This capability allows the multinomial model to forecast the total number of bidders participating in an upcoming auction and not just the new bidders (although not that accurately). For that purpose, the model resorts to a regression-based estimate of the population size of all potential bidders. This is also possible by the urn model described later. Indeed, we will show that the urn model estimates compare favorably with the Multinomial estimates.

The Multinomial model is substantially more mathematically complex than the Exponential model we propose. This is partially unavoidable as the former includes other capabilities that are of no interest when forecasting the number of new bidders. Our analysis here focuses on comparing the Multinomial and Exponential models in terms of predicting the number of new bidders in an upcoming auction as well as in estimating the population size of all potential bidders.

Materials and methods

In this section, we present the Exponential model, as well as the representative urn analogy upon which the model is built.

Notation

The proposed model makes use of the following terminology, some of which has already been presented:

i The i^{th} auction

N Total population of all potential bidders.

N_i Number of bidders participating in auction i.

N_1^i* Number of different bidders participating in auctions 1 to i.

N_i* Number of new bidders in auction i (they had not submitted a single bid in auctions 1 to i-1). It can be calculated as $N_i* = N_1^i* - N_1^{i-1}*$. By definition, $0 \le N_i* \le N_i$.

An urn analogy

The number and proportion of new bidders found in an auction i (or i+1) can be assimilated to an urn containing N different balls (each one representing one bidder). Each auction i is represented by a draw of N_i balls (total number of participating bidders in that auction). After each draw, all balls are returned to the urn and they can be drawn again in future (sampling with replacement after each draw).

As successive auctions (draws) take place (i=1,2,3…), the number of balls that have not been drawn before are quantified as new bidders (N_i*). If we simulate the proportion of new balls versus the total number of balls drawn, that is N_i*/N_i, the expected values of these ratios are represented in Figure 1.

Figure 1 was generated using 12,000 (Monte Carlo) simulations assuming a population of N=100 bidders (balls). Each line represents the successive values of N_i*/N_i as auctions (trials) progress (i=1,2,3…) and for different number of bidder (balls drawn) per auction (N_i=1, 2, 3, …, 100 bidders).

[Insert Figure 1 here].

Similar results are obtained with alternative population sizes, irrespective of the total number of trials (i) or the size of each draw (N_i). From the graph it appears that the average values of N_i*/N_i are well represented by an Exponential function. These empirically obtained expressions all have the same generic form if we choose the Euler's number as the logarithmic base:

$$\frac{N_i*}{N_i} = a \cdot e^{b \cdot i} \qquad (1)$$

where a and b are the two coefficients that define the Exponential regression line.

All lines cross the same point (i=1, N_i*/N_i=1), as all bidders (balls) are necessarily new (N_i*=N_i) in the first auction (i=1). By exploiting this boundary condition and taking log values, one of the regression parameters can be expressed as a function of the other as follows:

$$b = -LN\,a \tag{2}$$

where $LN\,a$ is the natural logarithm of a.

Next, by substituting (2) into (1) we obtain:

$$\frac{N_i\,*}{N_i} = a \cdot e^{-i \cdot LNa} \tag{3}$$

Therefore, in terms of bidding, when using expression (3) to forecast the proportion of new bidders in an upcoming auction, it will be necessary to compute the value of a from previous auctions. Fortunately, the value of a can be easily obtained from expression (3) taking log values:

$$a = \left(\frac{N_i\,*}{N_i}\right)^{\frac{1}{1-i}} \tag{4}$$

where N_i*/N_i is the observed value of new versus total bidders participating in auctions up to i.

Inserting expression (4) into (3), the mathematical expression for the proportion of new bidders in auction i+1 is therefore:

$$\frac{N_{i+1}\,*}{N_{i+1}} = a \cdot e^{(1-i)LNa} = \left(\frac{N_i\,*}{N_i}\right)^{\frac{1}{1-i}} e^{\frac{i+1}{i-1}LN\left(\frac{N_i*}{N_i}\right)} \tag{5}$$

Analogously, if the total number of participating bidders in auction i+1 could be somehow anticipated, expression (5) can be reorganized to compute, instead of the proportion, the number of new bidders:

$$N_{i+1}* = N_{i+1} \cdot a \cdot e^{(1-i)LNa} = N_{i+1} \left(\frac{N_i *}{N_i} \right)^{\frac{1}{1-i}} e^{\frac{i+1}{i-1} LN \left(\frac{N_i *}{N_i} \right)} \tag{6}$$

Finally, the previous expressions also allow the population size of potential bidders, that is N, to be estimated. The exponents of the three regression equations in Figure 1 of 0.01 (for N_i =1), 0.02 (for N_i =2), and 0.03 (for N_i=3) were obtained assuming a bidder population of 100 (N=100), indicating that a, N_i and N are empirically related as follows:

$$N = \frac{N_i}{LN\, a} \tag{7}$$

Again, this expression is identical for alternative sets of a, N_i and N values in the urn model. However, despite the fact that expressions (5) to (7) depend on a parameter which is (theoretically) constant (parameter a can be obtained with expression (4) by means of a single auction), a more accurate estimate of a can be obtained by analyzing a larger number of auctions. Similarly, in practice, neither the number of participating bidders (N_i) nor the population size of all potential bidders (N) are constant for all auctions over long periods of time. Therefore, it is practical to resort to "average" values of a and N_i to improve the quality of the estimates of (5) to (7), namely:

$$a_{avg} = \frac{1}{i} \sum_i a_i \tag{8}$$

$$N_{avg} = \frac{1}{i} \sum_i N_i \tag{9}$$

This leads to the following expressions that can be used for forecasting purposes:

$$\frac{N_{i+1} *}{N_{i+1}} = a_{avg} \cdot e^{(1-i)LNa_{avg}} \tag{10}$$

$$N_{i+1} * = N_{avg}\, a_{avg} \cdot e^{(1-i)LNa_{avg}} \tag{11}$$

$$N = \frac{N_{avg}}{LN\, a_{avg}} \quad (12)$$

Among other applications described earlier, expression (12) can be used to monitor the time-varying population size of potential bidders (N) as more auctions are completed.

Model validation

We empirically test the proposed Exponential model on three published datasets of construction auctions. The three datasets, containing all economic bids (not used in this study) and the numerical codes representing the bidders' identities, are provided in the *Supplemental Online Material* file.

Table 1 shows that the major features of the three datasets are vastly different. Dataset 1 contains bidder data in the London area over a short period of time. Dataset 2 contains bidder data in a wider area of the north of England for much smaller projects and over a longer time period. Dataset 3 contains bidder data in the highly competitive Hong Kong construction market, where as many as 33 bidders may be bidding for a single contract.

[Insert Table 1 here]

For all tender datasets, we forecast the proportion (when N_{i+1} is assumed known) and the number (when N_{i+1} is assumed unknown) of new bidders in auction i+1 using the proposed Exponential model. To make the best forecasts possible for auction i+1, all information up to and including auction i is utilized.

Table 2 compares the performance of the proposed Exponential model versus the Multinomial model of Ballesteros-Pérez & Skitmore (2016). The detailed auction-by-auction results of the Exponential model can be found in the *Supplemental Online Material* file. For the performance evaluation, the absolute, instead of squared, errors are preferred since absolute errors allow the error magnitude to be expressed in a more meaningful way, i.e. number of bidders (instead of number of bidders squared).

[Insert Table 2 here]

The results shows that the proposed Exponential model is superior to the Multinomial model, generating less than half the (sum and mean) absolute estimation errors. Given that the Exponential model is mathematically significantly simpler, this improvement is remarkable.

Distribution of the number of new bidders

So far we have implicitly made the assumption that N_i* is a deterministic variable whose average proportion or quantity can be approximated respectively by the Exponential regression line described in equations (10) and (11). However, it is clear that given any number of participating bidders N_i>0, N_i* can also take on different values. Particularly, N_i*=0,1,2, … N_i.

Anticipating the distribution of possible N_i* outcomes is equally important. First, to anticipate how likely it is that each outcome happens. Second, because future bidding models that incorporate randomly generated artificial bids will require a clear set of rules for modelling different number of new bidders.

Further analysis from the simulation results from the urn model reveals that, for any given value of N_i and i, the (unconditional) probability of N_i*=0, 1, 2, … N_i follows a Binomial distribution with a number of trials n=N_i and probability of success p=N_i*/N_i, obtained from expression (10). That is:

$$Distribution\left(N_i\,*\right) = Binomial\left(n = N_i,\, p = \frac{N_i\,*}{N_i}\right) \quad (13)$$

Proof of this can be found in the *Supplemental Online Material* file ("Binomial fit" tab), which provides a number of examples of the simulations used to create Figure 1.

In the urn analogy, however, we assumed that N_i is constant throughout all auctions. This is not usually the case in real auctions, where this number tends to vary substantially and, most of the time, is difficult to anticipate. Hence, it remains necessary to check whether a Binomial distribution can simulate closely enough the number of new bidders in real auctions. Doing this involves converting expression (13) to a sum of Binomials, where each Binomial contributes in the same proportion as auctions with different values of N_i appear in the dataset:

$$Distribution\left(N_{i+1}*\right)=\sum_{j=1}^{N_i}\left\{Freq_j\cdot Binomial\left(n=j,p=\frac{N_i*}{N_i}\right)\right\} \tag{14}$$

$Freq_j$ represents then the proportion of auctions with $j=N_i$ bidders in the tender dataset. By definition, the sum of all $Freq_j$ values from 1 to N_i must equal 1.

The application of expression (14) to dataset 1 is shown in Figure 2. Shown are the distribution fits for different auction group sizes. The figure shows that the average number of new bidders ($N_{i+1}*$) tends to decrease over time (as we analyze more auctions). Although not reported, the results for datasets 2 and 3 are similar to those of dataset 1 (see *Supplemental Online Material* file for results of those datasets).

[Insert Figure 2 here]

It is worth noting that the curves in Figure 2 are the result of a multitude of relatively heterogeneous auctions. Each auction has a different number of participating bidders, N_i, and population of bidders, N, that is constantly growing (see later) which affects the number of new bidders, N_i*, per auction. This causes both the number of trials, n, and the probability of success, p, to change in each of the Binomial distributions (as per expression (14)). Nevertheless, the visual appearance of the goodness of fit of the binomial expressions is excellent. Similarly, all things considered, the *p*-values provided also seem to suggest a satisfactory fit.

Finally, when introducing equation (13), we mentioned that it corresponds to the *unconditional* probability of finding a given number of new bidders in the next auction. This means that that expression is valid when the actual numbers of new bidders observed in previous auctions are considered in average (expected) terms. In general, this is a necessary assumption, as the population size of all potential bidders (N) is not known. However, if N was indeed known it would be possible to resort to a more accurate model than the one offered by the Binomial. That model corresponds to the Hypergeometric distribution, which actually fully represents the urn analogy, presented earlier. Because it is rare for N to be known (or to be accurately estimated) before the auction takes place, that model has been relegated to the Appendix.

The population of bidders

As described earlier in expression (12), the population size of all potential bidders (N) can be approximated as a function of the average values of a and N_i (a_{avg} and N_{avg}). Estimates of N have many applications, probably the most important estimating the market size of different construction sectors.

Calculating the value of N is mathematically very simple with the Exponential model. When describing the urn analogy, we assumed that the population of bidders remained constant. That is how we obtained the straight lines presented in Figure 1 (in log scale). However, this is not usually the case in real auctions, where the identity of participating bidders changes over time, and the total number of participating bidders either increases, remains stable, or decreases as auctions are completed.

Monitoring potential variations in N in the proposed model is straightforward. All that is necessary is to plot the values of the N estimates by applying expression (12) after the completion of each auction i. By observing the trend line, it is possible to infer whether the population of bidders is shrinking, remaining stable, or increasing. Another sign of a significant change in the population size is the coefficient a in expression (3) or a_{avg} in expressions (10) and (12) changing across auctions. These variations can also be plotted, but its application has no physical meaning beyond serving as a proxy of N changes.

Figure 3 plots the estimates of a and N over time for the three datasets. It shows that for all three datasets, the population of potential bidders has been growing over time. A few regression lines fitting the values of N as a function of i are also included for illustrative purposes to better visualize the major trends in successive N estimates.

[Insert Figure 3 here]

Of particular interest is the plot of N in dataset 1. A similar regression line was also shown by Ballesteros-Pérez & Skitmore (2016) for their Multinomial model (Figure 2 in their paper). Whilst their regression model produced a similar curve ($y=N=38.477 \cdot x^{0.396}$) as ours, their estimate of N had a high variation (i.e. $R^2=0.5611$). Fortunately, the proposed Exponential model is capable of producing substantially more accurate estimates of N using a significantly smaller database (number of auctions) and with higher coefficients of determination (i.e. $R^2=0.9639$).

Discussion

Since Friedman (1956) and Gates' (1967) seminal work on Bid Tender Forecasting, many other forecasting models have followed. Most of these try to anticipate the probability of several bidding-related outcomes (the number of participating bidders, lowest and/or average bids submitted, the presence of abnormally high or low bids, etc.), all with the intention of either gaining a competitive edge (from the contractor's perspective) or implementing better awarding criteria (from the contracting authority's perspective).

The construction industry is currently becoming both more competitive and more specialized. Additionally, there is increasingly easier, quicker, cheaper, and more transparent access to all kinds of information. Bidding information is the same. For example, the United States (with the data.gov website), the UK (with data.gov.uk), and the European Union (with the European Public Sector Information Platform), are examples of entities that have recently launched initiatives to make non-personal government data available as open data. Each of these platforms provides access to tens of thousands of governments-related datasets. The procurement and bidding information of local, regional, and national contracting authorities constitute a significant proportion of these datasets. Moreover, these datasets are constantly growing and periodically updated.

In this context, Bid Tender Forecasting Models (BTFMs) are very likely to thrive and become essential tools for enhancing construction projects and services procurement. BTFMs work with historical information to make predictions about the future. Companies and governments that take advantage of this increasingly massive amount of information will be able to make much better decisions. For construction contractors this might mean making more profits by being awarded more contracts and/or anticipating the contracts for which the competition may be less intense. For contracting authorities, this might mean fine-tuning the contract awarding criteria and allow a higher discrimination power over a population of potential bidders whose size and composition can be monitored.

These are just a few examples of inferences that BTFMs implementing the variables analyzed in this study will allow. Mathematical expressions have been provided for each of those variables. Further contexts, implications, and limitations are also discussed here.

In particular, an urn model is proposed to model the number of new bidders in an upcoming auction, N_{i+1}*. Despite assuming a series of relatively stable conditions – a constant number of bidders per auction, N_i, and a constant population of potential bidders, N – the model is empirically accurate at anticipating N_{i+1}* both when N_{i+1} is known and unknown.

Anticipating the total number of participating bidders in upcoming auctions (N_{i+1}) to make better estimates of N_{i+1}*, though, is not always possible. As noted in the literature review section, most models for anticipating N_{i+1} have not gained significant improvements over the pure random case, even when contract sizes of future auctions are known in advance. However, it is relatively common, from past records or just because bidders regularly meet each other, to know the number of future competitors with varying degrees of certainty. There are also situations when the number of bidders is certain. This happens, for instance, when the owner shortlists a specified number of bidders from some prequalification stage.

Finally, there are situations when either the maximum or minimum number of bidders is known. This usually happens, respectively, when the total number of invitations extended by the owner is known (although not all bidders may submit a bid), or when the owner states that unless there is a minimum level of competition (a minimum number of bids received), the contract will not be awarded. Therefore, both N_{i+1} known and unknown cases are worth considering and predictions may require complementary information from both cases sometimes.

Finally, the advanced reader may think that, once a first estimate of the population of potential bidders N is available, this variable can be used in turn to improve future N_{i+1}* estimates. This is, for example, a necessary assumption to implement the Hypergeometric model presented in the Appendix. Indeed, there is apparently no reason why forecasting future values of N should not be possible, as the growth of N in Figure 3 tends to be mostly relatively smooth. However, addressing this problem this way involves additional complexities and limitations.

Concerning the additional complexities, once a series of (past) N estimates is available, it is necessary to find out how to forecast the value of N accurately in auction i+1. This may need to be achieved in a number of ways as the N estimates generally experience some degree of (local) volatility. Possible approaches may involve weighting more heavily the most recent estimates of N, or just taking a linear regression estimate of the value of N at auction i+1. Obviously, constant updates of the latest N estimates may be necessary after every new auction has been completed.

There are two further limitations. First, a relatively large number of past auctions may be required to obtain a relatively stable estimate of N (between 20 and 30 in the three auction datasets analyzed). This could make it impractical, as it would be significantly more information-greedy than the model proposed in this paper. Second, even if we tried to forecast N_{i+1}* using N, it would still be necessary to resort to expressions (10) to (12), as these provide the only means of updating the N estimates. Therefore, any future model that tries to take advantage of N estimates will need to make use of the model presented here first.

Conclusions

We propose an Exponential model based on an urn analogy to predict the number of new bidders that will participate in an upcoming auction. This, or alternatively predicting the proportion of the new versus total number of bidders participating in the next auction, is of significant value to a number of construction stakeholders, mostly to enhance competitiveness.

Tests on three construction auction datasets shows that the absolute deviation errors of the proposed Exponential model are around 50% smaller than those produced by Ballesteros-

Pérez & Skitmore's (2016) Multinomial model – the only model with a similar aim found in the literature. Moreover, the proposed model is mathematically simpler and has a much lower computational cost. Indeed, the calculations could be carried out manually if necessary. Furthermore, we show that the statistical distribution of new bidders closely resembles the sum of a series of Binomial distributions. We also analyzed the variation (growth) of the population of potential bidders by means of the same Exponential model. Finally, we briefly outlined a number of applications of the model.

Regarding the model limitations, the exponential model heavily relies on a relatively accurate estimate of the *total* number of participating bidders in the upcoming auction. That is not a problem when we are only interested in the *proportion* of new versus total bidders, but it is certainly limiting when we are interested in the *absolute* number of new bidders. Many models in the past have attempted to come up with reliable estimates of the total number of participating bidders and some examples have been reviewed in this paper. Besides the early Poisson distribution model, many of them have resorted to multivariate regression. However, most of these models have shown to be strongly context-specific (same country, economic environment, client, type of project, etc.) and hardly provide reliable results in more generic settings. Hence, future research may investigate multivariate approaches, particularly on trying to accommodate the information of subsequent bids from previous bidders.

Appendix

The (conditional) probability of k new bidders participating in auction i+1 (i.e., $N_{i+1}*$) in the urn analogy actually follows a Hypergeometric distribution. The Hypergeometric distribution is a discrete distribution that describes the probability of getting k successes (drawing k objects with a particular feature) in n draws, *without replacement*, from a finite population of size *N*. That population *N* contains exactly *K* objects with that feature (in this case that the bidder is new). Therefore, each draw is either a success or a failure.

Using the previous notation, the Probability Mass Function (PMF) of the Hypergeometric can be expressed as follows:

$$Distribution\left(k = N_{i+1}*\right) = Hypergeometric\begin{pmatrix} k = n^{o}\, of\ sucesses = N_{i+1}* \\ n = n^{o}\, of\ trials = N_{i+1} \\ N = population\, size = N \\ K = n^{o}\, of\ objects = N - N_1^i* \end{pmatrix} =$$

$$= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\binom{N-N_1^i*}{k}\binom{N_1^i*}{N_{i+1}-k}}{\binom{N}{N_{i+1}}} = \tag{15}$$

$$= \frac{\left(N - N_1^i*\right)!}{k!\left(N - N_1^i* - k\right)!} \cdot \frac{\left(N_1^i*\right)!}{\left(N_{i+1}-k\right)!\left(N_1^i* - N_{i+1} + k\right)!} \cdot \frac{N_{i+1}!\left(N - N_{i+1}\right)!}{N!}$$

The problem when implementing expression (15) is, obviously, that *N* is generally not known, what is more, it may be changing over time. If we try to infer *N* from the random observations (number of new bidders from past auctions) we face another problem. The mean of the Hypergeometric distribution corresponds to $n\cdot(K/N)$. Note that this mean coincides with the mean of the Binomial distribution, which is $n\cdot p$. Thus, as from expression (13) we can infer that

$$np = n\frac{N_{i+1}*}{N_{i+1}} = n\frac{N - N_1^i*}{N} = n\frac{K}{N} \tag{16}$$

Therefore, when implementing the Binomial model suggested earlier when trying to infer the probabilities of finding a given number of new bidders in the next auction we are already using the best estimates we have available. These estimates will be very accurate as long as $N_i << N$, which is generally the case in real contexts.

However, the Hypergeometric model also offers a new way of estimating *N* after each auction has occurred. Working with expression (16) we can infer that once auction *i* has been completed, the best estimate of *N* corresponds to the one that fulfils the following equality:

$$N_i \frac{N - N_1^{i-1}*}{N} = N_i* \qquad (17)$$

Therefore, by obtaining the variable *N* using expression (17) we can infer that the best estimate of *N* after auction *i* corresponds to:

$$N = \frac{N_i \cdot N_1^{i-1}*}{N_i - N_i*} \qquad (18)$$

The problem with this expression is that it suffers from high volatility. It is almost identical to that proposed by Ballesteros-Pérez & Skitmore (2016), which has proven to be more disadvantageous than the ones proposed in expressions (7) and (12). Therefore, as a general rule, the latter are preferred over expression (18).

Acknowledgements

This work was partially supported by "Estancias de movilidad en el extranjero José Castillejo para jóvenes doctores, 2017" from the Ministerio de Educación, Cultura y Deporte, Gobierno de España (Spain).

Data Availability

The data generated or analyzed during the study are available from the corresponding author by request.

References

Ahmad, I., and Minkarah, I. (1988). "Questionnaire Survey on Bidding in Construction." *Journal of Management in Engineering*, American Society of Civil Engineers, 4(3), 229–243.

Al-Arjani, A. H. (2002). "Type and size of project influences on number of bidders for maintenance and operation projects in Saudi Arabia." *International Journal of Project Management*, 20, 279–287.

Azman, M. A. (2014). "Number of Bidders in Small and Medium Public Construction Procurement in Malaysia." *Applied Mechanics and Materials*, 567, 595–600.

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

Ballesteros-Pérez, P., González-Cruz, M. C., Fuentes-Bargues, J. L., and Skitmore, M. (2015a). "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. (2012). "The iso-Score Curve Graph. A new tool for competitive bidding." *Automation in Construction*, 22, 481–490.

Ballesteros-Pérez, P., and Skitmore, M. (2016). "Estimating the number of new and repeated bidders in construction auctions." *Construction Management and Economics2*, In press.

Ballesteros-Pérez, P., Skitmore, M., Das, R., and del Campo-Hitschfeld, M. L. (2015b). "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 Gutiérrez-Bahamondes, J. H. (2016a). "Improving the estimation of probability of bidder participation in procurement auctions." *International Journal of Project Management*, 34(2).

Ballesteros-Pérez, P., Skitmore, M., Pellicer, E., and Gutiérrez-Bahamondes, J. H. (2016b). "Improving the estimation of probability of bidder participation in procurement auctions." *International Journal of Project Management*, Elsevier Ltd and Association for Project Management and the International Project Management Association, 34(2), 158–172.

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

Benjamin, N. B. H. (1969). *Competitive bidding for building construction contracts. PhD dissertation. Stanford University*.

Capen, E. C., Clapp, R. V., and Campbell, W. M. (1971). "Competitive Bidding in High-Risk Situations." *Journal of Petroleum Technology*, 23, 641–653.

Carr, R. I. (1982). "General bidding model." *Journal of the Construction Division. Proceedings of the American Society of Civil Engineers*, 108(CO4), 639–650.

Drew, D. S., and Skitmore, R. M. (2006). "Competitiveness in bidding: a consultant's perspective." *Construction Management and Economics*, E. & F.N. Spon, 10(3), 227–247.

Dyer, D., Kagel, J. H., and Levin, D. (1989). “Resolving uncertainty about the number of bidders in independent private-value auctions : an experimental analysis.” *RAND Journal of Economics*, 20(2), 268–279.

Engelbrecht-Wiggans, R., Dougherty, E. L., and Lohrenz, J. (1986). “A Model for the Distribution of the Number of Bids on Federal Offshore Oil Leases.” *Management Science*, 32(9), 1087–1094.

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*.

Fu, W. K., Drew, D. S., and Lo, H. P. (2002). “The effect of experience on contractors’ competitiveness in recurrent bidding.” *Construction Management and Economics*, Taylor & Francis Group, 20(8), 655–666.

Fu, W. K., Drew, D. S., and Lo, H. P. (2003). “Competitiveness of Inexperienced and Experienced Contractors in Bidding.” *Journal of Construction Engineering and Management*, American Society of Civil Engineers.

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

Hu, A. (2011). “How bidder’s number affects optimal reserve price in first-price auctions under risk aversion.” *Economics Letters*, Elsevier B.V., 113(1), 29–31.

Klemperer, P. (2004). *Auctions: Theory and Practice. Princeton University Press*.

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.

Lan Oo, B., Lo, H.-P., and Teck-Heng Lim, B. (2012). “The effect of bidding success in construction bidding.” *Engineering, Construction and Architectural Management*, 19(1), 25–39.

Levin, D., and Ozdenoren, E. (2004). “Auctions with uncertain numbers of bidders.” *Journal of Economic Theory*, 118, 229–251.

Liu, X., Wang, D., and Jiang, Z. (2015). “Simulation and Analysis of Bid Evaluation Behaviors for Multiattribute Reverse Auction.” *IEEE Systems Journal*, Institute of Electrical and Electronics Engineers Inc., 9(1), 165–176.

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

Mohlin, E., Östling, R., and Wang, J. T. (2015). “Lowest unique bid auctions with population uncertainty.” *Economics Letters*, Elsevier, 134, 53–57.

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

Ngai, S. C., Drew, D. S., Lo, H. P., and Skitmore, M. (2002). “A theoretical framework for determining the minimum number of bidders in construction bidding competitions.” *Construction Management and Economics*, Taylor & Francis Group, 20(6), 473–482.

Oo, B.-L., Drew, D. S., and Lo, H.-P. (2010a). “Modeling the Heterogeneity in Contractors’ Mark-Up Behavior.” *Journal of Construction Engineering and Management*, 136(7), 720–729.

Oo, B.-L., Drew, D. S., and Runeson, G. (2010b). "Competitor analysis in construction bidding." *Construction Management and Economics*, Routledge, 28(12), 1321–1329.

Rickwood, A. K. (1972). *An investigation into the tenability of bidding theory and techniques, and proposal for a bidding game. Msc Thesis. Loughborough University of Technology*.

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

Runeson, G., and Skitmore, M. (1999). "Tendering theory revisited." *Construction Management and Economics*, Taylor & Francis Group, 17(3), 285–296.

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

Shash, A. A. (1993). "Factors considered in tendering decisions by top UK contractors." *Construction Management and Economics*, E. & F.N. Spon Ltd., 11(2), 111–118.

Skitmore, M. (1981). *Bidding dispersion – an investigation into method of measuring the accuracy of building costs estimates. Msc Thesis. University of Salford*.

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

Skitmore, M. (1988). "Fundamental Research in Bidding and Estimating." *Proceedings British/Israeli Seminar on Building Economics; The International Council for Building Research Studies and Documentation CIB W-55,*.

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

Skitmore, M. (2013a). "Generalised gamma bidding model." *Journal of the Operational Research Society*, Palgrave Macmillan, 65(1), 97–107.

Skitmore, M. (2013b). "kmth price sealed-bid auctions with general independent values and equilibrium linear mark-ups." *Journal of the Operational Research Society*, Palgrave Macmillan.

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.

Takano, Y., Ishii, N., and Muraki, M. (2014). "A sequential competitive bidding strategy considering inaccurate cost estimates." *Omega*, Elsevier, 42(1), 132–140.

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

Dataset	***Source***	***Description***	***Period***	***Nº bids***	***Nº contracts***	N_{avg}
1	Skitmore (1986)	London building contracts	1976-77	1,915	373	5.13
2	Skitmore (1986)	North of England public works contracts	1979-82	1,235	218	5.67
3	Fu (2004)	Hong Kong Administrative Services Dept. contracts	1991-96	3,445	266	13.30

Table 1: Descriptive summary of the datasets of construction tenders

Dataset	*Absolute deviations*	N_{i+1} *known*		N_{i+1} *unknown*	
		Exponential	***Multinomial***	***Exponential***	***Multinomial***
	Sum	*50.53*	*103.22*	*280.96*	*497.50*
1	*Average*	*0.14*	*0.28*	*0.76*	*1.33*
	Maximum	*1.00*	*1.28*	*4.36*	*5.85*
	Sum	*25.64*	*32.39*	*144.25*	*274.32*
2	*Average*	*0.12*	*0.15*	*0.67*	*1.26*
	Maximum	*0.74*	*0.90*	*5.09*	*5.11*
	Sum	*8.63*	*10.15*	*124.94*	*144.54*
3	*Average*	*0.03*	*0.05*	*0.47*	*0.65*
	Maximum	*0.67*	*0.77*	*8.87*	*10.17*

Table 2: Performance of the Exponential and Multinomial models.

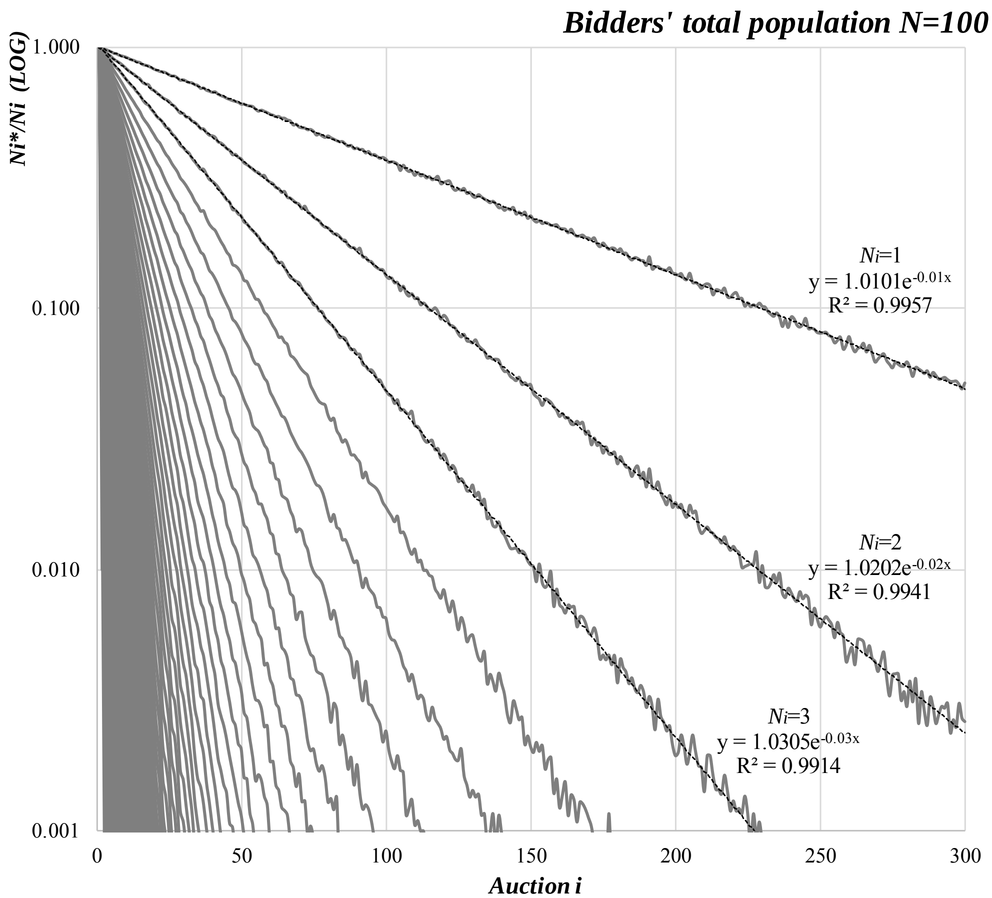

Fig. 1. Urn analogy with N=100 bidders and N_i curves 1 to 100.

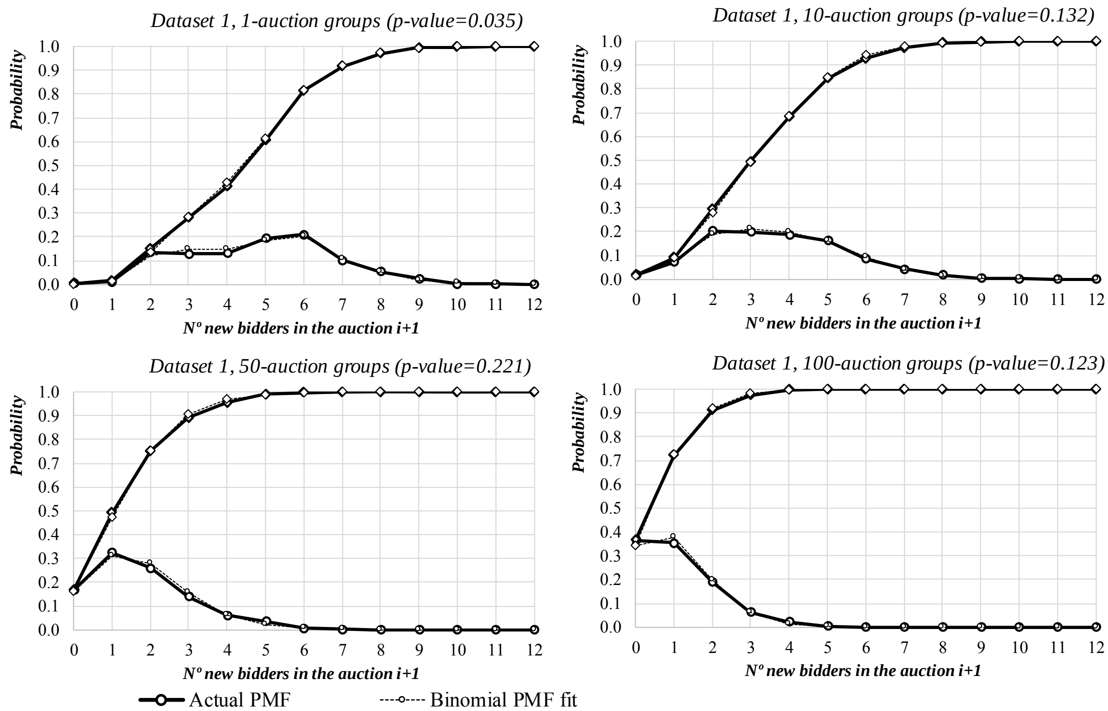

Fig. 2. Binomial distribution fit to the N_{i+1}* values when auctions are analyzed in groups of 1, 10, 50 and 100.

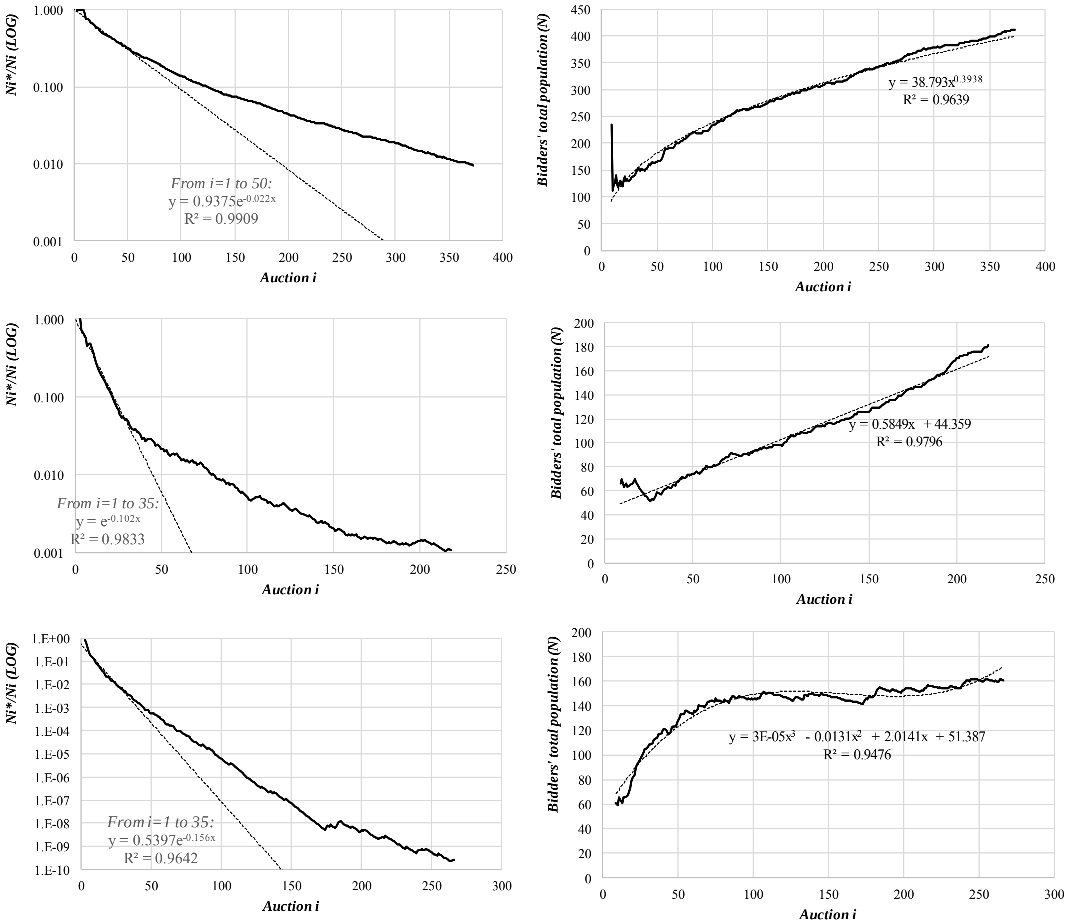

Fig. 3. Variation of a coefficients and population of bidders, N, for the three datasets of construction tenders.