

# Censored Data Analysis of Tea Auction Pricing

Subhrajyoty Roy  
Roll: MB1911



A report presented in partial satisfaction of the course “Survival Analysis”  
as a part of “Master of Statistics (M. Stat.) Programme”

Indian Statistical Institute, Kolkata  
June 12, 2021

# 1 Introduction

Tea (*Camellia sinensis*) is a manufactured drink that is consumed across the world. Since the tea crop has specific tropical agro-climatic requirements, it is mostly produced in China, India, Sri Lanka, Vietnam etc., the Asian countries. The tea industry is one of the oldest organized industries in India with a large network of tea producers, retailers, distributors, auctioneers, exporters and packers. It provides a vital role in the Indian economy by being one of the major foreign exchange earners [1].

Tea is heterogeneous with a wide range of variety being produced throughout India. Almost 88% of the tea produced in India are of CTC (Cut, Tear and Curl) type (the black tea leaf), with about 10% being of an orthodox variety, and rest 2% are green tea [2]. CTC tea is largely graded as Broken Leaf, Dust and Fannings. Each of these grades has a dozen sub-grades based on the size of the grain etc. Several factors influence the demand for tea, including the price and income variables, demographics, cultural background. Apart from consumption, other main drivers of international tea prices are trends and changes in per capita consumption, market access, the potential effects of pests and diseases on production, and changing dynamics between retailers, wholesalers and multinationals. To regulate this volatility in tea pricing, it is generally sold to the retailers through some regulatory e-Auction houses, which acts as an intermediary between the tea producers and the retailers.

## 1.1 e-Auction and its relation to censoring

Starting from the tea leaves produced in the tea gardens of Assam or Darjeeling (the major tea producers in India) to a beverage lying in our cup, there are many steps that it must go through. The tea leaves handpicked by the labours working in the garden is first moved to the tea factory associated with the garden, where the leaves are cut, torn and dried for several weeks, and then curled into tea dust which is packaged. Then, these packages are sent to auction houses, where these packages are opened and samples are collected. These tea samples are brewed and several experienced tea tasters taste them, and assigns some quality ranking, which is then transferred into a basic valuation for the corresponding packet. These samples are sent to some of the potential bidders as well. Finally, on the day of the auction, the packages are again sealed and are put for the auction where the retailers can bid [3]. These retailers, upon winning the auction bids, procure these tea packets from the auction houses and sells them in the market to consumers like us.

The censoring mechanism in the bidding comes from the peculiar setup of the second price closed bidding auction system. In the auction, the auction house poses an initial valuation of the tea packets based on the subjective judgement of its quality obtained from the appointed tea tasters. Since, tea is a perishable good, to increase the sell-ability of the packet, bidders are allowed to bid less than the initial valuation. As long as the final bid remains less than the initial valuation, the highest bidder needs to pay equal to his bid amount to the auction house. However, if the final bid exceeds the initial valuation, the highest bidder only needs to pay the amount equal to the second-highest bid. Therefore, the price which is recorded by the auction houses is either an uncensored observation of buyers' own valuation, or a censored observation of the buyer's own valuation censored by the second-highest bid. Let,  $V_1, V_2, \dots, V_n$  be the initial valuations for  $n$  packets declared by the auction house, and let  $P_1, \dots, P_n$  be the selling price, then the observations are

$(P_i, \delta_i)$ , where

$$\delta_i = \begin{cases} 1 & \text{if } V_i \geq P_i; \\ 0 & \text{if } V_i < P_i; \end{cases} \quad \text{for } i = 1, 2, \dots, n, \quad (1)$$

denoting whether the underlying variable (i.e. the own valuation of the bidders, which is an indicator of the ultimate selling price to the consumers) is uncensored or not.

## 1.2 Existing Works

Classical analysis of auction theory viewed auction as a game between the buyers and sellers where some incomplete information is provided through different bidding structures. Vijay krishna [4] has provided detailed mathematical treatment of various auction mechanisms such as sealed envelope auction, first and second-price auction, European, English and Dutch auction etc from a game-theoretic point of view where the payoff functions are explicitly known. Compared to this, viewing the selling price as a censored data in the context of auction theory is fairly new in practice. In many real-life scenarios, the payoff function is not known apriori, hence one must estimate these payoffs based on the censored bidding history available. George and Hui [5], provide an ingenious way to estimate demand in the auction market, under the Independent Private value model second-price auctions. The Private value model assumes that each of the signals about the actual price of the item to each of the bidders is independent. However, in most of the cases, the bidders are well-informed, hence their respective signals have a dependency, hence an independent private value model assumption is not justified.

Other than demand estimation, censored data on bidding history can prove useful in analyzing optimal bid for the bidders as well. In the online advertising industry, different companies bid for screen spaces of websites to show their advertisements, and a prediction about the winning price in the bid helps the bidders to secure screen spaces optimally [6].

## 1.3 Objective of the study

In the context of the tea auction of India, analysis shows that the auctioneer's valuation, which should have been only a measure of the quality of the item, becomes a causal variable for the demand signal of the bidders [7]. Such an indication of causality and influence of the base price is also available in the context of tea auction in Kenya, Mombasa etc. [8], so the authorities have decided to hold offline auctions (where the bidders get to test the tea packets themselves) rather than "e-Auction" held in online mode. One of the primary objectives of the study is to establish whether auction house provided valuation plays a significant role in the determination of the valuation of the bidders in the market.

Also, since the auction house provided valuations are subjective and depend on the preferences and experiences of tea tasters, several bidders find it uninteresting to stay in the industry. Similar problems were also apparent in Coal block auctions till 2015 [9], which led the authorities to find a suitable substitute index number. This serves as a representative of its quality and well correlated with the bidder's proposed valuation, but remains free of any kind of subjective evaluation. The secondary goal is to provide a way to obtain such an index of quality for different types of tea.

## 2 Description of the dataset

We use the data<sup>1</sup> on e-Auction statistics of Kolkata dust tea provided by J. Thomas & Co. Pvt. Ltd, which is the largest auctioneer in the world, handling over 200 million kg of tea a year, about one-third of all tea auctioned in India [7]. The dataset contains e-Auction statistics from 2018 January to 2019 March. The bidding happens once a week, hence information about the packages displayed for bidding for that week is available, whether that package was sold or not. Apart from identification information on the tea package, the origin of the tea (the tea garden where it came from), its grade of quality [10], the net weight of each tea packets and the quantity, as well as the initial valuation provided by the auction house and ultimate sold price. Note that, no information about the bidding history is recorded by the auction houses since these auctions generally happen online via the internet with open access and there could be potentially a large number of bidders available. In the dataset, we have 25 grades of tea available with an enormous 238 tea gardens (293 including their Clonal, Royal, Gold and Special variants). A detailed description of these can be found in Dalal et al. [7]. The presence of these many factors of a categorical variable (source and grade) complicates any further analysis, hence Dalal et al. [7] perform EM-GMM clustering on these two categorical variables. Based on the outcome of the clustering method and subjective knowledge on the demographic and geographic conditions of the tea-producing districts of Assam, they provided 6 clusters for tea grade and 7 clusters for tea gardens. We shall refer to these cluster variables as “Gradecluster” and “Sourcecluster” respectively.

## 3 Methodologies and Data Analysis

Let  $(P_i, \delta_i)$  for  $i = 1, 2, \dots, n$  be the censored data that we observe as described in Eq. (1). Also, for each package  $i$ , we observe the valuation  $V_i$  provided by the auction house, the grade quality  $G_i$ , the origin or source  $O_i$  and the volume of the tea package  $Q_i$ . Also, let us assume  $B_i$  be the unobserved highest possible bid that a bidder is willing to bid for the  $i$ -th tea packet. Since according to auction theory, a reasonable bidder exactly bids his or her own valuation about the item [11],  $B_i$ ’s can be interpreted as the bidder’s own private valuation of the item.

To achieve the goal mentioned in Section 1.3, we shall use theories of survival analysis of the censored data setup given in Eq. (1). Although these censored variables do not necessarily denote time to some event, they can be analyzed through the mechanism of survival analysis, since the underlying variable  $B_i$  is a non-negative random variable. Also, the analysis by Dalal et. al. [7] showed that the distributions of these valuations generally follow a log-normal distribution or mixture of log-normal random variables, which is common in “time-to-event” variables. A precise description of this setup in the context of survival analysis is as follows:

1. Let  $B_{i2}$  be the second-highest bid for the  $i$ -th item.
2. The censoring random variable is  $c_i = \max\{V_i, B_{i2}\}$ .
3. The observable price  $P_i = \min\{c_i, B_i\}$ , and  $\delta_i$  is the censoring indicator which takes value 1 if the observation is uncensored.

---

<sup>1</sup>The data is not publicly available. Hence, is not used before except this work [7].

4. The survival function of  $B_i$  i.e.  $S_i(t)$  for  $t \in [0, \infty)$  denotes the probability of a bidders dropping out if the current bidding reaches  $t$ . We shall call it as “bid raising probability” at current bid price  $t$ .
5.  $B_i$  has a hazard rate  $\lambda_i(t)$  for  $t \in [0, \infty)$  which denotes the willingness of a bidder to bid a little more than  $t$ , given that the current highest bid is  $t$ .
6. Under the assumption that the auction houses does not cause bidders to change their own private valuation about the tea packet, and each valuation is simply an independent guess of the true value of the item, it follows that the above censoring scheme is random censoring.

In order to understand the relationship between the selling price  $P_i$  and the valuation  $V_i$ , we find that they has a Pearson’s product moment correlation coefficient equal to 0.962 and Spearman’s rank correlation as 0.958, which are extremely high. Such high correlation is also prevalent in different types of tea as clustered through the GradeCluster and SourceCluster variables; see Figure 1 for illustration.

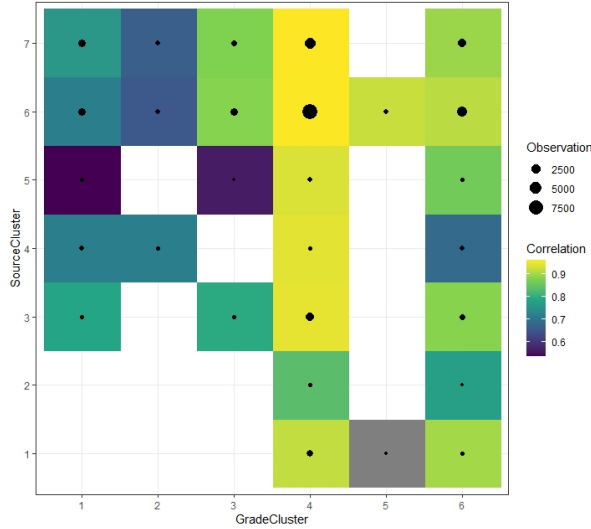


Figure 1: The correlation between Price and Valuation for different types of tea of varying grades and sources, with dots denoting the number of available observations.

### 3.1 Estimation of bid raising probability

Since, each of the types of tea could have a different dropping patterns among bidders, we could use the usual Kaplan-Meier Product Moment Limit (KMPL) estimator [12] to nonparametrically estimate the survival functions  $S_i(t)$ . Assuming we have  $P_{(1)} < P_{(2)} < \dots < P_{(k)}$  distinct selling prices without censoring, and at time  $P_{(i)}$ ,  $d_i$  bidders drops out of the remaining  $n_i$  bidders, the KMPL estimate turns out to be given by the formula

$$\hat{S}(t) = \prod_{i: P_{(i)} < t} \left(1 - \frac{d_i}{n_i}\right), \quad \forall t \geq 0.$$

Applying complementary log-log transformation to  $S(t)$  and using delta method, one can show that an asymptotic  $100(1 - \alpha)\%$  confidence interval of  $S(t)$  based on the KMPL estimate  $\hat{S}(t)$  can be obtained as

$$\left( \exp \left[ -e^{\hat{L}(t) + z_{1-\alpha/2} \hat{\sigma}(t)} \right], \exp \left[ -e^{\hat{L}(t) - z_{1-\alpha/2} \hat{\sigma}(t)} \right] \right)$$

where

$$\begin{aligned} \hat{L}(t) &= \log \left( -\log(\hat{S}(t)) \right) \\ \hat{\sigma}(t) &= \frac{1}{\log(\hat{S}(t))^2} \sum_{i: P_{(i)} < t} \frac{d_i}{n_i(n_i - d_i)} \end{aligned}$$

and  $z_{1-\alpha/2}$  is the  $(1 - \alpha/2)$ -th quantile of the standard normal distribution. The KMPL estimate along with 95% confidence intervals for the particular data has been computed and shown in Figure 2. The effect of the sources on different gradeclusters are very different, for example, for tea type of grade cluster 3, the cluster of the source or origin of the tea does not affect the bidder's behavior much, while the behavior for grade cluster 4 has dependence with different origin or sources.

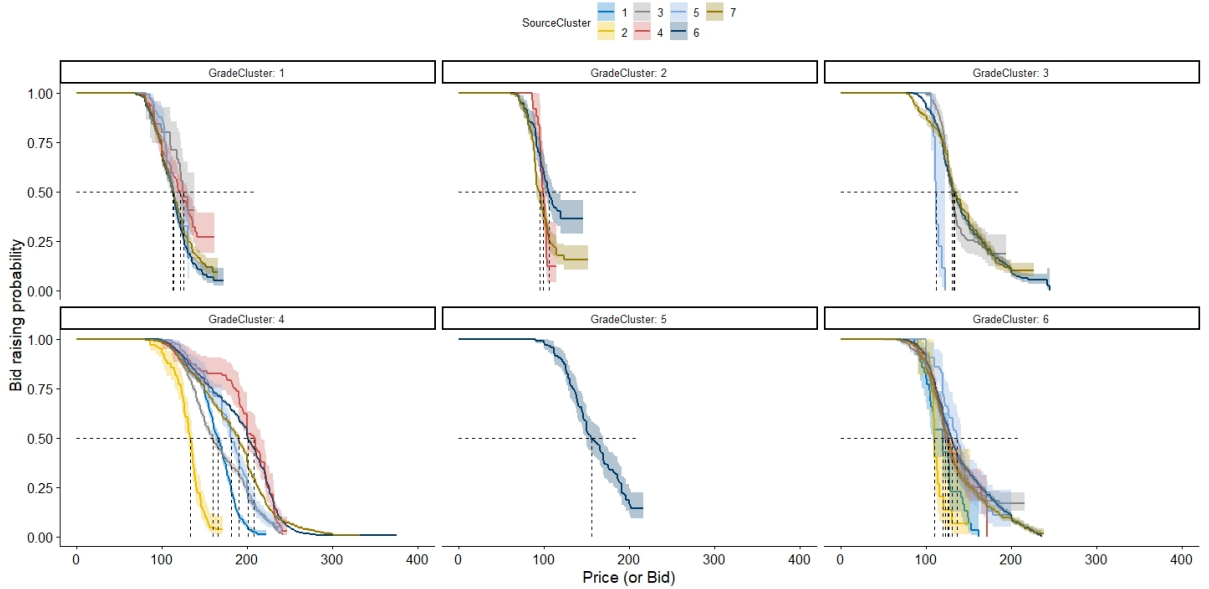


Figure 2: KMPL estimate of Bid raising probability as a function of current price (or bid) for all source and grade clusters

### 3.2 Estimating objective valuation index

Our secondary goal of obtaining a valuation index, which is representative of the quality of the tea packet can now be obtained from the estimated median of these survival curves. The median of the survival curve  $S(t)$  is estimated as the median of  $\hat{S}(t)$ , i.e.

$$\widehat{\text{Median}} = \hat{S}^{-1}(1/2) = \inf_t \left\{ t : \hat{S}(t) \geq 1/2 \right\}.$$

This median of the survival curve theoretically represents the exact price at which exactly half of the available bidders are ready to purchase the item. If the initial valuation by the auction house is set lower than this price, then more bidders will be interested to

make bids, hence the bidding will continue to grow up. On the other hand, if the initial valuation is set higher than this price, then less number of bidders will be interested, as a consequence, they will try to bargain at a lesser price than quoted. Therefore, this median represents an equilibrium price in the auction mechanism, hence could be used as a proxy of the subjective valuation provided by the auction house. For the combinations of Grade and Source where data is available, we have computed these estimates of the median, and they are summarized in Table 1. It shows that Grade 4 (consisting of D, CD, CHD etc.) is usually the most desirable grade of tea, hence is sown in all of the tea-making regions. On the flipside, Grade 2 (OCD, OD1 etc.) are usually sold at lesser price irrespective of the origin.

Source	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Cluster 1	NA	NA	NA	166	NA	120
Cluster 2	NA	NA	NA	133	NA	110
Cluster 3	125	NA	130	160	NA	123
Cluster 4	121	99	NA	208	NA	106
Cluster 5	114	NA	112	181	NA	136
Cluster 6	113	106	132	201	156	130
Cluster 7	114	95	133	190	NA	125

Table 1: Median bid for different types of tea of different source and grade clusters

### 3.3 Effect of subjective valuation on bid prices

To understand the effect of the subjective valuation provided by the auction house ( $V_i$ ) on the private valuation (or bid prices) by the bidders ( $B_i$ ), we start with Cox’s semi-parametric regression model. Under this model, we assume that  $\lambda_i(t) = \lambda_0(t) \exp(\eta_i)$  where  $\eta_i = \beta_1 \log(V_i) + \beta_2 \log(Q_i) + \sum_{j=1}^6 \beta_{3j} \mathbf{1}\{G_i = j\} + \sum_{j=1}^7 \beta_{4j} \mathbf{1}\{O_i = j\}$ . Using the maximum partial likelihood approach, one can estimate the regression coefficients, and using Breslow’s method, one can simultaneously estimate the baseline hazard  $\lambda_0(t)$ . With the same notation as before, if  $\mathcal{R}(P_{(i)}-)$  denotes the risk set at price  $P_{(i)}-$ , then the estimate of cumulative baseline hazard function becomes

$$\widehat{\Lambda}_0(t) = \sum_{i: P_{(i)} \leq t} \frac{d_i}{\sum_{k \in \mathcal{R}(P_{(i)}-)} \exp(\widehat{\eta}_k)} \quad (2)$$

where  $\widehat{\eta}_k$  is the  $\eta_k$  with beta coefficients replaced by their estimates. We call this as **Model 1**. Based on the available data, we obtain

$$\widehat{\eta}_i = -8.4291872 \log(V_i) + 0.3162530 \log(Q_i) + -0.0606649 \times \mathbf{1}\{G_i = 2\} + 0.1206609 \times \mathbf{1}\{G_i = 3\} + \dots + 1.1791956 \times \mathbf{1}\{O_i = 7\}. \quad (3)$$

The model showed that the log-valuation is a significant factor with a very low z-statistic  $-175.626$  (i.e. the null hypothesis  $\beta_1 = 0$  would be rejected based on asymptotic score test). However, a negative coefficient for  $\log(V_i)$  is doubtful, since it is expected that if the basic valuation increases, then a causal relationship would force the bidders to also



increase their own private valuations. Turning to verify whether the proportional hazard assumptions are satisfied or not, we obtain the cox-snell residuals as  $r_i = \widehat{\Lambda}_0(P_i) \exp(\widehat{\eta}_i)$ , which under the null hypothesis of proportional hazard should be a censored sample from the exponential distribution of mean 1. Therefore, the estimated cumulative hazard rate of the pseudo data  $(r_i, \delta_i)$  should remain close to  $y = x$  line. As seen from Figure 2(a), this is clearly not the case for model 1, hence we can say that proportional hazard assumption is not satisfied.

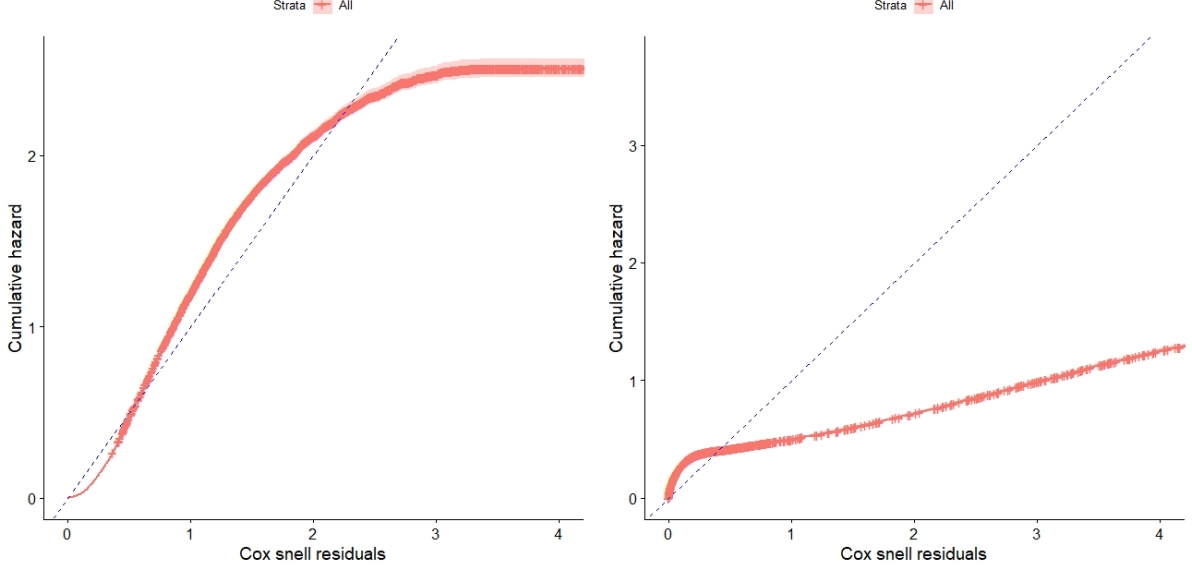


Figure 3: (a) [Left] Cumulative hazard rates estimated from Cox-snell residuals of first model; (b) [Right] Cumulative hazard rates estimated from Cox-snell residuals of second model

Since it is evident from Figure 2 that there are certain differences between the bid raising probability (or the survival function of  $B_i$ ) between different types of tea of varying grades from varying sources, it seems that the proportionality assumption could be violated due to different baseline hazard in different strata. Therefore, we consider the categorical variables  $(G_i, O_i)$  grouped together as a stratum and denote this by  $St_i$ , with  $St_i$  taking values in  $(1, 1), (1, 2) \dots (6, 7)$ . Based on this stratum in effect, we consider a simple strata-based extension of Cox's regression model (**Model 2**) where the hazard function is expressed as

$$\lambda_i(t) = \lambda_{0,St_i}(t) \exp(\xi_i)$$

where  $\xi_i = \beta_1 \log(V_i) + \beta_2 \log(Q_i)$  and  $\lambda_{0,St_i}(t)$  is the baseline hazard corresponding to the strata  $St_i$ . In this case, since  $\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_{0,St_i}(t)}{\lambda_{0,St_j}(t)} \exp(\xi_i - \xi_j)$ , the proportionality assumption of Model 1 is violated. Again, the partial likelihood can be obtained by multiplying the partial likelihood from each stratum, and then one can obtain maximum partial likelihood estimators of the beta coefficients. In this way, we estimate the new Cox's equation as

$$\widehat{\xi}_i = -10.5525692 \log(V_i) + 0.3925898 \log(Q_i). \quad (4)$$

Again, the model shows log-valuation as a significant covariate in the Cox's regression, however, the coefficient of  $\log(V_i)$  remaining negative is counter-intuitive. Therefore, we



again see if proportional hazard assumptions are satisfied using the means of Cox-snell residuals. Similar to before, Breslow's estimator of cumulative baseline hazard for each stratum can be obtained through Eq. (2) where the numerator will only consider  $d_i$ 's from the same strata but the denominator will remain the same as before. Here, the Cox-snell residuals are calculated as  $r_i = \widehat{\Lambda}_{0,St_i}(P_i) \exp(\widehat{\xi}_i)$ , and corresponding residuals and cumulative hazard based on  $(r_i, \delta_i)$  is plotted in Figure 3(b). Clearly, the residuals show even more deviation, suggesting that the Model 2 is not adequate for this data.

Next, we move towards a more general semi-parametric Accelerated Failure Time (AFT) model. Under this model, the private valuation (or bid prices)  $B_i$  is assume to have a log linear representation

$$Y_i = \log(B_i) = \alpha + \beta_1 \log(V_i) + \beta_2 \log(Q_i) + \sum_{j=1}^6 \beta_{3j} \mathbf{1}\{G_i = j\} + \sum_{j=1}^7 \beta_{4j} \mathbf{1}\{O_i = j\} + \sigma w$$

where  $w$  follows an unknown density function  $f$ . Such a representation is also in accordance to the log-normal distributional assumption mentioned in Dalal et. al. [7]. Since no distributional assumption is made for  $w$ , this remains a semi-parametric approach. Miller [13] suggested a least squares approach to solve this problem where only the uncensored part of the data. Buckley and James [14] proposed an alternative to use the least squares on the whole data, but treat  $Y_i^*$  as response variable where

$$Y_i^* = Y_i \delta_i + \mathbb{E}(Y_i | Y_i > \log(P_i))(1 - \delta_i)$$

where the conditional expectation is estimated using the KMPL estimate of the right censored data along with the most recent value of the estimate of the beta coefficients. Buckley and James suggested an iterative procedure that performs this estimate of  $\mathbb{E}(Y_i | Y_i > \log(P_i))$  and then minimizes the least square to update the beta coefficients and iterate. Koul et al. [15] extended that idea to include censoring distribution as well to compute the expectation. Miller [16] and Kleinbaum and Klein [12] provided an excellent review and comparison of these methods. For our purpose, we shall use Buckley and James estimator, computed via the following algorithm.

1. Obtain some initial estimate  $\widehat{\alpha}^{(0)}, \widehat{\beta}_1^{(0)} \dots \widehat{\beta}_{47}^{(0)}$ .
2. For  $t = 0, 1, \dots$ , repeat until convergence
  - (a) Compute the residuals  $e_i = \log(B_i) - \widehat{\alpha}^{(t)} - \widehat{\beta}_1^{(t)} - \dots \widehat{\beta}_{47}^{(t)} \mathbf{1}\{O_i = 7\}$ .
  - (b) Obtain KMPL estimate  $\widehat{F}_{KM}(e)$  based on  $(e_i, \delta)$ . Let, it has a jump  $w_i$  at  $e_i$ .
  - (c) Compute  $B_i^* = B_i \delta_i + (1 - \delta_i) \left[ \widehat{\alpha}^{(t)} - \widehat{\beta}_1^{(t)} - \dots \widehat{\beta}_{47}^{(t)} \mathbf{1}\{O_i = 7\} + \frac{\sum_{j: e_j > e_i} e_j w_j}{\sum_{j: e_j > e_i} w_j} \right]$
  - (d) Update the estimates as

$$\left( \widehat{\alpha}^{(t+1)}, \widehat{\beta}_1^{(t+1)} \dots \widehat{\beta}_{47}^{(t+1)} \right) = \arg \min \sum_{i=1}^n \left( Y_i^* - \widehat{\alpha}^{(t)} - \widehat{\beta}_1^{(t)} - \dots \widehat{\beta}_{47}^{(t)} \mathbf{1}\{O_i = 7\} \right)^2$$

An estimate of the standard error of the converged estimate  $\widehat{\beta}^* = (\widehat{\alpha}^*, \widehat{\beta}_1^* \dots \widehat{\beta}_{47}^*)$  can be obtained as

$$\widehat{\text{Var}}(\hat{\beta}^*) = \frac{1}{n_u - 14} \sum_u \left( Y_i - \hat{\alpha}^* - \hat{\beta}_1^* - \dots - \hat{\beta}_{47}^* \mathbf{1}\{O_i = 7\} \right)^2 (\mathbf{Z}^\top \mathbf{Z})^{-1}$$

where  $\mathbf{Z}$  is the design matrix,  $n_u$  is the number of uncensored observations and  $\sum_u$  denotes summation over only the uncensored observations. Based on the available data, the following regression equation is obtained

$$\begin{aligned} \hat{Y}_i = \log(B_i) = & -0.2541 + 1.1222 \log(V_i) - 0.0435 \log(Q_i) + 0.0185 \times \mathbf{1}\{G_i = 2\} \\ & - 0.0447 \times \mathbf{1}\{G_i = 3\} + -0.1393 \times \mathbf{1}\{G_i = 4\} - 0.087 \times \mathbf{1}\{G_i = 5\} - 0.0493 \times \mathbf{1}\{G_i = 6\} \\ & - 0.0198 \times \mathbf{1}\{O_i = 2\} + 0.0229 \times \mathbf{1}\{O_i = 3\} + 0.0726 \times \mathbf{1}\{O_i = 4\} \\ & + 0.0143 \times \mathbf{1}\{O_i = 5\} - 0.0103 \times \mathbf{1}\{O_i = 6\} - 0.0016 \times \mathbf{1}\{O_i = 7\}. \quad (5) \end{aligned}$$

We also found that the log-valuation and log-volume are significant factors here. Note that, here the coefficient of log-valuation is positive and coefficient of log-volume is slightly negative, which was expected. This means that when the auction house makes its valuation twice as large, then the private valuation  $B_i$  increases 2.177 times as large. The standard error of  $\hat{\beta}_1^*$  equal to as low as 0.0028 also suggests that the estimate is quite reliable. On the other hand, if the quantity of the tea package increases, the usual demand supply equation motivates the bidders to decrease their private valuation (because they should be able to procure a lot with less per unit price and perishable nature ensures that auction house would try to sell even if the price is low). Finally, to check the model assumptions, similar to linear regression, we plot the fitted  $Y_i$ 's and the residuals in Figure 4. As seen from Figure 4, except few outlying observations, most of the residuals are randomly and densely packed near 0, suggesting the goodness of the fitted AFT model.

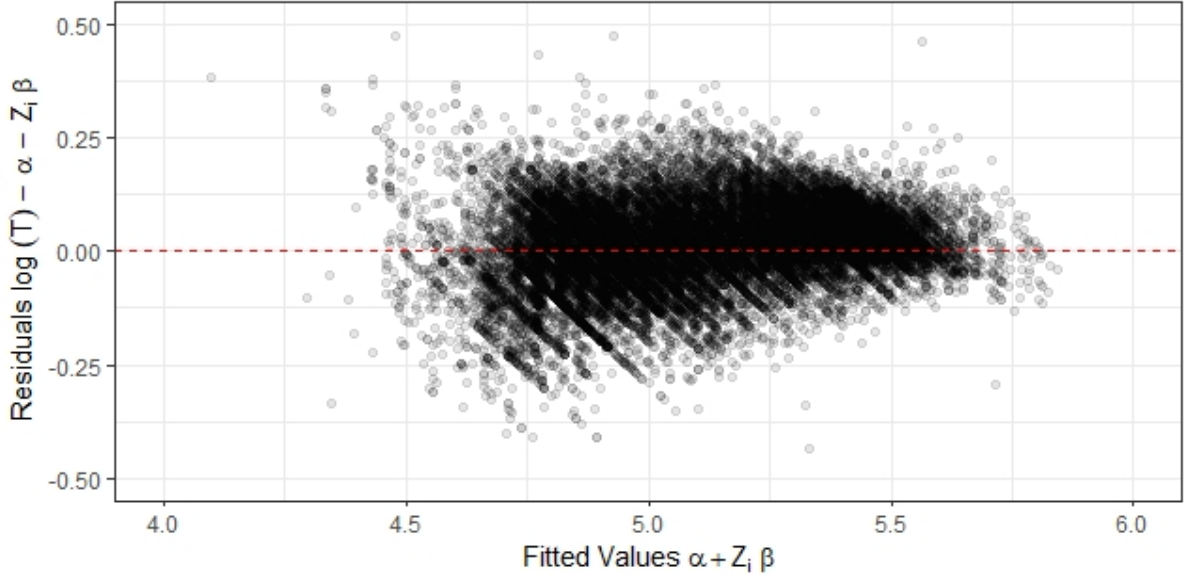


Figure 4: Fitted vs Residuals plot for Buckley James estimate of semi-parametric AFT model

## 4 Conclusion

The final model (i.e. the semi-parametric AFT model) suggested that it can correctly model the private valuations  $B_i$  as a linear function of the log-transformed subjective valuation  $V_i$  provided by the auction house, the volume of the package  $Q_i$  and different grade and origin-specific characteristics of the tea. We found that the log-transformed subjective valuation  $V_i$  is indeed a significant variable in determining the private valuations of the bidders. Therefore, the subjective valuation provided by the auction house is not necessarily just a quality measure of the tea package, but rather a carefully chosen price to increase their potential to sell. The bidders here are at a disadvantage, since the auction house has the ability to change the bidder’s perception about a tea package, irrespective of its quality. To this end, the estimated survival curves or the bid raising probabilities could provide insights into the quality of the tea packages. The indices mentioned in Table 1 should be able to help these bidders as a reference point to speculate whether the auctioneer is manipulating the price of some packages to their own profits, and take preventive measures.

## References

- [1] R Sivanesan. Tea industry in india-analysis of import and export of tea. *International Journal of Business and Management Inventions*, 2(8):09–15, 2013.
- [2] Nizara Arya. Indian tea scenario. *International Journal of Scientific and Research Publications*, 3(7):1–10, 2013.
- [3] Tea Board of India. e-auction process.
- [4] Vijay Krishna. *Auction theory*. Academic press, 2009.
- [5] Edward I George and Sam K Hui. Optimal pricing using online auction experiments: a polya tree approach. *The Annals of Applied Statistics*, pages 55–82, 2012.
- [6] Wush Chi-Hsuan Wu, Mi-Yen Yeh, and Ming-Syan Chen. Predicting winning price in real time bidding with censored data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1305–1314, 2015.
- [7] Abhinandan Dalal, Diganta Mukherjee, and Subhrajyoty Roy. The information content of taster’s valuation in tea auctions of india. *arXiv preprint arXiv:2005.02814*, 2020.
- [8] M. Graham. *Digital Economies at Global Margins*. International Development Research Centre. MIT Press, 2019.
- [9] Janki Hariani, Harshita Malladi, Jash Mehta, Manthan Dedhia, and Kaushik Bhutka. Auction of mines in india. *International Journal for Advance Research and Development*, 3(10):112–117, 2018.
- [10] TeaFountain.com. Tea leaf grades and production methods.

- [11] Dan Levin and James L Smith. Optimal reservation prices in auctions. *The Economic Journal*, 106(438):1271–1283, 1996.
- [12] David G Kleinbaum and Mitchel Klein. *Survival analysis*. Springer, 2010.
- [13] Rupert G Miller. Least squares regression with censored data. *Biometrika*, 63(3):449–464, 1976.
- [14] Jonathan Buckley and Ian James. Linear regression with censored data. *Biometrika*, 66(3):429–436, 1979.
- [15] H Koul, Vyaghreswarudu Susarla, John Van Ryzin, et al. Regression analysis with randomly right-censored data. *Annals of statistics*, 9(6):1276–1288, 1981.
- [16] Rupert Miller and Jerry Halpern. Regression with censored data. *Biometrika*, 69(3):521–531, 1982.