MASTER

Agent Based Simulation of Online Auctions

de Ronde, T.

Award date:
2022

Technische Universiteit Eindhoven
University of Technology

Department of Mathematics and Computer Science
Architecture of Information Systems Research Group

Agent Based Simulation of Online Auctions

Master Thesis

Tom de Ronde

Supervisors:
Dr. Yingqian Zhang - TU/e IE&IS
Dr.Ir. Laurens Bliek - TU/e IE&IS
Michiel Hilgeman - Troostwijk

Eindhoven, May 2022

Executive summary

This master thesis project was made possible by Troostwijk auctions. Troostwijk is one of the largest online auction marketplaces in Europe. They facilitate hundreds of thousands of online auctions annually across 150 countries. The focus of these auctions is mostly on industrial sectors such as agriculture, construction, food handling and metal.

The auctions on Trooswijk have design feature setting them apart from the most common online auctions on other marketplaces. The auction format mostly used is the English ascending bid auction. The most interesting feature is the soft-close format of an auction. In this format an online auction is extended by five minutes when a bid has been placed in the final five minutes of the auction.

Another interesting feature of the auctions on Trooswijk is the long duration. Online auctions on Troostwijk take place of the course of two weeks on average. While a one day online auction is perceived as a short format. And online auctions of 3, 7 or 10 days are perceived as long [22].

Perhaps the most important design feature of online auctions is the starting price 2.2.1. Troostwijk experimented with the starting price over the course of two separate data gathering periods. Two different strategies were tested during these periods. The low pricing strategy started an auction at roughly 10% of the estimated value. The other strategy started auctions at roughly 50% of the estimation.

These auction design choices play a key role in shaping the online auction process. This process is judged based on key performance metrics for online auctions. The first of two metrics used in this project are the result of an auction measured as the relation between the final bid price and the estimated value. The second metric is the number of bidders who placed a bid in an online auction.

The other side shaping the online auction process is the bidder perspective. Bidders behavior in auctions is an important area of interest in the online auction literature 2.3. Many studies defined the most popular strategies amongst bidders. Evaluators place bids during the early to middle stages of the auction duration, often only placing a single bid. Snipers or opportunists place their bids at the final stage of an auction, also often placing just a single bid. The Sniper strategy is the most commonly used strategy by bidders. Then a more active style of bidding are the participators. These bidders tend to be active throughout the whole duration of an auction and place multiple bids.

All research on bidder strategies had a similar data partitioning method to find clusters of bidders in their data. The common method is the K-mean algorithm 2.6.1. The k-means++ algorithm is an adapted version that changed the method of setting initial cluster centroids for the algorithm. The k-means++ algorithm proved to consistently find better clustering results with faster computation times. Therefore, the K-means++ algorithm was to partition the data in this project.

The project was executed using the Cross Industry Process for Data Mining model [50]. The model provides a framework to help with a structured execution of a data mining project. The CRISP-DM cycle consist of 6 phases.

The first phase is the business understanding phase. In this phase the online auction process is evaluated. Scientific literature on online auctions as well as bidding behavior was reviewed.

The second phase is the data understanding phase. The data was gathered by Troostwijk over the course of two periods where different strategies regarding the starting price of auctions were tested. The structure of the data, the available information, and the description of the data conclude this phase.

The third phase is the data preparation phase. For htis project all steps required to create the input for the simulation model were executed in this phase. Firstly, the data was cleaned. Then, a subset of the clean data was selected based on several criteria. One of those criteria was the auction had the have only one item. Another criteria was the auction had to result in a successful sale. With the appropriate subset of clean data many data transformation steps were done. The main transformation was the preparation of the data for the K-means++ algorithm. Other steps were to prepare data for bidder arrival fitting and personal valuation fitting.

The fourth phase is the modelling phase. In this phase the simulation model is built. The simulation model is built guided by the steps in the framework for agent based simulation models 2.4.1. Key elements in this framework are the agents, the relation and the environment.

The fifth phase is the evaluation phase. In this project the evaluation of the model was the validation method. The validation proved the simulation model accurately reflects the processes of online auctions with the low and medium pricing strategies.

This project showed how to create an agent based simulation model for online auctions. The simulation model accurately reflects the process of real online auctions using multiple starting price strategies. Furthermore, the bidder behavior analysis showed a new bidding strategy. The starter strategy was never defined in existing literature on bidding behavior. In the client base on Troostwijk the starter strategy proved to be a fairly popular strategy.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Troostwijk is one of the largest online auction marketplaces in Europe. They facilitate hundreds of thousands of auctions across over 150 countries annually. The vast majority of these auctions are in industrial items, as opposed to consumer type items more often seen on other online auction marketplaces. Some examples of such industries are the agriculture, construction, food handling, metal and more. Another aspect of Troostwijk that differs from other online auctions is the type of auctions they run, the most often seen auction type is the hard-close ascending auction where Troostwijk runs soft-close ascending auctions. In the soft-close auctions ran at Troostwijk the auction duration is extended by five minutes when a new highest bid has been placed in the final five minutes of an auction. A third distinctive feature of Troostwijk's auctions is the very long auction duration, on average auctions run for 15 to 16 days. Comparing this to the duration of auctions studied in the literature where auction of 3,7 or 10 days are viewed as long [22].

As online auctions increased in popularity so did the research in performance and optimisation of online auctions. As reported in the literature study 2.2.1 one main decision variable with a large impact on the online auction process and performance is the starting price. Initial studies show that lower starting prices increase the number of active bidders, and consequently increase the final bid price. This theory sparked the interest of Troostwijk to start experimenting with the starting price in relation to the estimated price of an item. The way Troostwijk set this up is to start the auction on 50 to 60 percent of the estimated price during one period, and at 10 percent during the other. The question that arises is: What does the online auction process look like using different starting price strategies?

In the literature study is shown that studies attended to answer this question by creating highly fixed data using a single item that is sold in online auctions with the auction design as identical as possible other than the variable at interest. This way the effect of that one variable on the outcome of online auctions can be investigated

A new way to approach online auction studies is a simulation model. Initial work on this shows that simulations can be at least as accurate as quantitative analysis methods, as discussed in the literature review 2.4.3. A major advantage of simulation models is that it allows for more flexibility in the design parameters of the online auction process. This gives that opportunity to gain more insight in the way each parameter impacts an online auction.

This thesis project builds upon the limited simulation work previously done in the field of online auctions. The way this project adds to the literature is by adding more of the commonly used bidding behavior strategies. These bidder strategies are required as input for the simulation model. In order to correctly program the behavioral characteristic for every type of bidder a full bidding behavior analysis is done similar to previous studies in the field of bidding behavior. This project adds to that field by discovering a new type of bidder strategy. Furthermore, the simulation runs multiple starting price strategies in order to research the effects of the starting price on the online auction process.

1.1 Problem Statement

Online auctions work very similarly to traditional auctions. An item is presented and sold to the highest bidder. In the auction format used by Troostwijk, an English soft-close type, bidders have the opportunity to bid on the item until nobody is willing to and the auction duration has passed. The soft-close rule means that when a bid is placed in the final five minute of an auction the auction is extended by an additional five minutes.

Troostwijk and the sellers share the goal of maximizing revenues. The key decisions they can make are the auction duration and the starting price of and auction. Much research has been done on the effects of these decision variables, with varying results as discussed in the literature study 2.2. The impact of the auction design has shown to depend on the marketplace and it's user base. Therefore, uncertainty remains on how to utilize the knowledge of these variable to design an optimal auction.

The other side of the online auction process is shaped by the participating bidders. The literature study shows that bidders will adopt a certain strategy over time. These strategies are used to optimize different goals, the main goal being winning likelihood and bidder's surplus. However, very little is known on how these strategies shape the auction process as a whole and how it impacts the auction outcomes.

Thus, uncertainty remains on the effects of auction design on the auction process and outcomes. A simulation model offers an accurate and flexible method to gain knowledge on the online auction process.

1.2 Research Questions

The goal of this project was to create a simulation model that accurately reflect the real online auction process. In order to give structure to the project the following research questions are proposed.

Main research question: How to build a realistic simulation model for online auctions?
In this project the online auction process is modeled using agent-based simulation modeling. These models consist of three main elements, the agents represent the bidders who act in an online auction. The relationship represents the human interactions. Finally, the environment is a set of boundaries in which the agents operate.

Sub-question 1: How can bidder behavior be modelled?
In the literature study is shown that bidder strategies are defined in terms of timing of a bid as well as the number of bids. The different strategies are likely to also differ in terms of the personal valuations of items.

Sub-question 2: How can a simulation model be used to evaluate the online auction process?
The objective of the simulation model is to acquire knowledge on the online auction process. The effects of decision variables as well as sensitivity to design parameters can be analyzed. Which can lead to further interesting research questions for new studies.

Sub-question 3: How can the simulation model be validated? In the literature study is shown that a major step in simulation research is the validation of the model. In this project the model is validated using confidence interval validation.

1.3 Project Approach and Report Outline

The framework used in this research project is the Cross Industry Process for Data Mining Model (CRISP-DM)[50]. The CRISP-DM process model serves to execute a data mining project in a structured manner. The cycle consists of 6 phases as shown in 1.1. The project starts with

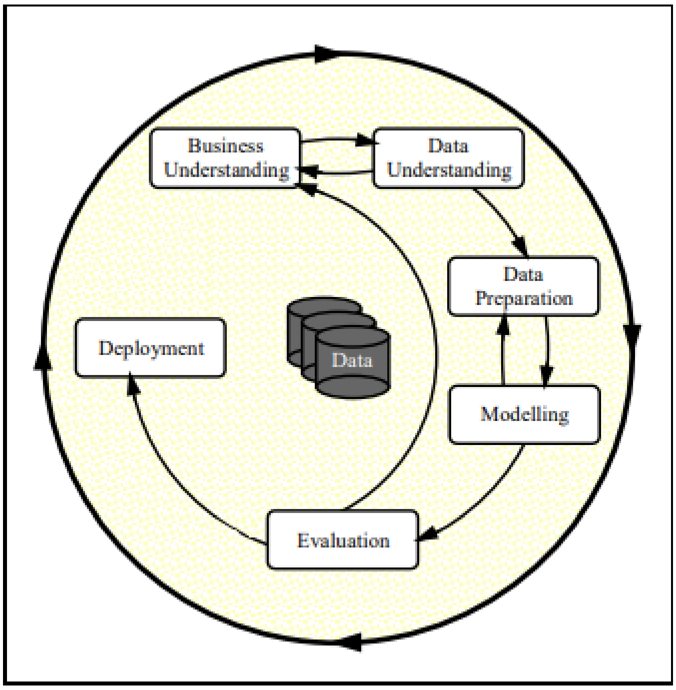


Figure 1.1: CRISP-DM Process Model [50]

the business understanding phase aimed towards the understanding of the project objectives and converting this into problem definition. This phase is discussed in the problem statement 1.1 as well as the literature study 2. The second phase is data understanding, in this phase the data collected data is observed and possible data quality issues are identified. Additionally the data structure is discussed, the data understanding phase is reported in 3. The third phase is the data preparation phase, this phase consists of all necessary steps to convert the raw initial data into a final data set ready to serve as input for the simulation model. Steps in this phase includes data cleaning, data selection, data transformation and more and are discussed in section 4 . With the prepared data the model can be created. The modeling and data preparation steps are often executed iteratively since often problems are discovered while modeling. In the modeling phase an additional framework for agent based simulation (ABS) will be used. The modeling phase is discussed in chapter 5. Finally, the simulation model will be validated in the chapter 6. This concludes the CRISP-DM cycle. In chapter 7 the project is discussed.

Chapter 2

Literature review

In This chapter a literature review will be reported. The domain will be online auctions, with additional focus on two opposing major perspectives: the seller's and the bidders perspectives. Then agent based modeling will be discussed as well as methods of validating such models. Finally, machine learning in online auctions is discussed with a focus on k-means clustering and k-medoids clustering.

2.1 Online auctions

Online auctions as a part of e-commerce has grown rapidly over the past two decades. As a result the scientific literature had a similar development. The online auction literature directly follows the literature of traditional auctions. In the traditional auction literature three main types of auctions are defined [39]: English auction, Dutch auction and sealed-bid auction. English and Dutch are ascending and descending bid auctions respectively. The sealed-bid auction where the highest bid wins and pays either the value of his bid or the second highest bid. In online auctions English auction is by far the most dominant type [36].

At the core of traditional auction literature are several valuation models aiming to understand bidder behaviour. Starting with the Independent Private Values Model discussed by Vickrey [45]. This model assumed that bidders know their valuation of an item, but are uninformed about competition valuations. The Common Values Model assumes the opposite, where a bidder does not know the value of an item [43], [49] but is informed about competition valuations. A Third approach is the Affiliated Values model [40], which has the more realistic assumption that bidders are influenced by both their own as well as the competing bidders valuations and item qualities. The AV model is confirmed many times by different studies on the impact of the number of bidders on auction outcomes, as shown in later sections.

In 2008 a framework for online auction research was proposed by [5]. Based on existing literature at the time they proposed 5 main dependent variable to determine the outcomes of online auctions.

Auction success	Auction results in a completed sale
Final closing price	highest bid in successful auction
Seller revenue	The gross amount of the sale
Price Premiums	Final bid compared to the average price of identical items
Number of bids	Total number of bids in an auction

Table 2.1: Auction outcomes

These variables depend on a lot of different factors which are categorized as Seller-, Auctioneer- and Bidder-controlled factors. The Auctioneer controlled factors consist of enabling certain auction types for sellers and design factors maximizing technology acceptance by both the buyers and the sellers. These factors are part of the widely researched field of technology acceptance and is beyond the scope of this research. The Seller and Bidder controlled factors will be discussed in more depth in the following sections.

2.2 Seller perspective

When listing items for an auction sellers have a lot of decisions to make. These effects of these decisions on the sellers goal, a successful sale for the highest possible final price, have been widely studied. Twelve of such decision variables are identified in a literature review [5] in 2008. Variables such as payment and delivery options relate to more practical implications for buyers ease of use. While sellers reputation rating and product information relate to buyers perceived risk. Similarly to most auctioneer controlled variables the previously mentioned seller controlled variables are very important regarding the buyers technology acceptance. Therefore these variables also are beyond the scope of this research. Pricing variables such as Initial bid price, and the availability of a reserve price or buy-now option have a more direct effect on auction performance and are widely studied. Other variables such as auction duration and auction end time are also directly linked to the auction performance. Finally, the auction mechanism is a key decision variable which can have an effect on revenue as well as bidder behavior.

2.2.1 Pricing

Choosing a **starting price** is probably the most important decision a seller makes. A low starting price encourages more bidders to participate in the auction with a higher risk of a lower final price. On the other hand a high starting bid reduces the risk of a low final price, but will increase the risk of not selling. The effects of starting price on auction outcome is the most extensively researched topic in the field of online auctions. One study [52] found for eBay like online auctions the optimal pricing strategy is to set a starting price as low as possible and to not have a reserve price. Another research [30] did six studies with different methodologies to find that low starting prices result in higher final bid prices. They found three reasons for this relation. Firstly, the low starting price reduces the barriers of entry, which increases the competition and results in higher final bids. The second reason is the low starting price leads to more time commitment by bidders creating sunk costs, and as a result escalates their bids. Finally, the increased traffic created by the low starting price can increase bidders valuation of an item. However, the effects of low starting price on final bids can be reversed by high barriers of entry that limit traffic. A high entry barrier can be misspelled brands, titles, or descriptions. Other entry barriers can be seller's reputation [53], [10]. And even low traffic can be interpreted as an entry barrier, since traffic leads to more traffic [30].

A second pricing decision sellers can make is whether or not to use a **(hidden) reserve price**. When the final bid of an auction is lower than the reserve price the seller will not sell. A seller can use a reserve price to protect itself from selling an item too cheap. However, this protection does come at a cost. Since, a study [27] of 50 matched pairs of Pokémon cards found that the use of a secret reserve price reduces the probability of a successful auction as well as the number of serious bidders and lowers the final bid. Another study [46] found the optimal reserve price to be 1 dollar and disclosed, with the goal to maximize auction interest and final sale price. A reserve price that low offers no protection to selling too cheap. Thus, one could argue that using no reserve price at all is the optimal setting. This view is confirmed by Bapna and Gupta [20] who found that "a zero reserve price provides higher expected profits than a reserve price greater than or equal to the auctioneers salvage value". Therefore, there is a clear trade-off between protection against selling too cheap and the expected revenue.

2.2.2 Auction duration

Auction duration is another key decision variable from the sellers perspective. Intuitively longer duration gives time to more bidders to participate in an auction. This view holds true when bidders arrival time is a random variable [22]. In the previous section potential reasons for more bidders leading to higher final bid prices are discussed. In addition the larger number of bidders can lead to herd behavior increasing the number of bids further and resulting in higher final bids [15]. Furthermore, an empirical study found that longer auctions result in significantly higher prices [13]. While longer auction generally expect more bidders to participate, they do not always perform better than short auctions. Some articles found that shorter auctions result in better final prices. A potential reason is impatience of bidders, which is found to be a cause of bidders bidding in increments larger than the minimum. This so called jump bidding is expected to positively increase revenue [22]. A second potential reason is that shorter auctions can generate more excitement or competitive arousal and as a consequence increase the bidding activity. This increased bidding activity is found to be the total number of bids, and not necessarily the number of bidders. Individual bidders placing multiple bids indicates a higher level of commitment and competitiveness and thus increases the final prices [2].

All of these studies on auction duration were done using field data collected from eBay. Only one study [22] compared the effects of auction duration using both data collected from eBay as well as from a local auction site. This way using eBay data the positive correlation between longer auctions and the number of bidders and bids was confirmed as well as a positive effect on the final bid prices. However, when performing the same methods on data collected from the local auction website the effects on final price was reversed. They found no correlation between the auction duration and the number of bidders or bids. A reason could be that the online auction has a more steady consumer base, while on eBay this is more randomly distributed. On top of that a negative correlation between the duration and the magnitude of jump bids was found. They combination of a stable number of bidders and bids with jump bids causes the final price to be higher in short auctions on the local site.

2.3 Bidder perspective

On the other side of the spectrum is the bidders perspective. Bidders and sellers have opposing goals regarding the outcome of the auction. Where sellers want the highest possible final price, buyers want the lowest possible final price that wins the auction. In order to do so bidders can adopt different strategies to achieve their goals. The model of bidder strategies and auction prices (MBSAP) [7] discovered a set of bidder strategies and their varying effectiveness in both winning likelihood as well as bidder surplus extraction. MBSAP uses three variables related to timing and number of bids to classify bidder strategies. Firstly, the time of entry (TOE) is the time at which a bidder decides to enter the auction. Secondly, the time of exit (TOX) is the time of their last bid. It is worth noting that bidders who place a single bid will have the same TOE and TOX. Then, the last variable is the number of bids (NOB) which is the frequency with which a bidder updates their bid. Several studies have used K-means cluster analysis [14], [7], [19] or segmentation analysis based on K-means [6] in order to define bidder strategies and evaluate their outcomes. In the following sections the different strategies and their effectiveness are discussed.

2.3.1 Evaluators

The first commonly identified bidder strategy is the evaluator [7], [6], [19] or early biding strategy [14]. Bidders using this strategy often place only one bid during the early or middle stages of an auction. This single bid most likely represents their maximum willingness to pay, indicating these bidders think they can estimate the true value of an item [7]. The evaluator strategy is found to be adopted more than opportunist strategy by single-unit bidders [19].

A benefit to an evaluator strategy is the minimized monitoring cost for an auction [7]. However, Cui et al. [14] found that compared to other strategies the early bidding strategy is the worst performing strategy regarding winning likelihood, bidder surplus and strategy satisfaction. Furthermore, evaluator strategies were found to be inferior to other strategies regarding both winning likelihood and surplus [7].

2.3.2 Opportunists

Another commonly identified bidder strategy is the opportunist or sniper strategy. These bidders often enter and exit an auction near the end and place a single bid [7], [6], [14], [19]. Sniping and evaluator strategies are the most popular among bidders. Intuitively sniping should be a good strategy to maximize winning likelihood since there is no time for competition to react. This effect is confirmed by [7] as well as [14], who also found sniping to be superior to other strategies regarding cost saving and satisfaction. Traditionally sniping is believed to be effective in hard-close auction, since there is no time for competition to react. However sniping is found to have an even stronger effect on revenue in soft-close auctions [11].

2.3.3 Participators

The final commonly identified bidder strategy is the participator. These bidders tend to enter an auction early and exit late and make multiple bids during an auction [7], [6], [19]. Participators spend more time on an auction and as a consequence have higher monitoring costs. Participators have a higher bidder surplus compared to opportunists and evaluators [7], which makes up for the higher monitoring costs. Participators are found to have a significantly lower winning likelihood compared to opportunists [7]. This low winning likelihood combined with high monitoring costs probably cause participators to change strategy in future auctions [19].

2.3.4 Discussion online auctions and bidder behavior

In the literature it is found that the client base is likely and important factor together with the auction design on the performance of online auctions. Therefore, for every data set the bidder behavior must be analyzed independently to describe the behaviors of that client base. Furthermore, the bidder behavior studies are performed mostly on eBay data, which is a hard close auction type with small increment bids and generally short duration. Specially when compared to Troostwijk where the duration is rather large averaging over two weeks per auction. Moreover, Troostwijk has soft-close auctions and relatively large bid increments. Thus a differently behaving client base with respect to current literature is expected. For these reasons the bidder behavior analysis might show new insights in the success of each bidder type or even find a completely new bidder strategy.

2.4 Simulation

In this section agent based modeling and simulation techniques are discussed. A framework for agent based modeling and simulation will be explained. Then, the applications of agent based modeling in economics and online auctions specifically is discussed.

2.4.1 Agent-Based Modeling and Simulation Framework

In this section a framework is provided that guides in the translation of a real life process to an appropriate agent-based simulation model. The framework is introduced by C. Macal and M. North in several articles two of which are used to explain the framework [37], [38]. The structure of ABMS models consist of three main elements. The first element is the agents, these represent the people in the real life process. Secondly, the relationships represent the human interactions. Finally, the environment represent the boundaries in which the real process should be modelled. Each of the elements are explained in more detail in the following sections.

Agents

One of the main elements in agent-based modeling are agents. Despite agents being they key for a model to be agent based, there is no agreement in the literature on the definition of what an agent is. With respect to modeling in practise [37] and [38] considered four essential characteristics that make an agent an agent. Firstly, an agent is self-contained, modular and uniquely identifiable. This means agents can be easily identified, distinguished from other agents and recognized. Secondly, an agent is autonomous. The autonomy of an agent allows it to act independently in the environment. The third essential characteristic for agents is the requirement of a state. The state of an agent is the set of its attributes and behaviors, which can vary over time. The final essential characteristic is that agents are social. This means that agents behavior can be influenced by interactions with other agents. On top of the four above mentioned essential characteristics there are some additional characteristics that are worth mentioning, these are more situational depending on the goal of the model. Some examples are that agents have goals to which the can modify their behavior. Agents can adept based on its individual experiences or the experiences of the population. Furthermore agents can be heterogeneous, meaning they can differ in behavior, goals, resources, and other attributes.

Relationships

The second main element for agent-based models is the relationships between agents. In the previous section the most important attributes and individual behavior of agents is discussed. This sections focuses on the agents interactions with each other. There are two primary issues in modeling the interactions. Firstly, the connectedness of agents need to be specified. This means which agents are connected with each other. Secondly, the mechanisms of the interaction dynamics need to be specified. The reason these two issues need to be specified is that agents make decisions based on their local information. This local information is obtained by agent's interactions. Agents are typically connected with a subset of agents and not with the entire population. A subset of connected agents are called neighbours, and can change during a simulation. The way agents are connected in a model is the topology. There are a few possible topologies for agents-based models, for example the agents location in a 2D or 3D space or the agents social group can be used to define a neighborhood. Independent of the topology the social interaction can only occur during a certain time with a limited number of agents.

Environment

The final key element for agent-based models is the environment. Similarly to interactions between agents, interactions with the environment provide agents with information. This information is typically in the form of boundaries and constraints as time, space and resources.

2.4.2 ABMs in economics

One of the main aspects that makes simulation studies valuable is the ability to explain real world phenomena that are costly or impossible to study in laboratory or field experiments [17]. Specifically in the field of economics ABMs offer two added values. Firstly, ABMs allow for more descriptive richness by being able to describe ecologies of agents interacting through non-obvious network structures, learning from information and competing in imperfect markets. Additionally, the flexibility in both input and output validation add value to the modelers of ABMs [17].

2.4.3 ABMs in online auctions

Agent-based simulation has proven its value in the field of online auctions in a wide range of applications. Several studies in the field of electricity auctions used agent based simulation in some way to model human bidding behavior [54], [16]. In this market bids have the form of a price-quantity pair. This illustrates the flexibility and efficacy of agent based modeling in online auction research. However, these types of auctions strongly differ from the ascending English auction which is the scope of this research. Another application of ABMs in online auction is shown in [23]. Here an agent based simulation optimisation methodology is proposed to determine an optimal online auction policy to control inventory in the agriculture supply chain.
More relevant to the scope of this research are the next applications of ABMs in online auctions. Where the auctions are of the single unit English ascending bid type. In preliminary agent based simulation work [41] attempt to model human bidder behavior that they anecdotally observed in some online auctions. There are exactly two strategies modeled, early bidders and snipers. Each have different behaviors regarding the timing of their bid and their valuation. This study aimed to study interactions of the two types of bidders as well as the performance of the strategies. They found that early bidders can win with lower prices than snipers on average, but with lower probability. They also found that winning prices and probabilities of snipers show that the strategy is effective. A study [42] built on their work by allowing snipers who bid in the last minute of an auction to enter a sealed bid auction. This allowed for all snipers to adjust their final bid. This additional auction resulted in an increase of the price by more than five percent on average.
Another study [32] compared the results of a Markov chain model with the results of an agent based simulation. They modelled bidder behavior by setting a fixed maximum bid and uniformly random willingness to pay for every bidder. Also the duration, minimum bid increment, reserved price, and the number of bidders are fixed in every auction. They performed four experiments which differ in maximum bid, and available information to bidders. These experiments showed that the simulation has some advantages over the Markov chain model. Firstly, the simulation performs equally well in approximating the expected revenue of single-item auctions. Then, it allows analyst to scale the auction complexity and still match well with mathematical analysis. It can model simple and complex bidder behaviors with realistic assumptions. Finally, it can be applied to a wide range of other scenarios and complexities.

2.4.4 Discussion simulation online auctions

The current state of the literature on online auction simulation is in its early stages. Where currently simulation models have two types of bidders, with a hard close ending, often with a fixed number of bidders. Bidder behavior is typically modeled in a specified way not allowing for much variance. The literature can be improved by building a simulation model with more bidder types, more price categories, and less fixed behavior. A more detailed and expanded simulation model can help understand the online auction process better as well as find possible effects of changing design parameters. Furthermore, potential limitations or other issues involving analysis of online auctions using simulation model can be found.

2.5 Validation

One of the most important tasks when conducting a simulation study is to determine the accuracy of the simulation model with respect to the real system [51]. Validation is defined as the process that determines whether the conceptual model is a reasonably accurate representation of the real world [31] and the output of simulations is consistent with the real world output. Moreover, verification is the process that determines whether the conceptual model is correctly programmed. And calibration is an iterative process of adjusting parameters in the programming implementation with the purpose of improving agreement with the real world. Together, validation, verification and calibration are at the core of creating acceptable simulation models. Where validation is achieved through calibration of the model until an acceptable level of accuracy is reached.
Modelers are responsible to find and choose appropriate techniques to assure accuracy and credibility of their model [51]. In the next parts some simulation validation approaches are introduced which can guide modelers in their development process. Additionally, a classification scheme for the level of empirical validity of an ABM is presented. And four dimensions to help decide the best validation techniques are discussed.
Empirical validation involves the comparison between data generated from a simulation model and data generated in the real world process [18]. This study identified methodological issues with empirical validation on ABMs. For example, how does one compare a single observed trace in the real world with the distribution of traces from the model? to what extend can we consider simulated traces to be stylized facts, or counterfactuals? Then stylized facts obtained under a specific set of inputs might not necessarily hold true under different arrangements and robustness analysis must be performed before using ABMs for policy analysis exercises [18]. Another issues is that there is no consensus amongst modelers on the validation techniques that should be used to construct and analyse their models. In order to make an informed decision regarding the empirical validation to be used [18] identified four dimensions in which ABMs differentiate:

1. The nature of the objects under study; This regards the empirically observed stylized facts that the model is seeking to explain
2. The goal of the analysis
3. Modeling assumptions
4. The method of sensitivity analysis
    (a) micro-macro parameters
    (b) initial conditions
    (c) across-run variability

a study [17] discussed three main approaches to guide modelers in the empirical validation of their models. The approaches are introduced below:

1. **Indirect calibration approach** the indirect calibration approach consists of four steps:

   step 1) Output validation

   step 2) Indirectly calibrate the model by focusing on the parameters that are consistent with the output validation

   step 3) Empirical evidence on stylized facts is used to restrict the space of parameters and the initial conditions if the model is non-ergodic

   step 4) Deepen understanding of causal mechanism that underlie the relevant stylized facts as well as exploring new stylized facts

2. **Werker-Brenner approach** The Werker-Brenner approach consist of three steps:

   step 1) Use existing empirical knowledge to calibrate initial conditions and parameter ranges.

   step 2) Empirical validation of the output for each of the model specifications from step 1).

   step 3) Another round of calibration using the surviving set of models.

3. **History friendly approach** The history friendly approach builds models on the basis of a range of data, from detailed empirical studies to anecdotal evidence to histories in the industry under study. This approach is used to assist in the modelling as well as the validation. Furthermore, it should guide in the specifications of agents, behaviors, interactions, environment, decision rules, initial conditions, parameters, key variables and so on. Finally, the data is used to empirically validate the simulation output with the actual history of the industry. An important limitation to note regarding this approach is the model is often built on the history of a single company and not an entire industry.

A general classification scheme judging the level of empirical validity of ABMs has been proposed by [4] and [8]. The scheme consists of four levels:

level 0: The model is a caricature of reality, as established through the use of simple graphical devices

level 1: The model is in qualitative agreement with empirical macro structures, as established by plotting the distributional properties of agent population. This is the easiest way to matching stylized facts.

level 2: The model produces quantitative agreement with empirical macro-structures, as established through on-board statistical estimation routines.

level 3: The model exhibits quantitative agreement with empirical micro-structures, as determined from cross-sectional and longitudinal analysis of the agent population

2.6 Data mining and machine learning in online auctions

Data mining and machine learning have a large role in the online auction literature. It has been utilized in a wide range of applications and goals. A lot of research has been done with the goal to detect any type of auction or bidder frauds. Some examples of auction fraud are auctioning defect items, incorrect information about the condition of items or failing to deliver an item [12], [1]. A common type of bidder fraud is shill bidding [33], [44], [47], [48]. Besides fraud the focus of many studies was the price dynamic of online auctions [25], [26], [29]. Finally, many studies are done on finding and defining bidder behaviors and strategies as is reported in section 2.3. The majortiy of studies used an expanded form of either k-means clustering or k-medoids clustering to partition their data. Where the main difference is that the centroids resulting from the k-means algorithm are not necessarily points in the real data, while the medoids resulting from the k-medoid algorithm are. These clustering techniques are discussed in more detail below.

2.6.1 K-means Clustering

A frequently occurring problem in any scientific field is the clustering problem. Where the goal is to find groups of data points in a given set. Such that each off these groups can be defined as a region in which the density of data points is locally higher than other regions [34]. Solving this problem exactly is NP-hard, which is why the k-means algorithm was proposed by Lloyd [35]. In practice the k-means algorithm is by far the most popular clustering method for scientific and industrial applications [9], [24]. However, it is the speed and simplicity that makes the k-means algorithm so appealing and not its accuracy [3]. Below the k-means algorithm is reported and explained. Furthermore, some issues with k-means clustering are discussed.

The k-means algorithm begins with k arbitrary centers chosen uniformly random from all data points. The k-means algorithm then searches locally optimal solutions with respect to the clustering error. Each data point in than assigned to its nearest center. Then new centers are recomputed as the mass of all data points assigned to a center. These steps are repeated until the process stabilizes. The algorithm is formulated below [3]:

1. Arbitrarily choose an initial k centers $C = \{c_1, c_2, ..., c_k\}$.
2. For each $i \in \{1, ..., k\}$, set the cluster C_i to be the set of points in X that are closer to c_i than they are to c_j for all $j \neq i$.
3. For each $i \in \{1, ..., k\}$, set c_i to be the center of mass of all points in $C_i : c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
4. Repeat step 2 and 3 until C no longer changes

Where we are given an integer k and a set of n data points in $X \subset \mathbb{R}^d$.

One major issue causing the inaccuracy of the k-means algorithm is the sensitivity to the initial centers [34], which are chosen uniformly random. In practise the workaround for this issue is to run the algorithm multiple times with different initial centers and keep the best result [3]. On top of this improvements can be made with the initial selection of clusters. A new method is suggested by [3] where they suggest a combined algorithm called k-means++. Their suggestion is to select the first cluster uniformly random, which is similar to the k-means method. The difference is in the remaining k-1 initial clusters, which are selected and added one by one with weighted probability. These probabilities are weighted with respect to the distance of all data points to their nearest cluster. This is repeated until k initial clusters are selected. The next steps are the same as the original k-means algorithm. The algorithm is formulated below [3]. The k-means++ algorithm has shown to consistently find better clustering with faster computation times than the original k-means.

1a. Take one center c_i chosen uniformly at random from X.

1b. Take a new center c_i, choosing $x \in X$ with probability $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$.

1c. Repeat step 1b. until we have taken k centers altogether.

2-4. Proceed as with the standard k-means algorithm.

Another major issue is to decide the number of clusters k. This decision is often made ad hoc based on prior knowledge, assumptions and practical experience. The are several studies focused on optimizing the number of clusters [21]. This is a very interesting topic on its own, however this is beyond the scope of my project and therefore will not be further discussed.

2.6.2 K-medoid Clustering

Another popular data partitioning technique is the k-medoid technique [28]. The k-medoid algorithm consists of two phases which are presented below. In the first phase k initial objects are selected as follows.

1. Consider an object i that has not yet been selected.
2. Consider a non selected object j and calculate the difference between its dissimilarity D_j with the most similar previously selected object. Then calculate its dissimilarity $d(j, i)$ with object i.
3. If the difference is positive, object j will contribute to the decision to select object i. Therefore we calculate $C_{ji} = max(D_j - d(j, i), 0)$.
4. Calculate total gain obtained by selecting object i : $\sum_j C_{ji}$
5. choose the not yet selected object i that maximizes $\sum_j C_{ji}$

This process is repeated until k objects have been found. In the second phase the set of selected objects is improved by considering all pairs of objects (i,h) for which i is selected and h is not. To calculate the effect of a swap between i and h on the value of clustering the first two steps are carried out.

1. Consider a non selected object j and calculate its contribution C_{jih} to the swap.
   - a If j is more distant from both i and h than from one of the other representative objects C_{jih} is zero.
   - b If j is not further from i than any other representative objects $d(j, i) = D_j$ two situations must be considered:
   - b1 j is closer to h than to the second closest representative object: $d(j, h) < E_j$, where E_j is the dissimilarity between j and the second most similar representative object. In this case the contribution of object j to the swap between i and h is $C_{jih} = d(j, h) - d(j, i)$.
   - b2 j is atleast as distant from h than from the second closest representative object $d(j, h) \geq E_j$. In this case the contribution of object j to the swap is $C_{jih} = E_j - D_j$.
   - c j is more distant from object i than from at least one of the other representative objects but closer to h than to any representative objects. In this case the contribution to the swap is $C_{jih} = d(j, h) - D_j$.
2. Calculate the total result of a swap by adding the contributions $C_{jih} : T_{ih} = \sum_j C_{jih}$.
3. Select the pair (i,h) that minimizes T_{ih}.
4. if the minimum T_{ih} is negative the swap is carried out and the algorithm returns to step 1. If T_{ih} is zero or positive the value of the objective cant be decreased and the algorithm stops.

2.6.3 Machine learning discussion

The k-medoid clustering algorithm has the benefit of dealing with outliers better than the k-means algorithm. Also it doesn't depend on the order in which the objects are presented. However, a key downside is the time and memory requirements [28]. This causes issues especially with large data sets. Furthermore, the literature discussed in 2.3 all found clearly defined bidder behavior using k-means clustering. For this reason and the fact that performing the k-means++ algorithm solves some key issues from the original k-means algorithm, the data partitioning is done based on k-means++ clustering.

Chapter 3

Data Exploration

In this chapter the available auction data is discussed. Firstly. the method of gathering the data is explained. Then, the structure of the data is presented. Next, the information contained in each table is presented.

3.1 Data Gathering

The data used in this project is real online auction sales data from Troostwijk in two distinct time frames. The data contains information on the auctions, the lots and bids. Over the course of two years Troostwijk has experimented with different strategies on the starting price of auctions. In 2018 data was collected in the period 12/21/17 until 6/30/18. In this period a starting price of roughly $50-60\%$ of the estimated value was mostly used. The 2019 data was collected in the period 3/28/19 till 6/28/19, in this period the starting price of approximately 10% of the estimated value was mostly used. In the figure 3.1 the total number of auctions, lots, bids and bidders from both years are shown. As we can see the number of lots auctioned in these periods is similar, but the number of bids and bidders is larger in 2019. This shows an expected difference between the two periods.

	2018	2019
Auctions	1100	455
Lots	147346	144502
Bids	1175044	2018011
Bidders	34358	44785

Figure 3.1: Data gathered

3.2 Data Structure

The data is provided in 6 separate tables, 3 tables from data gathered in 2018 and 3 tables for data gathered in 2019. These tables contain information on the auctions that ran in the periods considered, information about the lots being auctioned, and finally information about the bids on those lots. These tables can be connected trough their primary keys, this results in the structure shown in 3.2.

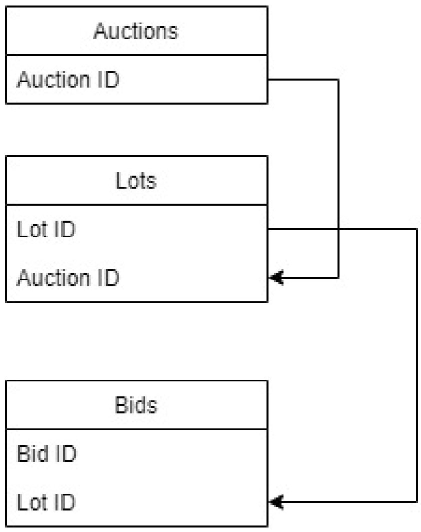


Figure 3.2: Data structure

The tables are connected as explained above and result in a relational database, creating the relational database is done in Microsoft Access. The reason for the use of this type of database is that filtering based on attributes from the lots can also filter the bids data. For example if a filter on lots data with an estimated value is required, the bids data on only those lots can be filtered simultaneously. Furthermore, information stored in one table can be added to the an other table using the relationships presented in figure 3.2. This allows for efficient data cleaning as well as filtering and selection of data.

3.3 Available information

In this section the available information in each of the tables is presented. As explained in the previous section the information is stored in three separate tables for each of the periods of data gathering. In the auction tables 3.1 the identifiers are auction ID and the auction title. Further information is the duration of an auction stored as a starting data and a closing date. In 2019 a branch description of the lots sold in the auction was added.

Attributes	**Data Type**
Auction ID	Numerical
Auction Title	String
Auction Start Date	DateTime
Auction Closing Date	DateTime
Branch Description(2019)	String

Table 3.1: Auction Data

Then in both tables of the lots data the relevant information is presented in 3.2. Firstly, the Lot ID is the unique identifier for the lots as well as the lot title. Lot numbers identify lots within an auction and can repeat for other auctions, as such those are not unique values. Then, the auction ID is the same for all lots in the corresponding auction.

The status of an auction explains whether or not that lot was sold or the auction was unsuccessful. There are multiple reasons for the sale of a lot to fail. The number of items states the amount of items sold in a singular auction.

The buyer account id is the unique identifier for the winning bidder of an auction with the buyer country being the country where that bidder lives.

The estimated value of an item is the estimation made by Troostwijk of the value of that lot. The starting price is the price at which the auction starts, and the reserve bid is the minimum price required for a successful sale of that lot. Then, the current bid is the final bid on that lot made by the winner.

Finally, the main category and sub category are the categories assigned by Troostwijk in which the lots are advertised.

Attributes	**Data Type**
Lot ID	Numerical
Auction ID	Numerical
Lot Number	Numerical
Status	Category
Number of items	Numerical
Buyer Account ID	Numerical
Buyer Country	Category
Estimated Value	Numerical
Starting Bid	Numerical
Reserve Price	Numerical
Current Bid	Numerical
Main Category	Category
Sub Category	Category

Table 3.2: Lot Data

Then the relevant information in both bids data tables is presented in 3.3. In the bids tables information on the corresponding auction and lots is stored in the attributes: Auction ID, Lot ID, and Lot number. Then the account ID is the unique identifier of the bidder placing that bid, the bid price is the value of that bid. Finally, the bidding date time is the exact moment the bid was placed and the closing date time is the time the auction of the corresponding lot is closed.

Attributes	**Data Type**
Auction ID	Numerical
Lot ID	Numerical
Lot Number	Numerical
AccountID	Numerical
Bid Price	Numerical
BiddingDateTime	DateTime
ClosingDateTime	DateTime

Table 3.3: Bid Data

3.4 Data Description

In this section initial descriptive statistics from the data is reported and discussed. The first tables are shown below 3.3 and 3.4 and contain information on the number of items sold in one auction, the estimated value, the starting price, reserve price and the final sale price.

attributes	means	SDs	prt25	prt50	prt75	minimum	maximum
Number of Items	14	194	1	1	2	1	24000
Estimated Value	272	2958	0	0	50	0	350000
Starting Price	171	1311	10	15	75	0	200000
Reserve Price	1,4	49	0	1	1	0	10000
Final Bid Price	349	4972	10	35	130	0	1600000

Figure 3.3: Descriptive statistics A 2018

attributes	means	SDs	prt25	prt50	prt75	minimum	maximum
Number of Items	16,7	959	1	1	2	1	312000
Estimated Value	555	5765	10	50	200	-1	1300000
Starting Price	96	3845	10	10	10	0	1300000
Reserve Prices	0,97	0,41	1	1	1	0	100
Final Bid Price	490	3888	15	55	190	0	860000

Figure 3.4: Descriptive statistics A 2019

The first thing to discuss is the number of items in auctions. The results are very similar for both years, with over 50% of the auctions are just 1 item. The mean number of items are heavily skewed by a couple of very large numbers. The most important thing is that there are enough auctions with just 1 items since the focus will be on those auctions for the rest of the project. Another attribute is the estimated value of an auction.

There are some clear differences between the results in 2018 and 2019. Firstly, in 2018 the median estimated value is 0. This means that over half of the auctions are not estimated, a reason could be that in 2018 only the more valuable items were estimated. This also causes the mean estimated value in 2018(€100) to be smaller compared to 2019(€555). In 2019 the first quantile is 10, this indicates that also the lower valued items were estimated in 2019. The minimum valuation is -1, this is a value that was used for items that were not estimated. Further analysis showed that the amount of items with valuation in 2018 and 2019 are 67.401 and 130.472 respectively.

The starting price is a major attribute for this project. Troostwijk has adopted different strategies in 2018 and 2019. We can see the value of both first quantiles is €10, the differences start with the medians of €10 and €15. The difference in the third quantile is ever larger with €75 and €10. However, to identify differences in starting price strategy the starting price needs to be viewed in relation to the estimated value. In 2018 we have a higher mean starting price with lower estimated value compared to 2019. This indicates a clear difference in the starting prices and confirms the difference in strategies.

The final bid prices follow similar patterns in 2018 and 2019. The values in 2018 are slightly lower in all quantiles as well as the mean. Moreover, they both are heavily skewed by a few very expensive items.

The reserve price does play a large role in online auction literature, however in this table it clearly shows that most auctions have a reserve price of 1. Since the reserve price is lower than the starting price it is irrelevant and will be ignored for the rest of the project.

Two more tables show additional information on the data. Information on the buyers, item categories and auction outcomes is shown in the tables 3.5 and 3.6.

attributes	Unique values	Top value	Top frequency
Buyer Country	78	NL	48297
Buyer ID	15582	-	
Main Category	122	Living Furniture and Accessories	12845
Sub Category	1025	Further Inventory	5392
Status	7	sold	99583

Figure 3.5: Descriptive statistic B 2018

attributes	Unique values	Top value	Top frequency
Buyer Country	89	NL	51678
Buyer ID	17236	-	
Main Category	133	Catering and Horeca	11535
Sub Category	1255	Further Hand Tools	3955
Status	7	sold	104131

Figure 3.6: Descriptive statistics B 2019

In both data sets the most occurring nationality for the winning bidder is Dutch. Furthermore, auctions were won by bidders from 78 countries in 2018 and 89 in 2019. In 2018 there were 15.582 unique winners out of 34.358 participating bidders. In 2019 there were 17.236 unique winners out of 44.785 participating bidders. Moreover, the amount of main and sub categories is similar for both years.

The final and most important attribute is the status. The status of an auction refers to the outcome of that auction. A sold auction is a successful auction. The other 6 values are failed auctions for a variety of reasons, as shown in figures 3.7 and 3.8. For the remainder of the project only the data from sold auctions will be used.

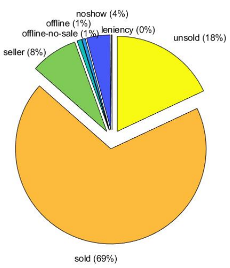


Figure 3.7: Pie chart Status 2018

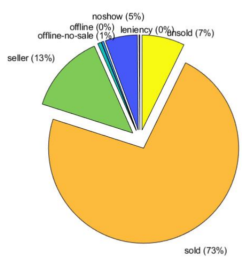


Figure 3.8: Pie chart Status 2019

Chapter 4

Data Preparation

In this chapter the steps taken to prepare the data to be used as input for the simulation model are reported. The first section explains the data cleaning steps taken, and reports the final data set used for in the project. Then the data transformation steps on this final data is reported. Moreover, the results from the K-means algorithm are reported and discussed. Finally, the resulting bidder strategies are defined and their behavior is reported.

4.1 Data Cleaning

In this section some issues with the data and their solutions are reported. A major issue with the data was the lacking or inconsistent unique identification for the lots. This was resolved by automatically creating new unique ID's for the lots such that the data structure explained in section 3.2 can be realized. Additionally, the bid data was presented without any unique identification and is added as such.

4.1.1 Lots

For the lots data the first issue was incompatibility between the data sets from the two years. One incompatibility was on the attributes Main/Sub category in the lots data. In the 2018 set the values were displayed in Dutch while in the 2019 set the values were in English. To solve this the Dutch values were translated to English using the category names on the website. A couple more small issues were that the data contained some missing values in relevant fields such as a missing winner account ID as well as some negative winning bids, which is incorrect data. These lots, and all the related bids were removed from the data.

4.1.2 Auctions

The first issue with the auction data was that there were a lot of duplicates, these were removed. After the data structure was realized it became clear that some auctions completely lacked information on lots or bids, these auctions were removed as well. Another issue was that some auctions did not fully fit within the previously defined periods in time of which the data was gathered. As a consequence the related lots related bids data may not be complete, with the possibility of missing bids the data does not reflect the real process of auctioning those lots. Therefore, those auctions, their related lots and related bids were removed.

4.2 Data Selection

The selection of data is aimed at finding the most appropriate subset of the data to use as the foundation of the simulation modeling process. The simulation process will be of single unit auctions. Hence, the first selection criteria of the data is all lots with the value on a the attribute Number of Items to be 1.

Another selection criteria is the estimated value of lots larger or equal to 5. The lower range of values show highly inconsistent data which makes it less appropriate to built a simulation model.

The final criteria of the data selection is that the status of lots have to be sold. This is because the simulation model will reflect the real process of successful auctions.

4.3 Final Data

After the data cleaning and data selection the following data 4.1 is used for the data transformation, simulation model and validation.

	2018	**2019**
Auctions	278	426
Lots	35.335	71.011
Bids	420.508	1.182.152
Unique Bidders	22.186	37.805
Unique Winners	9.260	14.827

Table 4.1: Final data

From this table some interesting results can be deducted. In 2018 the average number of bids placed in a successful auction of a lot was $\frac{420.508}{35.335} \approx 11,9$ and $\frac{1.182.152}{71.011} \approx 16,6$ in 2019. Then these bids are placed by 22.186 and 37.805 unique bidders in 2018 and 2019 respectively. Which means bidders placed $\frac{420.508}{22.186} \approx 19.0$ and $\frac{1.182.152}{37.805} \approx 31.3$ average bids in an auction they participated in. From these active bidders only 9.260 and 14.827 bidders won an auction. This means that most active bidders participating in auctions do not win any auction.

4.4 Data Transformation

In this section all steps that were required to prepare the data to be suitable as input for the simulation model are explained.

4.4.1 Pricing Strategies

As previously explained in the introduction 1 Troostwijk used different pricing strategies during both periods of data gathering. However, no such information is stored in the data as presented. The starting price and estimated value of items are the attributes used to categorize the pricing strategies and price category. The price category is categorized using all estimated values between 10 and 100, 100 and 1000, 1000 and 10.000 and finally 10000 or higher. Then the pricing strategies are categorized in a low, medium and high pricing strategy. The values used to categorize the strategies are the ratio between the starting price and the estimated value. The resulting values between 0 and 33 are considered a low pricing strategy, the values between 33 and 66 are medium pricing strategy and the values between 67 and 100 are high pricing strategy. These two categorical attributes are considered in pair which results in 15 possible categories, for all five price categories all three pricing strategies were used.

4.4.2 Standardizing Bid Timings

The timing of bids available in the data is in the date time format. While this is an accurate way of storing that information it is not usable across auctions with different start and end times. Therefore, the timings of bids are transformed into a standardized format scaled between 0 and 10. This way the bid timings of all lots in both periods of data gathering can be used equally. In order to standardize bid timings the date time format is converted to a milliseconds format. The three attributes required to then standardize the bid timings are: The start time of an auction, the closing time of an auction and the exact moment the bid was placed.

The information on the start time of and auction is missing in the bids data as presented by Troostwijk. Therefore, the start date from the auction data was added to the bids data. The closing time as well as the exact moment the bid was placed are present in the bids data. With all required information present in the milliseconds format the bid timings are standardized using the following formula: $\frac{closingtime - bidtime}{closingtime - starttime} * 10$
This results in the bid timings scaled to fit the range $[0, 10]$. These standardized bid timings are used to compute the time of entry and the time of exit for the k-means algorithm in the following section 4.4.3.

4.4.3 K-means Tuple

The standardized bid timings were used to compute the values for Time of Entry (TOE), Time of Exit (TOX), and the number of bids (NOB) from a bidder in one auction. These values were required as input for the k-means algorithm that is used to define the bidder strategies as explained in the literature review 2.6.1.

For every bidder participating in an auction the tuple (TOE, TOX, NOB) is created. The TOE is the standardized bid timing of the first bid placed by a bidder in that auction. The TOX value is the standardized bid timing of the last bid placed by a bidder in that auction. The values TOE and TOX can be equal when a bidder placed only one bid in an auction. The value for NOB is the total number of bids placed by a bidder in that auction.

4.4.4 Bidder Arrival Fitting

The information stored in the tuple (TOE, TOX, NOB) as computed in 4.4.3 was used to fit the bidder arrival process. The information is stored as one row for every unique bidder who participated in an auction of a lot. Hence, when bidders participate in multiple auctions there will be a row with this bidder ID for every auction. This also means that auctions with multiple bidders have multiple rows. Therefore, by counting the number of rows with the same Lot ID gave the number of bidders who participated in that auction. This resulted in a column of number of bidders in an auction. That column was then combined with all the categories as explained in subsection 4.4.1.

N way ANOVA has been used to determine whether the grouped subsets using the grouping variables price category and pricing strategy have different means. The results in figure 4.1 shown that all groups have significantly different means. Therefore, a gamma distribution fit is done for every group independently. These results are shown in table 4.2

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
PriceCategory	15646.6	3	5215.5	437.78	0
PricingStrategy	42123.2	2	21061.6	1767.86	0
PriceCategory*PricingStrategy	32544.6	6	5424.1	455.29	0
Error	1257525.8	105554	11.9		
Total	1609450.7	105565			

Constrained (Type III) sums of squares.

Figure 4.1: N way ANOVA

Category	count	mean bidders	Variance bidders	Standard error	a	b
10 Low	944	3,40	8,25	0,09	1,76	1,93
10 Medium	28.945	6,31	13,90	0,02	2,67	2,36
10 High	10.147	4,12	8,42	0,03	2,15	1,91
100 Low	14.690	4,15	11,89	0,03	1,78	2,33
100 Medium	17.677	4,43	7,25	0,02	2,60	1,71
100 High	23.283	3,61	6,37	0,02	2,35	1,53
1000 Low	11	3,09	12,09	1,05	1,56	1,99
1000 Medium	822	15,13	69,45	0,29	2,71	5,59
1000 High	306	7,79	22,44	0,27	2,45	3,18
10.000 Low	148	6,10	12,54	0,29	2,66	2,29
10.000 Medium	5.974	9,90	34,18	0,08	2,61	3,79
10.000 High	2.619	5,76	15,63	0,08	2,09	2,76

Table 4.2: Bidder Arrival Distributions

These numbers are computed based on the bidders who arrived at an auction and actively participated. This does not include bidders that arrive at an auction and decided not to place a bid. Bidders can enter auctions and not place a bid for a lot of reasons. For example, a bidder can enter an auction and value the item at a price lower than the highest current bid on that item. To account for the probability of a bidder arriving at an auction and not placing a bid the resulting expected bidders from the gamma distribution will be multiplied to more accurately reflect the real participating bidders in every category.

4.4.5 K-means Bidder Strategies

The tuple (TOE, TOX, NOB) as computed in subsection 4.4.3 is used to find and define the bidder strategies used by the bidders in the data. At this point the data from both periods is combined to keep the strategies adopted in both periods consistent. The partitioning method used to cluster the bidder strategies is the K-means++ 2.6.1 algorithm as explained in the literature review, the algorithm is the default K-means function in Matlab. The K-means++ algorithm find the best clusters on the data by searching locally optimal solutions with respect to the clustering error. The clustering error used is the squared euclidean distance of data points to the nearest centroid. The algorithm is replicated 20 times with a different initial cluster centroid position. The solution with the lowest final clustering error is returned.

One of the difficulties with the k-means algorithm is deciding the number of clusters K. This decision was made based on the bidder strategies literature 2.3 as well as practical knowledge. The expected strategies were sniper, evaluators and participators. However, we have practical knowledge on participators that there is a high variability in the number of bids placed by bidders who are expected to be in this cluster. Therefore, the expectation is that the participators can be split into two distinct strategies. This resulted in the 4 expected strategies snipers, evaluators and two types of more active bidders. Hence, the K in the K-means++ algorithm is set to 4.

An important strategy that is mentioned in the literature 2.3 is the automated bidder strategy. In this strategy bidders leave the bidding to automated software based on preset commands such as as maximum bid and the increment used to overbid other bidders when possible. This way the bidder is not required to actively monitor the status of auctions. Bids placed by automated bidder software can be recognised by the timing of their bid and the previous bid being exactly identical in milliseconds. After the data was clustered some of these bids were individually analysed. This way these type of bids were found to be placed by all different clusters. In several instances these type of bids were placed by one bidder in an auction, who later in that auction manually placed a bid. This indicates that even though the automated bidding software is used those bidders still actively monitor the auction. This individual analysis of some of the automated bids gives reason to believe the automated bidding is not a strategy but rather a bidding tool adopted by all types of bidders. This supports the strategy not being considered in the decision to set the cluster K to 4.

The data was partitioned in 4 strategies. The behavior of bidders in these strategies can be defined by the tuple (TOE, TOX, NOB). In the figures 4.2, 4.3, 4.4 below the box plots for these attributes are shown. It is clear that the difference in behavior between all 4 strategies is significant.

Time of Entry

In figure 4.2 the differences in behavior between the 4 strategies with respect to the time of entering an auction is shown. The TOE values are the standardized timing of the first bid placed by a bidder in an auction. The first cluster is the participator strategy. Bidders using this strategy mostly enter auctions during the second half of the auction duration. The median value is larger than 9, this indicates half of the participators enter and auction later than that. The auctions on Troostwijk have an average duration of around two weeks. Therefore, the participators often enter auctions in the final day or two of an auction. The behavior of participators entering an auction during the later stages was not expected from the participator strategies found in the literature 2.3

The second strategy is the piranha strategy, these bidders enter auction during any stage. With the quartile values at approximately 2.5, 6 and 9 bidders the time of entering an auction slightly tilts towards the second half of the duration.

The sniper strategy is expected to be at the very end of auctions, this is confirmed by the data partitioning. However, with the first quartile at approximately 9 not all bidders in this cluster enter the auction in the last day. Furthermore, the shape of the box plot as well as the many outliers indicate a strong negative skew. Therefore, the behavior of snipers with respect to entering an auction is not perfectly represented by this data.

Finally, the evaluators enter the auction during the earlier part of an auction up until slightly later than the midpoint of the auction. This behavior is in line with what was expected from evaluators in the literature 2.3.

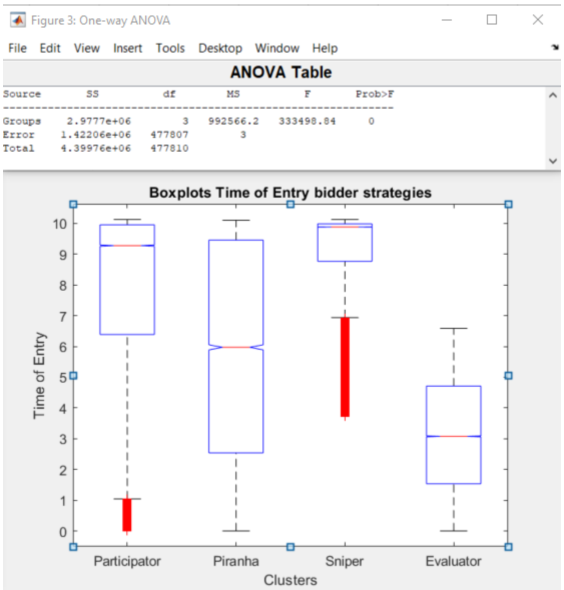


Figure 4.2: Box plot Time of Entry

Time of Exit

In figure 4.3 the differences in behavior between the 4 strategies with respect to the time of leaving an auction is shown. The TOX values are the standardized timing of the last bid placed by a bidder in an auction. The participators, piranhas and snipers have very similar behavior regarding the time of exit. All three of these types of bidders leave the auction towards the end of the duration. The differences between the three are mostly the lower two quartiles. All participators except outliers have a time of exit in the last day of an auction. While the piranha and snipers approximately 25 percent of the bidders leaving an auction before the last day. However, the outliers and shape of the box plots indicate a strong negative skew for both of those strategies. This is a clear indication that the behavior of those strategies is not perfectly represented in these plots.

Finally, the evaluators show the expected behavior. With a third quartile value of approximately 6 most evaluators have left in the early stage of an auction.

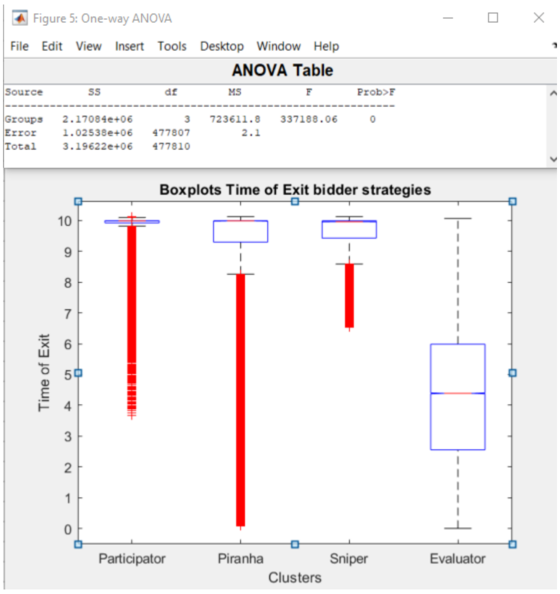


Figure 4.3: Box plot Time of Exit

Number of Bids

In figure 4.4 the differences in behavior between the 4 strategies with respect to the number of bid in an auction by a bidder is shown. The NOB values are all bids placed by a bidder in a single auction. The participators are active bidders throughout the auction placing between 5 and 12 bids in an auction. The piranhas are far more active with most placing between 11 and 28 bids, with some instances of up to 80 bids in an auction.

The sniper and evaluators show very similar behavior with respect to the number of bids placed in an auction with the majority of bidders placing only one bid in an auction. The only noticeable difference between those strategies is the evaluator strategy having more and higher outliers.

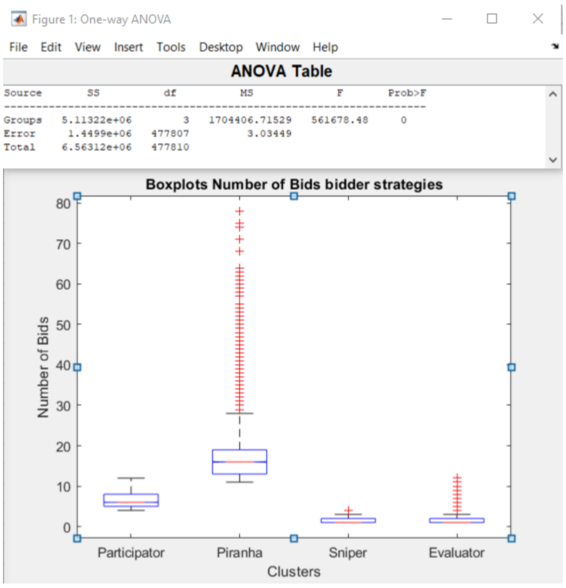


Figure 4.4: Box plot Number of Bids

K-means Strategies Overview

To get a more complete perspective on the behavioral differences between the 4 strategies that resulted from the k-means++ algorithm a 3D-plot is shown in figure 4.5. In this figure the strategies are color coded, this way the clusters were clearly visualized.

The first thing to notice is the linear line across the TOE and TOX axis. Since it is impossible to exit and auction before entering it, all values of TOX are larger or equal to the values of TOE. On the same axis a clear cutoff line between the clusters sniper and evaluator is shown. This cutoff indicates that bidders with a similar low number of bids are partitioned into either sniper or evaluator clusters based on the combination of TOE and TOX values. All time of entry values below approximately 5 are partitioned as evaluators, and above all time of entry values of approximately 7 are partitioned as snipers. In between those time of entry values bidders are partitioned as snipers when the time of exit is sufficiently later than the time of entry, and evaluators otherwise.

Another interesting cutoff is between evaluators and piranhas. Where bidders with extremely early time of entry are partitioned as evaluator for a number of bids lower than approximately 15 and piranhas with more bids. Moreover, a similar cutoff exists between participators and piranhas. Where, bidders are partitioned as participator with lower number of bids compared to piranhas.

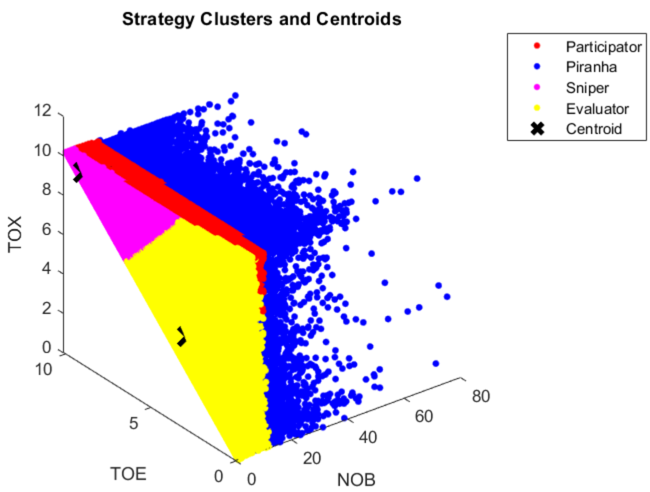


Figure 4.5: 3D plot bidder strategies

4.4.6 K-means discussion

In the section 4.4.5 the results from the K-means++ algorithm are reported. The strategies that resulted from the data partitioning were introduced and the differences between them are explained with respect to the tuple (TOE, TOX, NOB) values.

The first attribute reported is the Time of Entry of the resulting bidder strategies. Some unexpected behavior was found with the sniper strategy, there were many bidders in this cluster that entered the auction way before the final day. This is not in line with the sniper strategy as defined in the literature 2.3. Then, the participators seem to enter auctions rather late, mostly during the final two days. Again, this is unexpected since in the literature participators are found to enter the auction during any stage.

The second attribute is the Time of Exit of the resulting bidder strategies. The behavior from the strategies is mostly in line with the expectations. The evaluators seem to leave the auction as soon as their bid is placed during any stage of an auction. The participators, piranhas and snipers all leave the auction towards the end of the duration. These strategies have a large amount of outliers ranging to rather early times of exit. These outliers are not in line with what can be expected from the bidders in these strategies.

Then the behavior with respect to the number of bids is as expected. The participators and piranhas show more activity during auctions, where the snipers and evaluators mostly place one bid.

After the individual behavior of each attribute the combined results were reported. The complete behavior of strategies showed strange cutoffs at the edges of clusters. These cutoffs illustrate a downside to the K-means algorithm, which is sensitivity to outliers. As a consequence those edges are a bad way to describe the behavior of bidders in those strategies. Despite that the centroids of the clusters show exactly the behavior to be expected from each strategy. Hence, in the modelling of bidder behavior more value is given to the centroids.

4.4.7 Starter Strategy

In the section 4.4.5 four bidder strategies were identified using a K-means algorithm. General results on the behavioral attributes of bidders explained how these bidders act in auctions. A more detailed analysis into the behavior of snipers and evaluators showed a different behavior. Many of the bidders partitioned into these two strategies made only one bid. This is the expected behavior from these two strategies. However, many of those bidders placed their one bid to start an auction specifically. This is likely to happen randomly by either of those strategies. However, these instances of auctions started by a bidder placing only the first bid of that auction is 57.731. Furthermore these auctions were started by only 14.303 unique bidders who showed this behavior. This behavior is adopted by a sufficient number of bidders and started a sufficient amount of auctions to be considered as a bidder strategy on it's own. Therefore, this new 'Starter' strategy is added as a fifth cluster by converting the cluster values of the bidders in auctions that showed the specific behavior of placing only the first bid in an auction.

4.5 Resulting strategies

With the addition of the starter strategy the final five strategies to be used in the simulation are complete. The strategies are not equally popular amongst bidders and have differences in effectiveness. In the following figures 4.6, 4.7 the pie charts of the amount of bidders using strategies and the winning strategies are reported. The most used strategy is are the snipers, who won the most auctions as well. The evaluator strategy is the second most popular strategy, but have very low winning chances. Then the starter and participator strategies are similar in popularity, with good winning probability. Finally, the sniper strategy is not popular but does have high winning likelihood. These figures show the strategies and their odds to win auctions across all auctions.

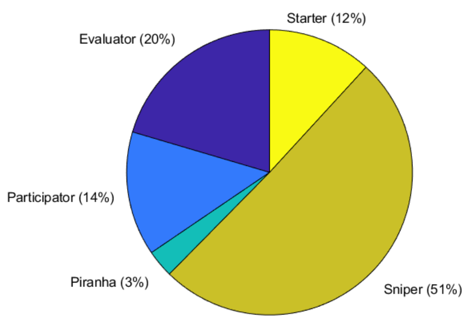


Figure 4.6: Pie chart bidder strategies

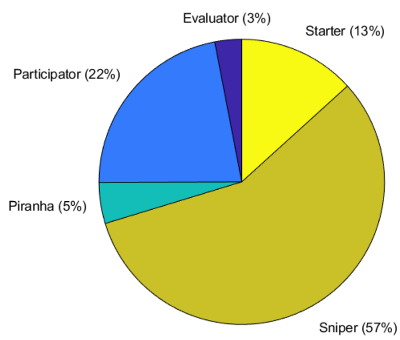


Figure 4.7: Pie chart winning strategies

The results of bidder strategies and performance reported above are aggregate results over all selected auctions during the data gathering periods. However, the likelihood of bidders using certain strategies might be different depending on the item being auctioned. In the following figures 4.8, 4.9 and 4.10 the adopted strategies in auction with certain estimated value ranges are shown. In this figures clear differences in strategy adoption in each price range can be found. Two clear trends that can be seen are the likelihood of bidders using a sniper or starter strategy decreases when the estimated value of the item increases. On the other hand the strategies participator, evaluator and piranha become more popular when the estimated value of an item is higher.

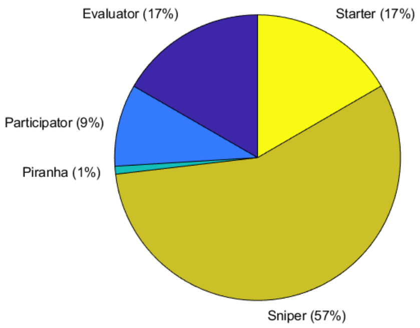


Figure 4.8: Bidder strategies estimation 10

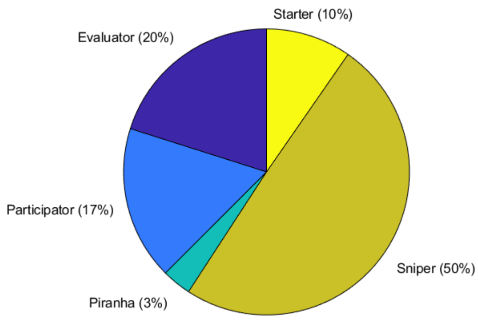


Figure 4.9: Bidder strategies estimation 100

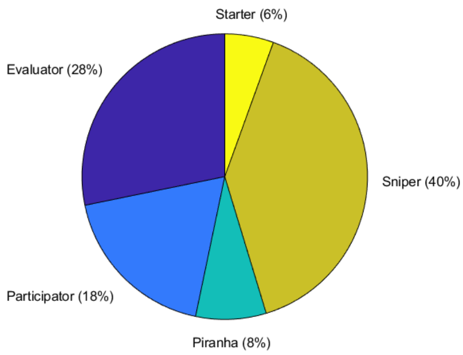


Figure 4.10: Bidder strategies estimation 1000

As a result of the clear differences in strategy adoption with the different price ranges, the performance of those strategies is different in these ranges as well. In the figures below 4.11, 4.12, and 4.13 the winning strategies are shown. It is clear that the starter strategy, which is more often adopted with lower estimation performs well in that price range. The snipers have very high winning chances in the lower two ranges and decent odds in the highest range. The participators and piranhas have strongly increasing winning chances when the estimated value of the auction increases.

Strategy adoption and their performance is an important part of the accuracy of the simulation model. The outcomes reported here are real the results from the real process and should be accurately reflected in the outcome of the simulation model.

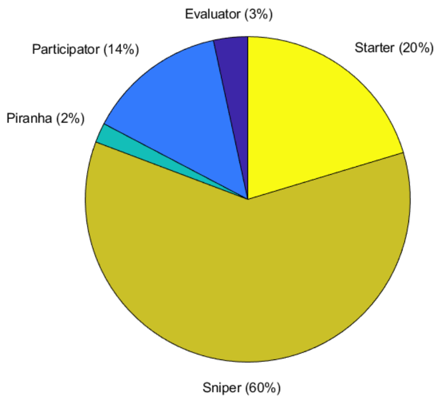


Figure 4.11: Winner strategies estimation 10

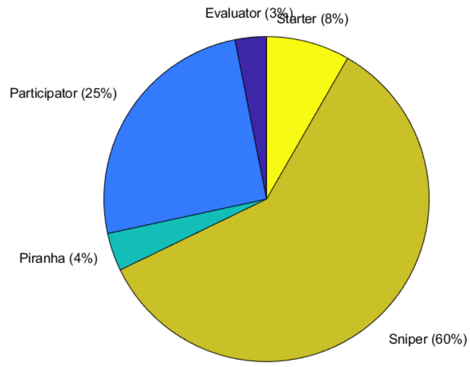


Figure 4.12: Winner strategies estimation 100

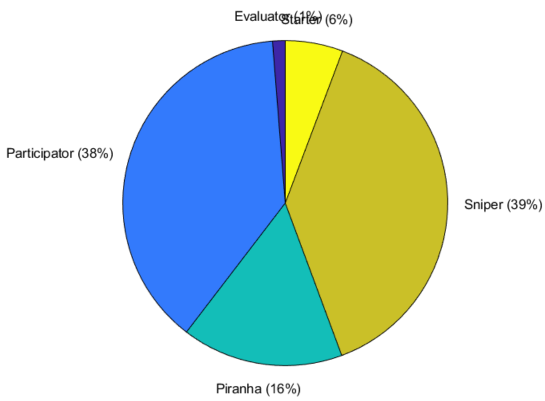


Figure 4.13: Winner strategies estimation 1000

4.6 Bidding behavior distribution fitting

In the previous sections the behavior of bidders in all strategies is presented and discussed. In this section this behavior is fitted with the best distribution fit for every strategy. These fitted distributions are used in the simulation model to program the bidder strategies. This strategies were partitioned based on the tuple (TOE, TOX, NOB). The behavior of the strategies is modelled based on the time of entry, the number of bids and the personal valuation of items. The personal valuation is computed based on the estimated value of items and the highest bid placed by a bidder in an auction. This relational valuation is then used to fit personal valuation distributions for every strategy.

4.6.1 Evaluators

The evaluator strategy behavior is explained by the personal valuation of an item, the maximum number of bids and the time of entering an auction. The personal valuation serves as a maximum bid value a bidder is willing to place. Therefore as long as the current bid plus the minimum bid increment is equal or lower than the personal valuation that bidder will remain active in an auction, this is the first of three constraints for evaluators to be active. In figure 4.14 the fitted distributions on the personal valuations of evaluators is shown. The distributions have a good fits as illustrated by the cumulative probability plot and the numerical fitting results in figure 4.15. The log-logistic fit is slightly better than the lognormal distribution as can be seen by the log likelihood values. However the log-logistic distribution has infinite variance which caused it to be a bad model to program the behavior of bidders. Therefore the best fit in practise is the lognormal fit.

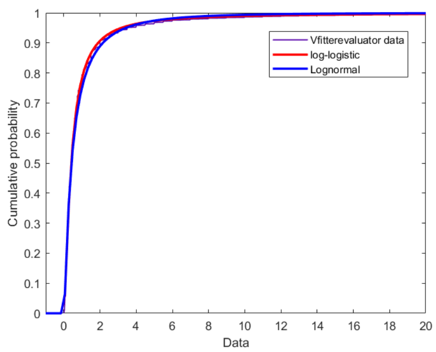
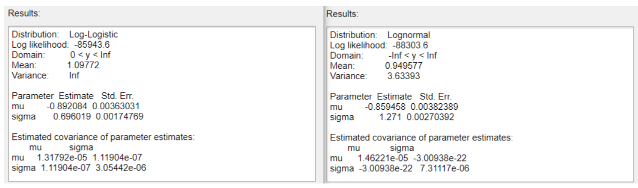


Figure 4.14: Personal Valuation Evaluator



Results:

Distribution: Log-Logistic
Log likelihood: -85943.6
Domain: 0 < y < Inf
Mean: 1.09772
Variance: Inf

Parameter	Estimate	Std. Err.
mu	-0.892084	0.00363031
sigma	0.696019	0.00174769

Estimated covariance of parameter estimates:

	mu	sigma
mu	1.31792e-05	1.11904e-07
sigma	1.11904e-07	3.05442e-06

Results:

Distribution: Lognormal
Log likelihood: -88303.6
Domain: -Inf < y < Inf
Mean: 0.949577
Variance: 3.63393

Parameter	Estimate	Std. Err.
mu	-0.859458	0.00382389
sigma	1.271	0.00270392

Estimated covariance of parameter estimates:

	mu	sigma
mu	1.46221e-05	-3.00938e-22
sigma	-3.00938e-22	7.31117e-06

Figure 4.15: Personal Valuation evaluator results

The second constraint for evaluators to remain active in auctions is the number of bids. The expected number of bids placed by an evaluator is fitted using the Poisson function as shown in figure 4.16 with numerical results in 4.17. In the cumulative probability plot large gaps can be seen between the fitted distribution and the real data. The probabilities of lower value up to 3 bids in an auction is underestimated by the Poisson fit. As we know from the literature, this is the part that is supposed to best explain the behavior of evaluators with respect to the amount of bid revisions. Furthermore, in the K-means results 4.4.5 a large number of outliers was found for the number of bids of evaluators. The fit seems to over fit to the tail end of the number of bids.

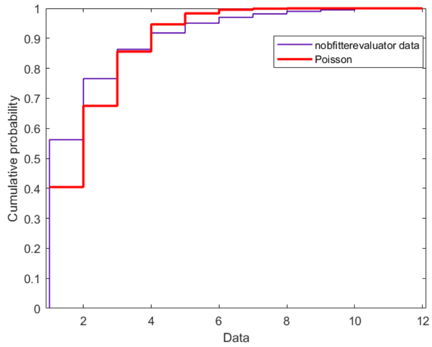


Figure 4.16: Number of Bids Evaluator

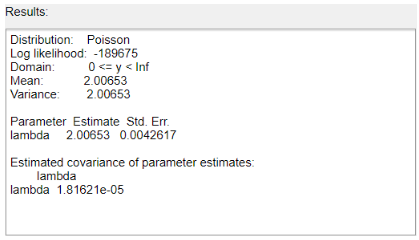


Figure 4.17: Number of Bids evaluator results

The third and final constraint of evaluators to remain active is the time of entry. The duration of an auction must be equal or later than the time an evaluator enters the auction. The time of entry values are shown in the histogram in 4.18. This clearly shows a uniform distribution up to a time of entry value of 6. As briefly discussed in 4.4.5, the cutoff point from the clusters sniper and evaluator is between 6 and 7 depending on the time of exit. This is reflected by the smaller bar on the right side.

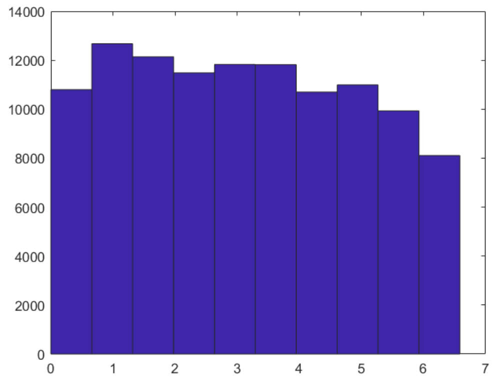


Figure 4.18: Time of entry evaluator

4.6.2 Snipers

Similarly to the evaluators the behavior of snipers is explained by the personal valuation, the number of bids and the time of entry. The first constraint is the personal valuation. The personal valuation was fitted using Log-Logistic and lognormal distributions as shown in figures 4.19 and 4.20. The cumulative probability plot shows the lognormal fit is slightly worse than the log-logistic giving predicting slightly lower probabilities between the values 2 and 6 compared to the data, where the log-logistic is fit better for the lower values. However, as mentioned before the log-logistic distribution has infinite variance and proved to be a bad way to model behavior.

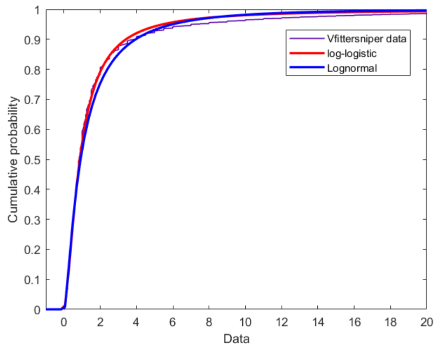
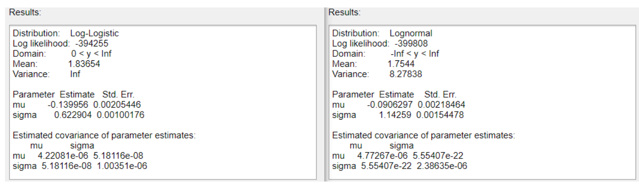


Figure 4.19: Personal Valuation Sniper



Results:

Distribution: Log-Logistic
Log likelihood: -394255
Domain: 0 < y < Inf
Mean: 1.83654
Variance: Inf

Parameter	Estimate	Std. Err.
mu	-0.139956	0.00205446
sigma	0.622904	0.00100176

Estimated covariance of parameter estimates:

	mu	sigma
mu	4.22081e-06	5.18116e-08
sigma	5.18116e-08	1.00351e-06

Results:

Distribution: Lognormal
Log likelihood: -399808
Domain: -Inf < y < Inf
Mean: 1.7544
Variance: 8.27838

Parameter	Estimate	Std. Err.
mu	-0.0906297	0.00218464
sigma	1.14259	0.00154478

Estimated covariance of parameter estimates:

	mu	sigma
mu	4.77267e-06	5.55407e-22
sigma	5.55407e-22	2.38635e-06

Figure 4.20: Personal Valuation sniper results

The second constraint for snipers to remain active is the number of bids. The number of bids are fitted using a Poisson distribution as shown in figures 4.21 and 4.22. In the cumulative probability plot the possible values of 1 to 4 bids in an auction are shown to be slightly underestimated by the fitted distribution.

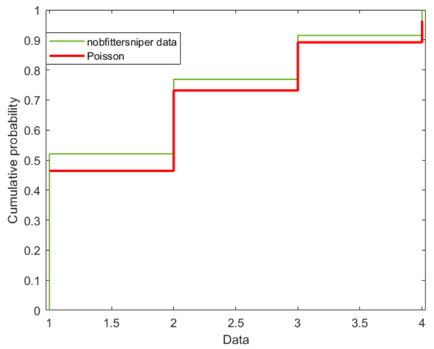
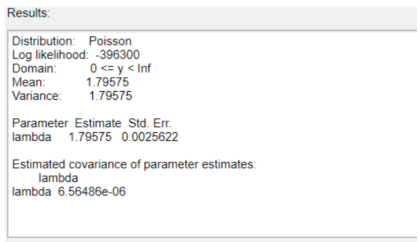


Figure 4.21: Number of bids Sniper



Results:

Distribution: Poisson
Log likelihood: -396300
Domain: 0 <= y < Inf
Mean: 1.79575
Variance: 1.79575

Parameter	Estimate	Std. Err.
lambda	1.79575	0.0025622

Estimated covariance of parameter estimates:

	lambda
lambda	6.56486e-06

Figure 4.22: Number of bids Sniper results

The final constraint for snipers is the time of entry. All time of entry values by snipers are shown in the histogram in figure 4.23. The bars up to 9 are extremely low compared to the most right bar. The shape of the histogram confirms the conclusion made in 4.4.5 that the data does not accurately reflect the behavior of snipers. Therefore, fitting this data to a distribution to model the behavior of snipers doesn't make sense. A manual fitting approach ignoring the lower value ranges resulted in a folded normal distribution with mean 10 and sigma 0.1.

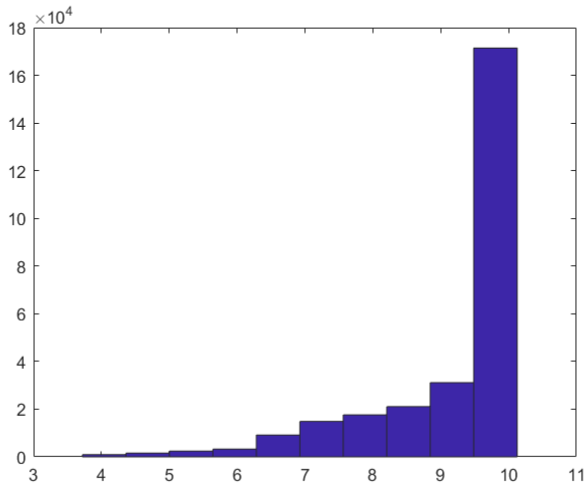


Figure 4.23: Time of Entry Snipers

4.6.3 Starters

The behavior of starters is more fixed than the other strategies. The number of bids is either 0 or 1 since they either start an auction or they don't. Therefore, no distribution fitting is required for the number of bids. Similarly, the personal valuation is exactly the same as the starting price and also doesn't require a distribution fitting. The only constraint that does require a distribution is the time of entry, which is shown in figures 4.24 and 4.25.

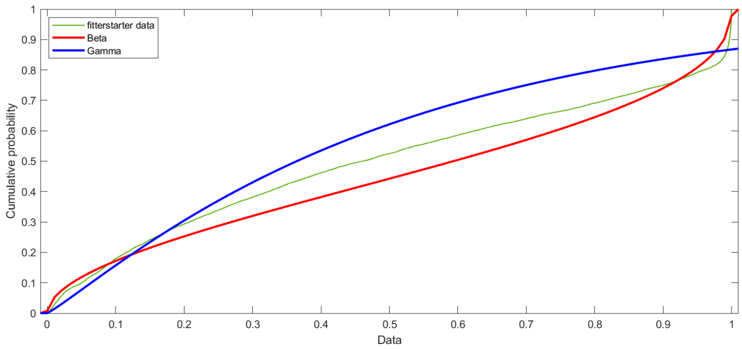


Figure 4.24: Time of entry starter

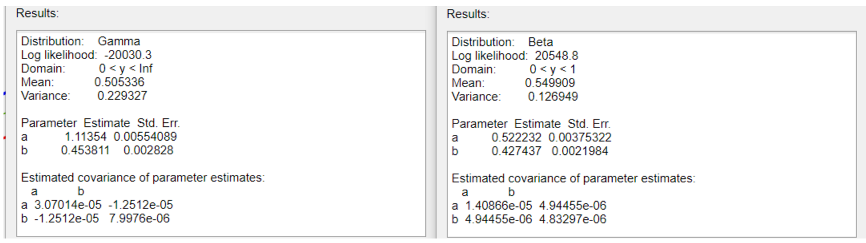


Figure 4.25: Time of entry starter results

In the cumulative probability plot the real data is fitted to two distributions. Both fits are inaccurate for the middle of the data, with the gamma distribution in blue predicts higher probabilities to the first half and then beta distribution predicts lower probabilities to the first half. The beta distribution has a slightly better fit according to log likelihood values. Moreover, the beta fit seem consider the tail end of the data, while the gamma distributions ignores that. Therefore, the behavior of starters with respect to entering an auction is modelled using the beta distribution.

4.6.4 Participators

The behavior of participators is more active than the previous strategies. Their behavior is explained by the personal valuations, the number of bids and the time of entry. The personal valuation of participators was best explained by the Log-Logistic distribution shown in figures 4.26 and 4.27. The cumulative probability plot shows a similar result as with sniper, where the lognormal fits slightly worse during the lower ranges of the data. However, its practical implications to program participator behavior is better and the fit is sufficient. Therefore, the lognormal distribution is used.

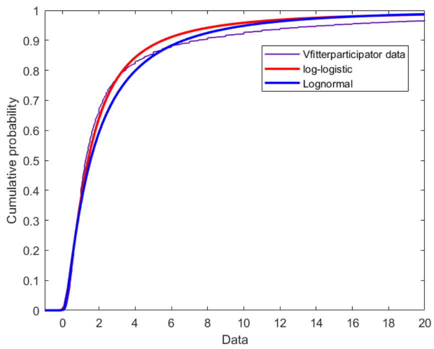


Figure 4.26: Personal Valuation Participator

```
Results:
Distribution:    Log-Logistic
Log likelihood:  -149587
Domain:          0 < y < Inf
Mean:            2.98316
Variance:        Inf

Parameter  Estimate  Std. Err.
mu         0.329972  0.00392754
sigma      0.628125  0.00191665

Estimated covariance of parameter estimates:
       mu           sigma
mu     1.54256e-05  4.3805e-07
sigma  4.3805e-07   3.67357e-06
```

```
Results:
Distribution:    Lognormal
Log likelihood:  -151147
Domain:          -Inf < y < Inf
Mean:            2.96368
Variance:        24.4538

Parameter  Estimate  Std. Err.
mu         0.421029  0.00417833
sigma      1.15361   0.00295456

Estimated covariance of parameter estimates:
       mu           sigma
mu     1.74585e-05  2.19993e-20
sigma  2.19993e-20  8.7294e-06
```

Figure 4.27: Personal Valuation participator results

The second constraint for participators to remain active in the auction is the number of bids. Participators are more active than the previous strategies as can be seen in the figures 4.28 and 4.29. The distribution is fitted using a Poisson distribution. The Poisson distribution slightly underestimates the probability of participators bidding 4 times in an auction and overestimates the probabilities of more than 6 bids.

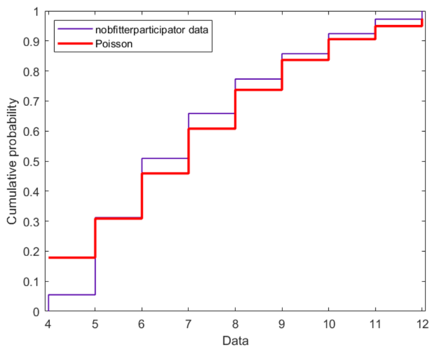


Figure 4.28: Number of bids Participator

```
Results:
Distribution:    Poisson
Log likelihood:  -165619
Domain:          0 <= y < Inf
Mean:            6.93867
Variance:        6.93867

Parameter  Estimate  Std. Err.
lambda     6.93867   0.00954078

Estimated covariance of parameter estimates:
        lambda
lambda  9.10264e-05
```

Figure 4.29: Number of bids participator results

The final constraint for participators is the time of entry. In figure 4.30 the histogram of entry times by participators is shown. The results show a similar shape to what was seen with snipers, but in a wider shape. For this data a distribution fit proved to give bad goodness of fit for every available distribution. Therefore, the time of entry is manually fitting using a folded normal distribution. This resulted in a folded normal distribution with mean 10 and sigma 1.65.

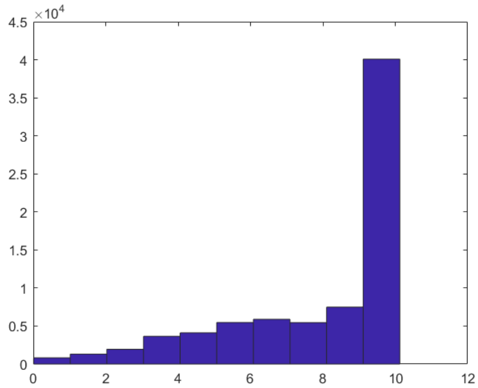


Figure 4.30: Time of entry Participator

4.6.5 Piranhas

The piranha is the most active strategy. These bidders place by far the most bids and do not seem to have a maximum amount. Their bids seem to be limited only by the personal valuation being exceeded. Therefore, the behavior of piranhas is modelled by their personal valuation and the time of entry. Similarly to the other strategies the personal valuation of piranhas is fitted to a log-logistic and a lognormal distribution as shown in figures 4.31 and 4.32. Again the lognormal fit is slightly worse in the lower values compared to the log-logistic.

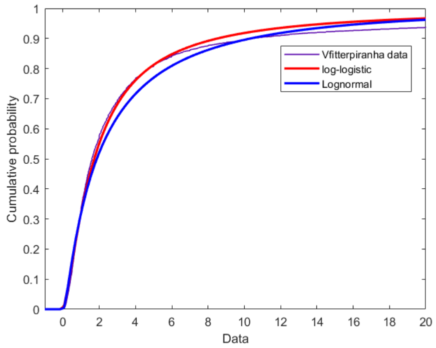


Figure 4.31: Personal Valuation Piranha

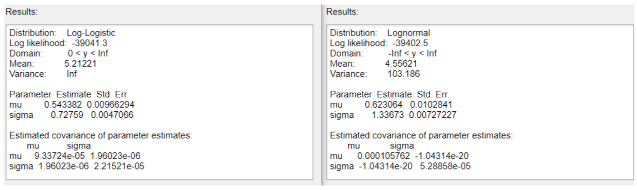


Figure 4.32: Personal Valuation piranha results

The second and last constraint on the behavior of piranhas is the time of entry. Similarly to the starters the time of entry is fitted with a beta and a gamma distribution. In the figures 4.33 and 4.34 the fits are shown. The beta function clearly fits the data better than the gamma as can be seen in the cumulative distribution. Therefore, the beta function is used in the simulation model.

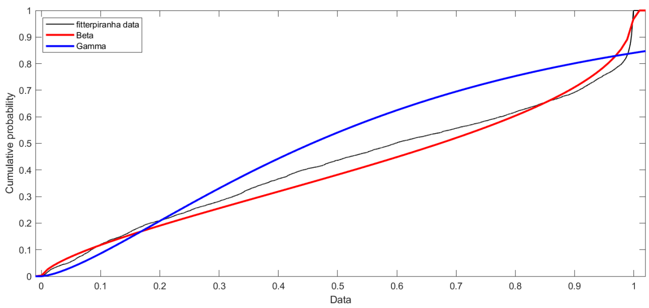


Figure 4.33: Time of entry Piranha

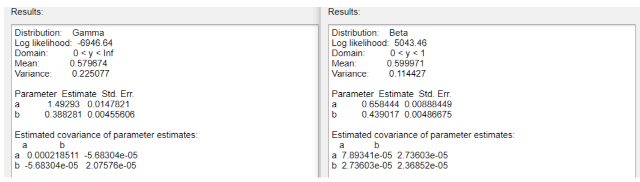


Figure 4.34: Time of entry Piranha results

Chapter 5

Simulation

In the literature review a framework for agent based modeling and simulation was presented 2.4.1. This framework was used to give structure to the simulation model for the online auction process. There are three main elements to agents based modeling: the agents, the relations and the environment. Each of these elements and the way the simulation model is build upon them is reported in the following sections.

5.1 Agents

The agents are the key part to agent based simulation. In the simulation model the agents represent the bidders in the real auction process. In the literature four main characteristics are explained. The agents must meet at least the requirements explained by those characteristics. The first characteristic is that the agent must be easily identified. This requirement is automatically met by Netlogo adding a number next to the type of agent upon creation.

The second requirement is the agent being autonomous. This means agents must be able to act independently in the environment. The bidders are able to place their bids based on the environmental attributes such as the current highest bid, the time, and their specified behavior.

The behavior is part of the third characteristic, the state. The state of an agent is the set of attributes and behaviors, which can vary over time. The way states are modelled is explained for each strategy in the following sections.

The final characteristic is that agents must be social. This means that the behavior of agents can be influenced by interactions with other agents. Agents interact with each other by competing in auctions.

All requirements for agents in an agent based simulation are met. The main part in the modelling is the state of an agent. Their attributes and behavior are modelled based on the fitted distributions as explained in section 4.6. In this section the exact way a bidder in each strategy is modelled is reported. Including the assumptions and justifications for the modelling of the attributes that were either manually fitted, or not fitted at all.

5.1.1 Evaluator agent

For evaluators the time of entry is generated uniformly random in the range from 0 to 9. This doesn't fully reflect the real data as shown in figure 4.18. The way the time of entry is modelled allows for evaluators to enter auctions slightly later than the partitioning results suggest. Translating the programmed range of entry times to the real auction process illustrates that evaluators are able to enter auction up to one or two days before the duration of an auction ends.

To allow evaluators enough time to place their bids and potential bid revisions the activity time is given. This activity time is set at 100 ticks, considering the duration of an auction is 1000 ticks this gives evaluators enough time.

Besides activity time another restriction for evaluator behavior is their personal valuation. The personal valuation is set using the lognormal distribution as fitted in section 4.6.1. The personal valuation can be interpreted as a maximum willingness to pay. When an evaluator places a bid their bid value is either the current bid plus the minimum bid increment or in case it is the first bid of an auction the start price.

The final attribute for evaluators is the maximum number of bids. As shown in section 4.6.1 the data was over fitted to match the tail end of the number of bids. These large number of bid values do not explain the behavior of evaluators correctly. Therefore, the maximum number of bids for evaluators is arbitrarily set to four.

5.1.2 Sniper agent

For snipers the time of entry is a key part of their behavior. Their strategy revolves around bidding very late in an auction. From the data partitioning this behavior was poorly explained as shown in figure 4.23. Which shows entry times by some snipers long before the last day of an auction. Since this does not accurately explain the behavior of real snipers a manual fit of entry times was done focused on the later range of values. To better predict the behavior of snipers entering auctions during the closing stages a manual fit was done. This fit resulted in a folded normal distribution with mean 10 and sigma 0,1.

Similarly, to evaluators sniper are given 100 ticks of auction time to place their bids. Since the entry time of snipers is likely to be late in the auction their activity time will last longer than the duration of an auction. This allows snipers to place bids during the added duration from the soft close format.

The personal valuation of snipers was fitted to a lognormal distribution as reported in section 4.6.2. Similarly to evaluators and other strategies this personal valuation is a maximum value a sniper is willing to bid. When a sniper decides to bid they will increase the current bid by the minimum increment bid, or when they place the first bid the start value.

Finally, the number of bid is constrained with a maximum of 4 bids in an auction. This is the same as evaluators, which is supported by the behavior seen in the literature review 2.3. The maximum of 4 bids in an auction for snipers is supported by the cumulative probability plot in figure 4.21.

5.1.3 Starter agent

The start strategy has the most restricted set of behavioral attributes. Their time of entry is predicted using the inverse cumulative distribution function with the fitted beta values shown in figure 4.24 and 4.25.

Starters are arbitrarily given 100 ticks of time to place their bid. This is a long time considering they only place the first bid of the auction or no bids at all. However, this long active period will not have an impact on the auction process due to the other restrictions.

Since starters only place the first bid of an auction potentially their personal valuation is set to the start price, and the maximum number of bids is set to one.

5.1.4 Participator agent

For participators the time of entry in auctions showed a similar shape as the snipers as reported in section 4.6.4. A manual fitting was done to explain their behavior. This resulted in a folded normal distribution with mean 10 and sigma 1.65. This means participators enter auctions fairly late which is consistent with figure 4.30. The activity time of participators is set to 2000 ticks, which is twice the duration of an auction. Therefore, the activity time is no restriction on the bidding behavior for participators.

There are two restrictions to the activity of participators in auctions. The first restriction is the personal valuation. Which is modelled using the fitted distribution in section 4.6.4. This personal valuation is the maximum willingness to pay. When participators place a bid they will increase the current bid with the minimum bid increment or by placing the first bid with a value of exactly the start price. The other restriction is the number of bids. The maximum number of bids placed by participators in an auction is set to 12 as suggested by figure 4.28.

5.1.5 Piranha agent

Fir the piranhas the time of entry is generated using the inverse cumulative distribution function with the beta parameter values fitted in figures 4.33 and 4.34. Similar to participators the activity time is set to 2000 ticks to ensure the behavior of piranhas is not restricted by time.

Piranhas are the only strategy not restricted by the number of bids which is set to 1000. Since, 1000 bid increments will exceed the personal valuations this is not a restriction. The behavior of piranhas in the real process suggests that they will not give up an auction unless the current bid exceeds their personal valuation. Which is generated using the results in section 4.6.5

5.1.6 Agent arrival

The arrival of agents was fitted in 4.4.4. These distributions are fitted on bidders that actively placed a bid in an auction. In the simulation model agents and may or may not actively place a bid depending on the simulation. As a result using the fitted distributions would lead to a lower number of actively bidding agents than in the real process. A multiplier on the fitted distributions is used to account for the possibility of agents not bidding in the simulation. This multiplier is a parameter used for calibration as explained in 6. Furthermore, for the same reason that bidders may not place a bid in an auction the probabilities of agents adopting certain bidding strategies can not be deduced from the data of actively participating bidders. Therefore, the likelihoods of agents adopting certain strategies are the second parameter.

5.2 Relations

The relations of agents are the second main element in agent based simulation. With the modelling of agents two key requirements must be met. Firstly, the connectedness of agents. The connectedness specifies which agents are connected to each other. In the case of an auction all agents who enter the auction are connected by competing for the same item during the same time.

The second requirement is that the mechanisms of the interacting dynamics need to be specified. These two requirements allow agents to make decisions based on their local information. Which is obtained through those interactions.

In the simulation model bidders directly compete to win the same auction. They interact by over bidding the other agents. Agents adept their behavior when being the highest current bidder by sitting idle until an other agent over bids.

5.3 Environment

The final key element for agent based simulation is the environment. The environment serves as a set of boundaries in which the agents can act. In this simulation model the boundaries of the online auction process are programmed as time, space and resource constraints.

5.3.1 Duration

The duration of an auction is modelled in terms of ticks. Every tick is a potential moment for an agent to place a bid. The duration is limited to 1000 ticks. The soft-close aspect of the online auction is modelled such that when a bid is placed in the final 5 ticks, the duration is extended by 15 ticks. This concludes the time constraints of the simulation.

5.3.2 Bidding

The way agents place their bids is modelled in two separate scenarios. In the first scenario the auction has started but no bids are placed. For the first bid agents in all strategies can only place a bid value equal to the start price. This behavior is confirmed by the data, where almost all auctions start have a first bid equal to the start price. This first bid is placed by the first active agent with a personal valuation of at least the start price. The start prices depend on the estimated value of an item and the pricing strategy. The following table 5.1 shows the starting prices for all estimation values and pricing strategies.

Estimation	**Strategy**	**Start Price**
10	low	1
	med	5
	high	10
100	low	10
	med	50
	high	100
1.000	low	100
	med	500
	high	1.000

Table 5.1: Start price

The second scenario is when at least one bid has been placed in the auction. In this case the agents will increase the highest current bid by exactly the minimum bid increment, if allowed by their personal valuation and maximum number of bids.

As such the minimum bid increments play a major role in the online auction process. The data shows that bidders from all strategies almost exclusively increase the current bid by this minimum increment. These minimum increments depend on the value of the highest current bid as reported in table 5.2.

Current Bid	Bid Increment	Percentage of current bid
$bid < 50$	5	$*50\% - 10\%$
$50 \leq bid < 200$	10	$20\% - 5\%$
$200 \leq bid < 600$	20	$10\% - 5\%$
$600 \leq bid < 2.000$	50	$8,33\% - 2,5\%$
$2.000 \leq bid < 4.000$	100	$5\% - 2,5\%$
$4.000 \leq bid < 10.000$	200	$5\% - 2\%$
$10.000 \leq bid < 15.000$	250	$4\% - 1,67\%$
$15.000 \leq bid < 25.000$	500	$3,33\% - 2\%$
$25.000 \leq bid < 60.000$	1.000	$4\% - 1,67\%$
$60.000 \leq bid < 100.000$	2.000	$3,33\% - 2\%$
bid $\geq$ 100.000	2.500	$2,5\%-$

Table 5.2: Minimum bid increments

As shown in the table the minimum bid increment is relative huge for items in the lower price ranges compared to the more expensive prices ranges. The bid increment has a strong impact in the online auction process as explained in the literature review 2.3.4. These bid increments are much bigger than other online auction platforms such as Ebay. Therefore, the online auction process is likely to be strongly shaped by the environment, specifically the bid increments.

Chapter 6

Validation

In this chapter the simulation model was validated. The validation of a model determines the accuracy of the model with respect to the real system. Validation is achieved through the calibration of a model until an acceptable accuracy is reached. The Calibration process was done iteratively by adjusting two parameters. The first parameter is the probability of an agent behaving according the programmed bidding behavior from each strategy. This parameter was adjusted mainly to calibrate the correct distribution of winning strategies. The second parameter is the multiplier. The arrival process of bidders resulted in fitted distributions on bidders that actively placed bids in auctions. However, the simulation model has an arrival process for agents that may or may not place bids depending on the simulation. Therefore, a multiplier parameter is used to ensure a number of agents to be created such the the number of agents actively bidding in the simulation accurately reflects the real online auction process.

The Calibration process aim to achieve validation of the model. The validation method used is the 95% confidence interval approach. Where the real data was used to generate confidence intervals for the two online auction performance metrics. The first metrics is the result of an auction, which is the final bid price relative to the estimated value of an item. The second performance metric is the number of bidders who placed a bid in an auction.

The results of the validation are reported in table 6.1. In this table the count shows the number of real auctions that occurred with the corresponding estimation value and pricing strategy. Then the confidence intervals and the simulated results for both performance metrics show the accuracy to which the simulation model reflect the real process. All simulations are ran 1000 times for each pricing strategy and estimation combination.

Estimation	**Strategy**	**Count**	95% **C.I. Results**	**Simulation Results**	95% **C.I. bidders**	**Simulation Bidders**
10	H	13599	9.8302 - 10.5894	1.5475	4.2123 - 4.3278	1.6730
	L	170	0.5901 - 1.5005	1.0707	1.4159 - 1.8194	2.2240
	M	194	1.7157 - 2.3843	1.9180	2.3804 - 2.8567	2.6650
100	H	143	2.6283 - 3.2466	1.2459	5.273 - 6.0284	1.7450
	L	9729	1.4727 - 1.6269	1.5548	5.5248 - 5.6597	5.5890
	M	804	1.3802 - 1.6594	1.5061	3.4689 - 3.8272	3.6620
1000	H	14	1.1838 - 2.6733	1.3448	2.6536 - 6.4893	2.5900
	L	1385	1.1479 - 1.3517	1.2623	8.0005 - 8.5064	8.1640
	M	419	1.4199 - 1.7296	1.5061	4.6845 - 5.3393	5.2080

Table 6.1: Confidence Interval Validation

First key result to notice is that the high pricing strategies are not validated. The reasoning for that is that the high pricing strategy with and estimation of 1.000 is ran only 14 times. This is too low of a number to validate a simulation model. Moreover, the high pricing strategy with an estimation of 10 showed an average results of 10,2. This means that real bidders payed on average 100 euros for items that are estimated to be worth 10 euros. This raises questions on the reliability of the data on the high strategy auctions, and the quality of those estimations in those instances. Therefore, the high pricing strategies were disregarded during the validation.

The two strategies of interest by Troostwijk were the low and middle pricing strategies. The focus in the two data gathering periods was on those strategies. The results in table 6.1 validate the simulation model for the low and medium pricing strategies for all estimation values using the 95% confidence interval validation method.

Besides the two performance metrics validated above, the calibration process aimed to accurately reflect the distribution of winning strategies. The validation is illustrated using the following pie charts aimed to accurately reflect the results reported in 4.5 in figures 4.11, 4.12, 4.13. The simulation resulted in the following winning strategies which accurately reflect the real process.

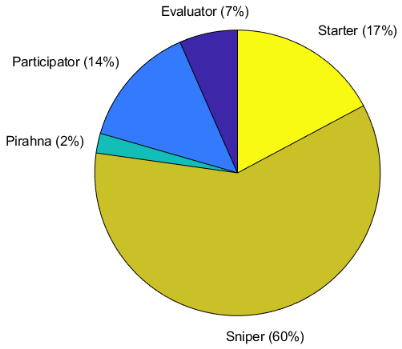


Figure 6.1: Simulation winners 10

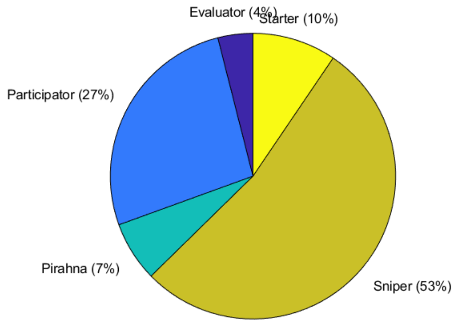


Figure 6.2: Simulation winners 100

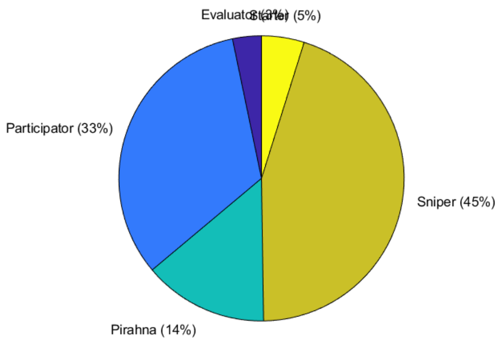


Figure 6.3: Simulation winners 1000

Chapter 7

Discussion

This project aimed to create a simulation model to accurately reflect the process of online auctions. Several pricing strategies were carried out in order to find out how the online auction process is shaped with different starting prices in relation to the estimated value of items being auctioned. The two strategies of interest for Troostwijk were the low strategy starting an auction at roughly 10% of the estimated value, and the medium strategy starting at roughly 50%. In order to structure the project the research questions were presented in section 1.2. The main research question was answered in the literature review 2.4 and the chapter 5 presented the practical steps taken to build the model. The first sub-question was discussed in depth in sections 4.4, 4.6, and 4.5. The second sub-question is discussed below. The third sub-question was discussed in chapter 6.

7.1 Limitations

Just like any other project not everything was perfect. The first limitation of this project is the fact that the bidder valuations are predicted using a distribution that was fitted based on bid values relative to the estimated value of items. These estimations were made by Troostwijk and human error will always be part of such estimates. As a result the simulated personal valuations might be skewed towards human biases in the estimation process. An example of the estimation being wrong was found with the real results using a high starting price strategy on items with an estimation of 10, as reported in table 6.1.

Another limitation is that the arrival process of bidders was founded on the active bidders from the real process. This would cause the simulated auctions to always have too few agents. This was compensated using a multiplication factor. An arrival process based on the number of bidders showing interest might lead to different results and a different simulation model.

Furthermore, the k-means++ algorithm resulted in imperfect cluster partitioning. Some of the behaviors could not be explained with knowledge on the strategies from the literature. Other data mining techniques to partition data might lead to better results with respect to the bidder behavior, and as such to the simulated behavior.

Finally, the high minimum bid increments severely limited the possible outcomes of the simulated auctions. A lower bid increment allows more bidders to participate in auctions and as such increases the potential activity.

7.2 Academic Relevance

This research adds to the literature on online auction simulation by simulating auctions with more possible bidder strategies. Another addition is the soft-close format, where previous simulation models used the more common hard-close format. Additionally, this research found and defined a new bidder strategy previously not discussed in the bidding behavior literature. The starter strategy is not yet discussed in the literature and is sufficiently common to have significant impact on online auctions.

7.3 Future Research

The simulation model was made for the potential to gain knowledge on the online auction process. Some interesting areas to discover are the effects of lower and higher bid increments on the auction process. Another potential area is to stimulate bidders adopting certain strategies. This could be used to manipulate the process of an auction. For example, bidders could be offered an incentive to place the first bid in auctions, or to place multiple bids.

Bibliography

[1] Miriam R Albert. E-buyer beware: Why online auction fraud should be regulated. *American Business Law Journal*, 39(4):575–644, 2002. 14

[2] Dan Ariely and Itamar Simonson. Buying, bidding, playing, or competing? value assessment and decision dynamics in online auctions. *Journal of Consumer psychology*, 13(1-2):113–123, 2003. 7

[3] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006. 14

[4] Robert Axtell and Joshua Epstein. Agent-based modeling: Understanding our creations. *The Bulletin of the Santa Fe Institute*, 9(4):28–32, 1994. 13

[5] Jeff Baker and Jaeki Song. A review of single-item internet auction literature and a model for future research. In *Electronic Government: Concepts, Methodologies, Tools, and Applications*, pages 3987–4012. IGI Global, 2008. 5, 6

[6] Ravi Bapna, Seokjoo Andrew Chang, Paulo Goes, and Alok Gupta. Overlapping online auctions: Empirical characterization of bidder strategies and auction prices. *MIS quarterly*, pages 763–783, 2009. 8, 9

[7] Ravi Bapna, Paulo Goes, Alok Gupta, and Yiwei Jin. User heterogeneity and its impact on electronic auction market design: An empirical exploration. *MIS quarterly*, pages 21–43, 2004. 8, 9

[8] Sylvain Barde and Sander Van Der Hoog. An empirical validation protocol for large-scale agent-based models. 2017. 13

[9] Pavel Berkhin. A survey of clustering data mining techniques. In *Grouping multidimensional data*, pages 25–71. Springer, 2006. 14

[10] Jennifer Brown and John Morgan. Reputation in online auctions: The market for trust. *California Management Review*, 49(1):61–81, 2006. 6

[11] Wen Cao, Qinyang Sha, Zhiyong Yao, Dingwei Gu, and Xiang Shao. Sniping in soft-close online auctions: empirical evidence from overstock. *Marketing Letters*, 30(2):179–191, 2019. 8

[12] Wen-Hsi Chang and Jau-Shien Chang. Using clustering techniques to analyze fraudulent behavior changes in online auctions. In *2010 International Conference on Networking and Information Technology*, pages 34–38. IEEE, 2010. 14

[13] Richard Guy Cox. Optimal reservation prices and superior information in auctions with common-value elements: Evidence from field data. *Ekonomia*, 8(2):142–167, 2005. 7

[14] Xiling Cui and Vincent S Lai. Bidding strategies in online single-unit auctions: Their impact and satisfaction. *Information & Management*, 50(6):314–321, 2013. 8

[15] Utpal M Dholakia, Suman Basuroy, and Kerry Soltysinski. Auction or agent (or both)? a study of moderators of the herding bias in digital auctions. *International Journal of Research in Marketing*, 19(2):115–130, 2002. 7

[16] Robert Entriken and Steve Wan. Agent-based simulation of an automatic mitigation procedure. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pages 92c–92c. IEEE, 2005. 11

[17] Giorgio Fagiolo, Mattia Guerini, Francesco Lamperti, Alessio Moneta, and Andrea Roventini. Validation of agent-based models in economics and finance. In *Computer simulation validation*, pages 763–787. Springer, 2019. 11, 13

[18] Giorgio Fagiolo, Alessio Moneta, and Paul Windrum. A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems. *Computational Economics*, 30(3):195–226, 2007. 12

[19] Paulo B Goes, Gilbert G Karuga, and Arvind K Tripathi. Bidding behavior evolution in sequential auctions: Characterization and analysis. *Mis Quarterly*, pages 1021–1042, 2012. 8, 9

[20] Alok Gupta and R Bapna. Online auctions: A closer look, 2002. 6

[21] Greg Hamerly and Charles Elkan. Learning the k in k-means. *Advances in neural information processing systems*, 16:281–288, 2004. 15

[22] Ernan Haruvy and Peter TL Popkowski Leszczyc. The impact of online auction duration. *Decision Analysis*, 7(1):99–106, 2010. iii, 1, 7

[23] Jingsi Huang and Jie Song. Optimal inventory control with sequential online auction in agriculture supply chain: An agent-based simulation optimisation approach. *International Journal of Production Research*, 56(6):2322–2338, 2018. 11

[24] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999. 14

[25] Wolfgang Jank and Galit Shmueli. Dynamic profiling of online auctions using curve clustering. *submitted for publication*, 2003. 14

[26] Wolfgang Jank and Galit Shmueli. Profiling price dynamics in online auctions using curve clustering. *Robert H. Smith School Research Paper No. RHS-06-004*, 2005. 14

[27] Rama Katkar and David Lucking-Reiley. Public versus secret reserve prices in ebay auctions: results from a pokémon field experiment. Technical report, National Bureau of Economic Research, 2001. 6

[28] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009. 15, 16

[29] Preetinder Kaur, Madhu Goyal, and Jie Lu. Pricing analysis in online auctions using clustering and regression tree approach. In *International Workshop on Agents and Data Mining Interaction*, pages 248–257. Springer, 2011. 14

[30] Gillian Ku, Adam D Galinsky, and J Keith Murnighan. Starting low but ending high: A reversal of the anchoring effect in auctions. *Journal of Personality and social Psychology*, 90(6):975, 2006. 6

[31] Averill M Law, W David Kelton, and W David Kelton. *Simulation modeling and analysis*, volume 3. McGraw-Hill New York, 2000. 12

[32] Anson Lee, Michael Lyu, and Irwin King. An agent-based platform for online auctions. In *Proceedings of the International Conference on Internet Computing, Las Vegas*. Citeseer, 2001. 11

[33] Bin Lei, Huichao Zhang, Huiyu Chen, Lili Liu, and Dingwei Wang. A k-means clustering based algorithm for shill bidding recognition in online auction. In *2012 24th Chinese Control and Decision Conference (CCDC)*, pages 939–943. IEEE, 2012. 14

[34] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003. 14

[35] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982. 14

[36] David Lucking-Reiley. Auctions on the internet: What's being auctioned, and how? *The journal of industrial economics*, 48(3):227–252, 2000. 5

[37] C M Macal and M J North. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3):151–162, 2010. 10

[38] Charles Macal and Michael North. Introductory tutorial: Agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference 2014*, pages 6–20. IEEE, 2014. 10

[39] R Preston McAfee and John McMillan. Auctions and bidding. *Journal of economic literature*, 25(2):699–738, 1987. 5

[40] Paul R Milgrom and Robert J Weber. A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society*, pages 1089–1122, 1982. 5

[41] Hideyuki Mizuta and Ken Steiglitz. Agent-based simulation of dynamic online auctions. In *2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165)*, volume 2, pages 1772–1777. IEEE, 2000. 11

[42] Kornelije Rabuzin and Nikola Bakos. Agent-based simulation model of online auctions in netlogo. In *Central European Conference on Information and Intelligent Systems*, page 75. Faculty of Organization and Informatics Varazdin, 2010. 11

[43] Michael H Rothkopf. A model of rational competitive bidding. *Management Science*, 15(7):362–373, 1969. 5

[44] Jarrod Trevathan. Detecting shill bidding in online english auctions. In *Handbook of research on social and organizational liabilities in information security*, pages 446–470. IGI Global, 2009. 14

[45] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961. 5

[46] Matthew JC Walley and David R Fortin. Behavioral outcomes from online auctions: reserve price, reserve disclosure, and initial bidding influences in the decision process. *Journal of Business Research*, 58(10):1409–1418, 2005. 6

[47] Wenli Wang, Zoltán Hidvégi, and Andrew B Whinston. Shill bidding in english auctions. *Emory University and the University of Texas at Austin, mimeo*, 2001. 14

[48] Wenli Wang, Zoltan Hidvegi, and Andrew B Whinston. Shill bidding in multi-round online auctions. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pages 7–pp. IEEE, 2002. 14

[49] Robert B Wilson. Communications to the editor—competitive bidding with disparate information. *Management science*, 15(7):446–452, 1969. 5

[50] Rüdiger Wirth and Jochen Hipp. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1. Springer-Verlag London, UK, 2000. iv, vii, 3

[51] Xiaorong Xiang, Ryan Kennedy, Gregory Madey, and Steve Cabaniss. Verification and validation of agent-based scientific simulation models. In *Agent-directed simulation conference*, volume 47, page 55, 2005. 12

[52] Hai Yu, Shouyang Wang, and Chuangyin Dang. Optimal starting price for ebay-like online auctions. *Journal of Systems Science and Complexity*, 19(1):9–21, 2006. 6

[53] Jie Zhang. The roles of players and reputation: evidence from ebay online auctions. *Decision Support Systems*, 42(3):1800–1818, 2006. 6

[54] NP Ziogos and AC Tellidou. An agent-based ftr auction simulator. *Electric power systems research*, 81(7):1239–1246, 2011. 11

Appendix A

Netlogo code

```
;;; creating constructs for auctions
breed [sellers seller]
breed [objects object]
breed [bidders bidder]

;;; assign construct owned variables
sellers-own [PriceStrategy Performance]
objects-own [StartingPrice EstimatedValue]
bidders-own [PersonalValuation Bid Strategy EntryTime ActiveTime behavior NOB MAXNOB]

globals [Time CurrentBid NOBidders a b StartPrice TotalBids FinalBid MarginalBid Leader xx yy
LLaplha LLbeta BidTiming ThisDuration Leaders winningstrategy Result finalbidders Bounda Bound



to setup
  clear-all
  reset-ticks
  set CurrentBid 0
  set TotalBids 0
  set ThisDuration Duration
  set leaders []
  create-sellers 1 [set shape "house" setxy 10 10 set color white set size 5]



;;; create objects with estimated value, marginal bid, and starting price according to the set
  create-objects 1 [set shape "car" setxy 10 5 set color white set size 2 set heading 90
    set EstimatedValue Estimation

    if PricingStrategy = "low" [set StartingPrice 0.1 * Estimation]
    if PricingStrategy = "medium" [set StartingPrice 0.5 * Estimation]
    if PricingStrategy = "high" [set StartingPrice Estimation]
;;; Marginal bids increase incrementally with respect to the current bid or starting price
;;; Here the marginal bid is initially set with respect to the starting price
   (ifelse
      StartingPrice < 50 [set MarginalBid 5]
      50 <= StartingPrice and StartingPrice < 200 [set MarginalBid 10]
      200 <= StartingPrice and StartingPrice < 600 [set MarginalBid 20]
      600 <= StartingPrice and StartingPrice < 2000 [set MarginalBid 50]
      2000 <= StartingPrice and StartingPrice < 4000 [set MarginalBid 100]
      4000 <= StartingPrice and StartingPrice < 10000 [set MarginalBid 200]
      10000 <= StartingPrice and StartingPrice < 15000 [set MarginalBid 250]
      15000 <= StartingPrice and StartingPrice < 25000 [set MarginalBid 500]
      25000 <= StartingPrice and StartingPrice < 60000 [set MarginalBid 1000]
      60000 <= StartingPrice and StartingPrice < 100000 [set MarginalBid 2000]
      StartingPrice > 100000 [set MarginalBid 2500]

     )

  ]
```

Figure A.1: Netlogo Code Part 1

```
  ;;; set PoissonMean according to the pricing strategy and estimated value
  ;;; also set startprice
  if PricingStrategy = "low" [set startprice 0.1 * Estimation

    if Estimation = 10 [set a 2.5961 set b 1.7055 set Multiplier 1]
    if Estimation = 100 [set a 2.6728 set b 2.3616 set Multiplier 1.7]
    if Estimation = 1000 [set a 2.6094 set b 3.7941 set Multiplier 1.55]
    if Estimation = 10000 [set a 2.7082 set b 5.5863 set Multiplier 1]]

  if PricingStrategy = "medium" [set startprice 0.5 * Estimation

    if Estimation = 10 [set a 2.3543 set b 1.5324 set Multiplier 2.2]
    if Estimation = 100 [set a 2.1545 set b 1.9102 set Multiplier 2.1]
    if Estimation = 1000 [set a 2.0888 set b 2.7589 set Multiplier 2.2]
    if Estimation = 10000 [set a 2.4475 set b 3.1831 set Multiplier 1]]

  if PricingStrategy = "high" [set startprice Estimation

    if Estimation = 10 [set a 1.7802 set b 2.3333 set Multiplier 1]
    if Estimation = 100 [set a 1.7635 set b 1.9264 set Multiplier 1]
    if Estimation = 1000 [set a 2.6648 set b 2.2896 set Multiplier 1]
    if Estimation = 10000 [set a 1.5569 set b 1.9853 set Multiplier 1]]

  ;;; create bidders according to a gamma arrival distribution with a mean derived above
  set NOBidders round(Multiplier * random-gamma a  (1 / b))
  if NOBidders = 0 [set NOBidders 1]


;;; here we model the behavior of all bidders as well as color code them for visuals
  ;;; set the boundary values for strategies based on estimation
if estimation = 10 [
    set BoundA 27
    set BoundB 42
    set BoundC 43
    set BoundD 50
  ]

if estimation = 100 [
    set BoundA 21
    set BoundB 40
    set BoundC 42
    set BoundD 52
  ]

if estimation = 1000 [
  set BoundA 22
  set BoundB 41
  set BoundC 44
  set BoundD 56
  ]
```

Figure A.2: Netlogo Code Part 2

```
create-bidders NOBidders [set shape "person" set size 2 set NOB 0
let choice random 100
  (ifelse
    ;;; Starter Strategy Yellow/ Arrival using beta distribution with a= 0.522232 and b = 0.
    choice <= BoundA [set Strategy "Starter" set color yellow
      set xx random-gamma 0.522232 1
      set yy random-gamma 0.427437 1
      set EntryTime round(xx / (xx + yy) * Duration)
      set ActiveTime 100
      set PersonalValuation StartPrice ;;;currently set PV equal to the global variable Star
      set MAXNOB 1

    ]
    ;;; Evaluator Strategy Green
    BoundA < choice and choice <= BoundB [set Strategy "Evaluator" set color green

      set EntryTime random (9 * Duration / 10)
      set ActiveTime 100
      ;;;Choose whether to use log-logistic or lognormal distribution for personal valuation
      ;;;set PersonalValuation round(estimation * random-LogLogistic -0.892084 0.696019)
      set PersonalValuation round(estimation * random-lognormal -0.859458 1.271)
      set maxNOB 4
    ]
    ;;; Pirahna Strategy Blue / Arrival using beta distribution with a = 0.658444 and b = 0.
    BoundB < choice and choice <= BoundC [set Strategy "Pirahna" set color blue
      set xx random-gamma 0.658444 1
      set yy random-gamma 0.439017 1

      set EntryTime round((xx / (xx + yy) * Duration))
      set ActiveTime 2000
      ;;;choose loglog or lognormal
      ;;;set PersonalValuation round(estimation * random-LogLogistic 0.543382 0.72759)
      set PersonalValuation round(estimation * random-lognormal 0.623064 1.33673)
      set MAXNOB 1000
    ]
    ;;; Participator Strategy Red/ Arrival using Folded Normal Distribution with mean 10 and
    BoundC < choice and choice <= BoundD [set Strategy "Participator" set color red
      set NormalParticipator random-normal 10 1.65
      if NormalParticipator <= 10 [
        set EntryTime round(NormalParticipator * Duration / 10)]
      if NormalParticipator > 10 [
        set EntryTime round((10 - (NormalParticipator - 10)) * Duration / 10)]
      set ActiveTime 2000
      ;;;Choose loglog or lognormal
      ;;;set PersonalValuation round(estimation * random-LogLogistic 0.329972 0.628125)
      set PersonalValuation round(estimation * random-lognormal 0.421029 1.15361)
      set MAXNOB 12

    ]
```

Figure A.3: Netlogo Code Part 3

```
    ;;; Sniper Strategy Pink/ Arrival Time using folded normal distribution with mean 10 and
    choice > BoundD [set Strategy "Sniper" set color pink
      set NormalSniper random-normal 10 0.1
      if NormalSniper < 10[
        set EntryTime round(NormalSniper * Duration / 10)]
      if NormalSniper >= 10 [
        set EntryTime round((10 - (NormalSniper - 10)) * Duration / 10)]
      set ActiveTime 100
      ;;;choose loglog or lognormal
      ;;;set PersonalValuation round(estimation * random-LogLogistic -0.139956 0.622904)
      set PersonalValuation round(estimation * random-lognormal -0.0906297 1.14259)
      set maxNOB 4
    ]
    )

  output-type strategy
  output-type " PV: "
  output-type personalvaluation
  output-type ", TOE: "
  output-type EntryTime
  output-type ",  Active: "
  output-print ActiveTime


 ]

 ;;; give all bidders an own spot
 let am 2
 let side sqrt NOBidders
 let step 2.2
 let x 0
 let y 0
 while [am < NOBidders + 2]
           [ if x > (side - 1) * step
             [ set y y - step
               set x 0]
             ask bidder am [setxy x y]
             set x x + step
             set am am + 1]

end ;;; end of setup
```

Figure A.4: Netlogo Code Part 4

```
to go
  tick
  if ticks <= thisduration [
if totalbids = 0 [
      place-firstbid
    ]
if totalbids > 0 [
  place-bid
    ]
  update
  ]
  if ticks >= thisduration [
      if leader = 0 [
      setup
      go
    ]
    if leader != 0 [
  terminate-auction
  ]
  stop
  ]


end
```

Figure A.5: Netlogo Code Part 5

```
;;; Model how the first bid is simulated
to place-firstbid


  let activebidder one-of bidders with [EntryTime < ticks and ticks < Entrytime + ActiveTime
    and PersonalValuation >= StartPrice]



  if activebidder != NOBODY [
  ask activebidder [
    if Strategy = "Starter"
      [
        set NOB 1
        set TotalBids TotalBids + 1
        set Bid StartPrice
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
    ]

    if Strategy = "Evaluator"
      [
        Set NOB NOB + 1
        set TotalBids TotalBids + 1
        set Bid StartPrice
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
      ]

    if Strategy = "Participator"
      [
        set NOB NOB + 1
        set TotalBids TotalBids + 1

        set Bid StartPrice
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
      ]

    if Strategy = "Pirahna"
      [
        set NOB NOB + 1
        set TotalBids TotalBids + 1
        set Bid StartPrice
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
      ]
```

Figure A.6: Netlogo Code Part 6

```
    if Strategy = "Sniper"
      [
        set NOB NOB + 1
        set TotalBids TotalBids + 1
        set Bid StartPrice
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
      ]
    ]
  ]

end

;;; Model how the bids are simulated
to place-bid

  ask leader [ht]

  let activebidder one-of bidders with [EntryTime < ticks and ticks < Entrytime + ActiveTime
    and PersonalValuation >= StartPrice and PersonalValuation >= Currentbid + marginalBid and

  ask leader [st]


  if activebidder != NOBODY and ActiveBidder != leader [
  ask activebidder [


    if Strategy = "Evaluator"
      [
        Set NOB NOB + 1
        set TotalBids TotalBids + 1
        set Bid CurrentBid + MarginalBid
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
      ]

    if Strategy = "Participator"
      [
        set NOB NOB + 1
        set TotalBids TotalBids + 1
        Set Bid CurrentBid + MarginalBid
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
      ]
```

Figure A.7: Netlogo Code Part 7

```
    if Strategy = "Pirahna"
      [
        set NOB NOB + 1
        set TotalBids TotalBids + 1
        set Bid CurrentBid + MarginalBid
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
      ]

    if Strategy = "Sniper"
      [
        set NOB NOB + 1
        set TotalBids TotalBids + 1
        set Bid CurrentBid + MarginalBid
        set label Bid
        set Leader activebidder
        set BidTiming  ticks
      ]
    ]
  ]

end


to update

  set CurrentBid max[Bid] of bidders
  ask sellers[
  set Performance CurrentBid / Estimation
;;; update marginal bid according to the latest bid
  (ifelse
      CurrentBid < 50 [set MarginalBid 5]
      50 <= CurrentBid and CurrentBid < 200 [set MarginalBid 10]
      200 <= CurrentBid and CurrentBid < 600 [set MarginalBid 20]
      600 <= CurrentBid and CurrentBid < 2000 [set MarginalBid 50]
      2000 <= CurrentBid and CurrentBid < 4000 [set MarginalBid 100]
      4000 <= CurrentBid and CurrentBid < 10000 [set MarginalBid 200]
      10000 <= CurrentBid and CurrentBid < 15000 [set MarginalBid 250]
      15000 <= CurrentBid and CurrentBid < 25000 [set MarginalBid 500]
      25000 <= CurrentBid and CurrentBid < 60000 [set MarginalBid 1000]
      60000 <= CurrentBid and CurrentBid < 100000 [set MarginalBid 2000]
      CurrentBid > 100000 [set MarginalBid 2500]

   )

  ]
  ;;; add soft close update
  if BidTiming >= ThisDuration - 5
  [
    set ThisDuration ThisDuration + 15
  ]
  ;;; create a list of leaders (to count the number of bidders who placed a bid)
  if not member? leader leaders
  [
    set leaders lput leader leaders
  ]

end
```


Figure A.8: Netlogo Code Part 8

```
to terminate-auction
  if member? 0 leaders
  [
    set leaders remove-item 0 leaders
  ]
  ;;; set perforamnce values
  set finalbid currentbid
  ask leader [set winningstrategy strategy]
  ask sellers [set Result performance]
  set finalbidders length leaders
  ;;; report on the auction results
  output-write "The winner of the auction is: "
  ask leader [output-show strategy]
  output-write "with a personal valuation of: "
  ask leader [output-show PersonalValuation]
  output-write "and a winning bid of: "
  ask leader [output-show bid]
  output-write "the auction completed succesfully with a performance of: "
  ask sellers [output-show performance]
  output-write "The auction had "
  output-write length Leaders
  output-write " bidders placing a total of "
  output-write totalbids
  output-write " bids"
  ask bidders [output-show nob]
end

to-report random-LogLogistic [mu sigma]
  let r random-float 1
  let q (r / (1 - r))
  let x (exp mu) * (q ^ sigma)
  report x


end


to-report random-lognormal [mu sigma]
  let x exp( mu + sigma * (random-normal 0 1))
  report x
end
```

Figure A.9: Netlogo Code Part 9