Web Appendix for Estimation of Beauty Contest Auctions

A Nonparametric Identification of Component Mixtures of Bid Distributions and Mixing Probabilities with Unobserved Auction Heterogeneity

We assume that bids arrive independently in this sealed bid setting. Let Q be the number of bids that an auction receives and K the total number of unobserved buyer/auction specific state variables. We present the proof for the case where price is the only bid attribute b_{ji}. Expanding to include more bid-specific variables is straightforward.

Let $\zeta \in \mathbb{R}$ be a cut-off in the positive Real line, and define:

$$\mu_{ji}(A_i,\zeta) = \begin{cases} 1, & \text{if } b_{ji} \leq \zeta \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Then, $\mu_i(A_i,\zeta) = \sum_{j=1}^{Q} \mu_{ji}(A_i,\zeta)$ is the number of times the i^{th} auction, with observed auction attributes A_i, received bids which are less than or equal to ζ. Based on this expansion, at each ζ, $\mu_i(A_i,\zeta), \mu_2(A_i,\zeta), \ldots, \mu_n(A_i,\zeta)$ are i.i.d. Then, we can write out the mixture model as:

$$\mathcal{L}(m,Q) = \sum_{k=1}^{\mathcal{K}} \pi_k \mathcal{B}(m;Q,\mathcal{T}_k(\zeta)) \tag{2}$$

where $\mathcal{T}_k(\zeta) = \int_0^{\zeta} \mathcal{G}_k(b|A_i,v^k)\mathrm{d}b$ is the CDF that a bid-draw b will be less than or equal to ζ, if the unobserved auction type is k, and $\mathcal{B}(m;Q,\mathcal{T}_k(\zeta) = \binom{Q}{m}[\mathcal{T}_k(\zeta)]^m[1-\mathcal{T}_k(\zeta)]^{Q-m}$ is the binomial probability function.

At any given ζ, the number of such equations is Q. The number of parameters to be identified consist of $K-1$ mixing probabilities and K CDFs, $\mathcal{T}_k(\zeta)$s. Thus, at each point in the non-negative real line, ζ, this system of equations is identified if:

$$Q \geq 2K-1 \tag{3}$$

Thus, the CDFs of the component distributions and the mixing probabilities can be retrieved at each point in the non-negative real line if the number of bids received is greater than or equal to $2K-1$. Note that this is a strong identification condition since it requires that the mixing probabilities be identified at each ζ, which is, of course, not necessary since identifying π_ks just once is sufficient for identification.

B Estimation with Assumption 7

B.1 Modified EM-like Algorithm in the First Step

Consider the following nonparametric EM-like algorithm to estimate the bid prices. At this stage, $\mathcal{G}^X(X_{ji}|A_i)$ is assumed to have been already estimated. Since it is not dependent on the unobserved type, its estimation is straightforward. Now, let $AX_{ji} = \{A_i, X_{ji}\}$ take on $\bar{H}$ possible levels, $AX_{ji} \in \{AX^1, AX^2, ..., AX^{\bar{H}}\}$. Then all the bids in the data can be partitioned into $\bar{H}$ groups based on observed auction and seller attributes AX_{ji}. Note that because seller attributes (X_{ji}s) can vary across bids within the same auction, bids from the same auction can belong to different partitions. This is one of the key ways in which this algorithm differs from the standard nonparametric mixture algorithms. Then, the three-step iterative algorithm is as follows:

B.1.1 KDE-Step

Let $\mathcal{G}^t_{h,k}(b|AX^h, v^k)$ denote the probability density function of observing bid price b at observed state AX^h and unobserved type v^k in iteration t. We now define:

$$\mathcal{G}^t_{h,k}(b|AX^h, v^k) = \frac{1}{\mu^t_h \left[\sum_{m=1}^{n_h} \lambda^{t-1}_{mhk}\right]} \sum_{m=1}^{n_h} \lambda^{t-1}_{mhk} \mathcal{K}\left(\frac{b - b_m}{\mu^t_h}\right) \tag{4}$$

where μ^t_h is the bandwidth for group h in iteration t, $\mathcal{K}(\cdot)$ is a univariate kernel such that $\int_R \mathcal{K}(b) d(b) = 1$.

B.1.2 E-Step

In the E-step, we update the posterior probabilities λ^t_{ik}s, for each auction, for this iteration, as follows:

$$\lambda^t_{ik} = \frac{\pi^{t-1}_k \prod_{j=1}^{q_i} \mathcal{G}^t_{h,k}\left(b_{ji}|AX_{ji} = AX^h, v^k\right)}{\sum_{k=1}^{K} \pi^{t-1}_k \left[\prod_{j=1}^{q_i} \mathcal{G}^t_{h,k}(b_{ji}|AX_{ji} = AX^h, v^k)\right]} \forall k \tag{5}$$

where π^{t-1}_ks is the population probability of unobserved type k from the previous iteration. An important point of note is that even when updating the posterior probabilities for a single auction, we might have to use many different bid distributions because the group that a bid belongs to depends on both A_i and X_{ji}.

B.1.3 M-Step

The Maximization step is the same as before.

We iteratively perform the above steps till convergence, at which point, we have consistent estimates of the population probabilities of unobserved types, posterior probability of an auction belonging to a given unobserved type, and $\bar{H} \times K$ probability density functions of bid prices $\mathcal{G}_{h,k}\left(b|AX^h, v^k\right)$.

B.2 Modified Second Step Estimation

The second step estimation is similar to that outlined in §4.2, with the following changes to Step i:

- We first make $(q_i - 1)$ draws of equilibrium seller attributes from $\widehat{\mathcal{G}}^X(X_{ji}|A_i)$. Denote these draws as: $\tilde{X}_{-ji} = \{\tilde{X}_{1i},\ldots,\tilde{X}_{(j-1)i},\tilde{X}_{(j+1)i},\ldots,\tilde{X}_{q_i i}\}$.
- Then, for each draw of X_{ji}, based on $AX_{ji} = \{A_i, X_{ji}\}$, make a draw of bid price from $\widehat{\mathcal{G}}_{h,k}(b_{ji}|AX_{ji} = AX^h, v^k)$. Denote these as $\tilde{b}_{-ji} = \{\tilde{b}_{1i},\ldots,\tilde{b}_{(j-1)i},\tilde{b}_{(j+1)i},\ldots,\tilde{b}_{q_i i}\}$. Together with j's own attributes and bid, this constitutes one simulation of auction i for $v_i = v^k$.

Continue with the estimation as before.

C Details of the First Step Estimation for the Freelancing Context

C.1 Non-Parametric Joint Distributions of Number of Ratings and Mean Rating

We model the joint distributions of these two attributes using bivariate kernel density functions. First, we classify all the bids according to the number of bids received in the corresponding auction (Table A1) and then sub-classify the bids based on the buyer's reputation (Table A2). These classifications allow us to capture the differences in sellers' expectations about the attributes of her competitors. For example, buyers with a large number of high ratings (sub-class 4) receive better bids (*e.g.*, low bid price, high bidder reputation), on average, compared to those with no ratings (sub-class 1). Ideally, of course, we would prefer a more fine-grained classification. However, further classification is not feasible given the size of our data.[1]

Classes	Number of bids
Class 1	$0 <$ Number of bids ≤ 5
Class 2	$5 <$ Number of bids ≤ 10
Class 3	$10 <$ Number of bids ≤ 20
Class 4	$20 <$ Number of bids

Table A1: The four classes of auctions.

Sub-classes	Number of ratings	Avg. rating
Sub-class 1	Number of ratings $= 0$	–
Sub-class 2	$0 <$ Number of ratings	Avg. rating ≤ 9.3
Sub-class 3	$0 <$ Number of ratings ≤ 10	Avg. rating > 9.3
Sub-class 4	Number of ratings > 10	Avg. rating > 9.3

Table A2: The four sub-classes of buyers.

Let $\{1,\ldots,T\}$ be the set of categories, where $T = 16$ since we have four classes for number of bids, each with four sub-classes. $t = 1$ denotes class 1 and sub-class 1, $t = 2$ denotes class 1 and sub-class 2, and so on. The observed bids in each category are indexed by $m \in \{1,2,..,M_t\}$, where M_t is the total number of bids in category t. The m^{th} bid in any category is denoted by vector Y_m, where the two elements of Y_m are the number of ratings and mean rating of the m^{th} bid. We model the probability density function at a point Y in the two dimensional space, in category t, using the multivariate kernel density estimator:

$$\psi_t(Y,\mu_t) = \frac{1}{M_t \mu_t^2 r(l_t,Y)^2} \sum_{m=1}^{M_t} \mathcal{K}\left\{ \frac{1}{\mu_t r(l_t,Y)} (Y - Y_m) \right\} \quad (6)$$

[1] Results are robust to modifications in cut-offs used to sub-classify the data.

where μ_t is the optimal bandwidth window for category t, $\mathcal{K}(\cdot)$ is the two dimensional kernel function satisfying the property $\int_{R^2}\mathcal{K}(Y)d(Y)=1$. $r(l_t,Y)$ is a scaling parameter that represents the Euclidean distance from Y to the l_t^{th} nearest point in the data.[2]

The choice of the bandwidth is crucial to the quality of the kernel estimator. We estimate the optimal bandwidths, $\mu_t \forall t$, using likelihood cross-validation (Silverman, 1986). Let $\hat{\psi}_t(\mu,Y)$ and $\hat{\psi}_{t,-m}(\mu,Y)$ be the PDF estimate of point Y from the t^{th} category using bandwidth μ and datasets $\{Y_1,...,Y_{M_t}\}$ and $\{Y_1,...,Y_{m-1},Y_{m+1},...,Y_{M_t}\}$, respectively. Then the cross-validation score of μ for category t is given by:

$$CV_t(\mu)=M_t^{-1}\sum_{m=1}^{M_t}\ln[\hat{\psi}_{t,-m}(\mu,Y)] \tag{7}$$

The likelihood cross-validation choice of the optimal bandwidth is the value that maximizes $CV_t(\mu)$. Intuitively, the cross-validation score $CV_t(\mu)$ is the log-likelihood of observing the dataset. $\hat{\psi}_{t,-m}(\mu,Y)$ is the probability of drawing the data-point A (assuming it is not part of the dataset). So $M_t^{-1}\Pi_{m=1}^{M_t}\hat{\psi}_{t,-m}(\mu,Y)$ is the total probability of observing the dataset.

While the maximization is conceptually simple, it is computationally intensive. To evaluate the likelihood at a given bandwidth, we need to evaluate the density at each data point at that bandwidth and then sum over the density contributions of all data points. Moreover, at each data point, we need to find the l_t^{th} nearest point in the Euclidean space to calculate its density contribution. This becomes prohibitively expensive as the size of the dataset and the number of dimensions increase. Moreover, this has to be done multiple times to reach the optimal μ_t. To address these computational issues, we follow the recent method proposed by Gray and Moore (2003), which is based on *k-d trees* and has been shown to be much faster than previous methods. We use the MATLAB-based KDE toolbox to perform the estimation (Ihler, 2003).

C.2 Multinomial Logit Model of Bidder's Geographic Region

Sellers can belong to one of four discrete geographic regions (Table 2). Conditional on buyer and auction specific state variables, and a given draw of bidder mean rating and number of ratings, we model the distribution of seller's geographic region using a Multinomial Logit model. Let $\mathcal{Z}_g(region_{ji}|X_{ji}^g,\theta_g)$ be the conditional probability of seller j's region, where X_{ji}^g is the set of variables that influences the draw of the bidders' geographic region and $\theta_g=\{\theta_{g1},\theta_{g2},\theta_{g3}\}$ are the parameter vectors associated with regions 1, 2, and 3, respectively. Then, the probability that bidder

[2] In the absence of $r(l_t,Y)$, the same bandwidth is used for all parts of the distribution. This is problematic in finite samples because it is difficult to pick one optimal bandwidth for the entire range; low bandwidths lead to spurious noise in the tails of the distribution, while high bandwidths cause over-smoothing in the main parts of the distribution (Silverman, 1986). Scaling the bandwidth locally using $r(l_t,Y)$ provides a simple but effective solution to this problem. Further, as is common in the literature, we set $l_t=\sqrt{M_t}$.

j in auction i belongs to geographic region k is:

$$\mathcal{Z}_g(region_{ji}=\gamma|X^g_{ji},\theta_g) = \frac{\exp(X^g_{ji}\theta_{gk})}{1+\exp(X^g_{ji}\theta_{g1})+\exp(X^g_{ji}\theta_{g2})+\exp(X^g_{ji}\theta_{g3})} \forall\gamma\in\{1,2,3\}$$

$$\mathcal{Z}_g(region_{ji}=4|X^g_{ji},\theta_g) = \frac{1}{1+\exp(X^g_{ji}\theta_{g1})+\exp(X^g_{ji}\theta_{g2})+\exp(X^g_{ji}\theta_{g3})} \quad (8)$$

$\theta_{g4}=0$ because $\gamma=4$ is the base region. The set of parameters to be estimated in this context is $\theta_g=\{\theta_{g1},\theta_{g2},\theta_{g3}\}$. The log-likelihood of drawing the sellers' geographic regions observed in the data is:

$$L_g(\theta_g)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{\gamma=1}^{4}\ln[\mathcal{Z}_g(region_{ji}=\gamma|X^g_{ji},\theta_g)^{I(region_{ji}=\gamma)}] \quad (9)$$

Maximizing the above log-likelihood gives us a consistent estimate of θ_g.

C.3 Logit Model of Buyer-Bidder Past Interaction Indicator

The indicator for whether a buyer and seller have interacted in the past is modeled using a binary Logit model. Let $\mathcal{Z}_l(int_{ji}|X^l_{ji},\theta_l)$ be the conditional probability of seller j's having positive past interactions with buyer i, given state variables X^l_{ji} and parameter vector θ_l. Then, the probability of drawing int_{ji} is:

$$\mathcal{Z}_l(int_{ji}=1|X^l_{ji},\theta_l)=\frac{e^{X_{ji}\theta_l}}{1+e^{X_{ji}\theta_l}}; \qquad \mathcal{Z}_l(int_{ji}=0|X^l_{ji},\theta_l)=\frac{1}{1+e^{X_{ji}\theta_l}} \quad (10)$$

The parameters of interest in this context are θ_l, which can be estimated using maximum likelihood. The log-likelihood of observing the indicators of past buyer-seller interactions observed in the data is given by:

$$L_l(\theta_l)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{\gamma=0}^{1}\ln[\mathcal{Z}_l(int_{ji}=\gamma|X^l_{ji},\theta_l)^{I(int_{ji}=\gamma)}] \quad (11)$$

C.4 Nonparametric Estimation of Mixtures of Bid Prices

Figure 4 depicts the kernel density estimate of all the bids in the data. Note that it is lumpy and does not resemble any parametric distribution. Therefore, we employ a fully nonparametric estimation method. To estimate the nonparametric mixtures of bid prices, we first distribute all the bids in the data into bins based on auction and seller attributes as follows:

- Three groups based on number of bids: (a) No. of bids ≤ 13, (b) $13<$ No. of bids ≤ 30, and (c) No. of bids >30.
- Three groups based on sum of seller ratings: (a) Sum seller ratings $=0$, (b) $0<$ Sum seller ratings ≤ 90, and (c) Sum seller ratings >90.
- Two groups based on seller geographic region: (a) Seller region $=1$, and (b) Seller region $\neq 1$.

Note that our binning is coarser than our state space. As discussed earlier, this is necessary in large state space settings with finite samples. We experimented with a large number of binning strategies before finalizing this one. The chosen binning strategy is the one which accomplishes the following three goals to the best possible extent – (1) Bin cut-offs should be such that the distributions of bid prices across bins should be as different as possible. (2) Bin cut-offs should be such that each bin has sufficient data for nonparametric estimation of K unobserved types. (3) Bin cut-offs should be such that each bin has approximately the same amount of data. Unbalanced binning will give rise to some bins with a large number of data points and others with very few data points. In such cases, estimation in smaller bins is likely to be biased.

Our binning strategy gives us a total of $3\times3\times2=18$ bins. Further, in all our analysis, we consider three unobserved types, $K=3$. So we estimate a total of $18\times3=54$ bid distributions. Since we have around 44,000 bids, this gives us approximately 1000 data points in each distribution, if the unobserved types are equally distributed (which they are not). With an unbalanced distribution of unobserved types, some distributions will have fewer points. This is the main reason why further binning is not feasible for this dataset; doing so would introduce significant finite sample bias. Nevertheless, we performed many robustness checks with different binning strategies, and found that the results are, in general, robust to modifications in the bin cut-offs and the number of bins used.

For each of the 18 bins, we employ the nonparametric EM-like algorithm described in detail in Web Appendix §B, and retrieve the 54 component distributions and their population probabilities.

C.5 Nested Logit Model of Buyers' Equilibrium Allocation Rule

Purely nonparametric models of the CCP of buyers' decision, $\mathcal{P}(\cdot)$, are not feasible in our large state space setting with finite data. Hence, we parameterize $\mathcal{P}(\cdot)$ using a Generalized Extreme Value (GEV) distribution and use a flexible nested Logit model. That is, in addition to the observed state variables, we introduce ϵ_{ji}, which is an unobserved seller-auction specific state variable that captures the unobserved preference of the buyer in auction i for seller j. ϵ_{ji} is a $(q_i+1)\times1$ mean-zero vector with support R^{q_i+1} and is assumed to be independent of observed seller attributes X_{ji}. ϵ_{ji} is buyer i's private information and not observable to sellers. Hence, sellers cannot condition their bids on the realizations of ϵ_{ji}s. We further assume that the errors, ϵ_{ji}s, are drawn from a Generalized Extreme Value (GEV) distribution. All the bid options are in one nest, and the cancel option is in a separate singleton nest. Let $\sigma\in[0,1]$ be the correlation of errors in the nest with the bid options, where $\sigma=0$ implies perfect correlation and $\sigma=1$ indicates no correlation. Errors across nests are not correlated.

The probability that buyer i will choose to cancel, and the probability that i will choose the bid from seller j, given state variables $X^n_{ji}=\{A_i,X_{ji},b_{ji},X_{-ji},b_{-ji}\}$ are given by:

$$\mathcal{P}(cancel_i|X^n_{ji},\theta_w,\theta_v,\sigma)=\frac{1}{1+\exp[\mathcal{W}(O_i,\theta_w)+\sigma\mathcal{I}(X_{ji},b_{ji},X_{-ji},b_{-ji},\theta_v,\sigma)]} \tag{12}$$

$$\mathcal{P}(\text{bid}_{ji}|X^n_{ji},\theta_w,\theta_v,\sigma)=\frac{\exp[\mathcal{W}(O_i,\theta_w)+\sigma\mathcal{I}(X_i,b_i,X_{-ji},b_{-ji},\theta_v,\sigma)]}{1+\exp[W(A_i,\theta_w)+\sigma\mathcal{I}(X_{ji},b_{ji},X_{-ji},b_{-ji},\theta_v,\sigma)]}\frac{\exp\frac{\mathcal{V}(X_{ji},b_{ji},\theta_v)}{\sigma}}{\sum_{k=1}^{q_i}\exp\left[\frac{\mathcal{V}(X_{ki},b_{ki},\theta_v)}{\sigma}\right]} \tag{13}$$

$\mathcal{W}(O_i,\theta_w)$ is a function of observable buyer/auction level variables and parameter vector θ_w that affects the upper level nest choice. $\mathcal{V}(X_{ji}, b_{ji}, \theta_v)$ is a function of seller j's attributes and bid price and parameter vector θ_v that dictates the choice of the seller within the bid nest. Finally, $\mathcal{I}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji}, \theta_v, \sigma) = \ln\left[\sum_{k=1}^{q_i} \exp\left(\frac{\mathcal{V}(X_{ki}, b_{ki}, \theta_v)}{\sigma}\right)\right]$ is the inclusive value of the bid nest. Thus, the probability that a specific bid j will be chosen (Equation (13)) is obtained by multiplying the probability that bid nest will be chosen with the probability that bid j will be chosen conditional on the nest choice.

This model is estimated using a maximum likelihood procedure. The parameters $\{\theta_w, \theta_v, \sigma\}$ are obtained from the following maximization:

$$\underset{\theta_w, \theta_v, \sigma}{\operatorname{argmax}} \sum_{i=1}^{n} \left[\ln\left[\mathcal{P}(cancel_i | X_{ji}^n, \theta_w, \theta_v, \sigma)\right]^{I(d_i)=cancel} + \sum_{j=1}^{q_i} \ln\left[\mathcal{P}(\text{bid}_{ji} | X_{ji}^n, \theta_w, \theta_v, \sigma)\right]^{I(d_i)=j} \right] \quad (14)$$

where d_i is buyer i's observed decision.

The estimates for this model are shown in Table A5, and discussed in §6.4.1 in the main text.

D Details of Second Step Estimation for the Freelancing Context – Seller Costs

Finally, using the first step estimates, we estimate seller costs. Since we consider three unobserved types and 18 bins of observed states when binning the bid prices, we estimate 54 kernel density functions of costs, as described in Web Appendix B.2.

A data issue that we face is that the bids are mainly clustered in certain regions (multiples of 50). A key question is whether such clumping is due to the underlying data-generating process being discrete or due to data deficiencies in the finite sample of a continuous distribution. If it is the former, then our FOC approach would be invalid. To ensure that it is not the case, we examined the data further. First, we found that there 372 unique prices between 85 and 500, *i.e.*, most integer prices are covered and the data is far from sparse. Second, we found that the average distance between two adjacent unique bids is only 1.398, *i.e.*, observed bids are mostly close by and hence likely to have come from a continuous distribution.

In principle, we have a choice of either adopting a continuous or discrete choice model here (both would be approximations). We chose the continuous set-up for two reasons. First, with 372 unique price observations, a continuous choice model seemed more appropriate (and feasible) compared to a discrete choice model. Second, from a conceptual standpoint, it is reasonable to assume cost distributions to be continuous.

With the assumption of a continuous distribution, we still have the problem of empty regions in the data. This could be either due to data deficiencies in finite samples or because these regions are off-equilibrium. We handled these gaps by smoothing both $P(\cdot)$ and $G(\cdot)$. We now clarify the need and the impact of this smoothing. Smoothing over off-equilibrium paths is both necessary and common in two-step estimators (Bajari et al., 2007). In our model, we don't need $G(\cdot)$s to be smooth, since they only represent sellers' equilibrium beliefs on the distributions of other bids in the market. We use kernel smoothing for bids simply because from a conceptual standpoint, we believe that cost distributions should have full support. However, we do need mild smoothness requirements for the probability of winning $P(\cdot)$ in order to calculate the derivative of $P(\cdot)$ at optimal observed bids (Assumption 2). To do that, we need estimates of agents' beliefs on winning probabilities in off-equilibrium paths arbitrarily close to the equilibrium path. For example, if we only see \$10 and \$20 bids in data, we not only need estimates of $P(\cdot)$ at 10 and 20, but also at points very close to 10 and 20 to obtain the derivative of $P(\cdot)$ at these values. Note that this smoothness requirement is weaker than that needed in two-step estimators for discrete choice models, where researchers need to smooth over all off-equilibrium paths. In terms of estimation, we tried to make $P(\cdot)$ as flexible as possible.[3]

E Endogenous Buyer Entry

We now present a model of buyer entry. In order to do so, we need to take a stance on buyers' choices. Specifically, we need to treat the estimates from the nested logit model of buyers' decision (Table A5) as structural parameters that define the utility of profit-maximizing buyers.

Before entering the auction, the buyer has uncertainty on both the number of bids he will get and their attributes. We have already modeled and retrieved the equilibrium distribution of bid attributes, $\mathcal{G}(X_{ji},b_{ji}|O_i,q_i,v_i)$. So we now model buyers' expectations on the number of bids they expect to receive after entry.

E.1 Bid Arrival Model

A Poisson model is appealing for two reasons. First, it models count data. Second, a key feature of data generated by Poisson arrivals is the equality of mean and variance, an empirical regularity shared by our data (see Table 4). The expected conditional probability of observing q_i bids as a function of auction specific observables is:

$$\mathcal{H}(q_i|O_i,r_i,\theta_p)=\sum_{k=1}^{K}\pi_k\frac{e^{\eta_{ik}}\cdot(\eta_{ik})^{q_i}}{q_i!} \tag{15}$$

where $\eta_{ik}=\exp(\{O_i,I(v_i=v^k),r_i\}'\cdot\theta_p)$, and θ_p is a parameter vector to be estimated. In the above equation, the unobserved type v_i is integrated out since it is not known to us.

For estimation, we include this bid arrival model within the EM-like loop, and recover both non-parametric estimates of bids and the parameters of the Poisson model as functions of the unobservable v_i, along with population distribution of types. These first step estimates are then used to recover the cost distributions.

[3]Of course, data sparseness is a non-issue if we take a parametric stance on buyers' decision since the derivative of $P(\cdot)$ would then be available from the parametric model.

E.2 Buyers' Decision to Post the Auction

Using the estimates from above, we now specify a buyer entry model. A buyer i chooses to post an auction/enter the market if his expected utility from doing so is greater than that from not entering. If we normalize the utility of not posting an auction to zero (similar to that from canceling), and assume that buyers' costs of making the actual post is negligible, we can write out i's entry decision as:

$$\mathcal{EU}_i + \epsilon_{i0}^{enter} > \epsilon_{i0}^{no-enter} \tag{16}$$

where the first term is the expected value of entering the auction (and making optimal decisions henceforth) and right hand side is the utility from not entering. The two error terms, ϵ_{i0}^{enter},$\epsilon_{i0}^{no-enter}$ are assumed to be i.i.d extreme value. The expected utility from entry can be expanded as:

$$\mathcal{EU}_i = \int_{q_i} \int_{(X_{ji},b_{ji})|_{j=1}^{q_i}} \left[\mathcal{U}_i \cdot \mathcal{G}\Big((X_{ji},b_{ji})|_{j=1}^{q_i} | O_i,r_i,q_i,v_i \Big) \cdot \mathcal{H}(q_i|O_i,r_i,v_i,\theta_p) \right] d\Big(X_{ji},b_{ji})|_{j=1}^{q_i} | O_i,r_i,q_i,v_i) \Big) dq_i \tag{17}$$

where $\mathcal{U}_i$ is the expected utility from posting the auction and receiving q_i bids with attributes $\{X_{ji},b_{ji}\}$ for $j \in \{1,q_i\}$. $\mathcal{U}_i$ is given by:

$$\mathcal{U}_i = 1 + \exp\left[\mathcal{W}(O_i,\theta_w) + \sigma \mathcal{I}((X_{ji},b_{ji})|_{j=1}^{q_i},\theta_v,\sigma) \right] \tag{18}$$

where $\mathcal{W}(O_i,\theta_w)$, $\mathcal{V}(X_{ji},b_{ji},\theta_v)$, and $\mathcal{I}((X_{ji},b_{ji})|_{j=1}^{q_i},\theta_v,\sigma)$ were defined in §C.5. $\mathcal{W}(O_i,\theta_w)$ is a function of observable buyer/auction level variables and parameter vector θ_w that affects nest choice. $\mathcal{V}(X_{ji},b_{ji},\theta_v)$ is a function of seller j's attributes and bid price and parameter vector θ_v that dictates the choice of bid within the bid nest. Finally, $\mathcal{I}(X_{ji},b_{ji},X_{-ji},b_{-ji},\theta_v,\sigma) = \ln\left[\sum_{k=1}^{q_i} \exp\left(\frac{\mathcal{V}(X_{ki},b_{ki},\theta_v)}{\sigma} \right) \right]$ is the inclusive value of the bid nest.

Since the buyer doesn't know how many bids or what kinds of bids he will get before entering the auction, we need to integrate $\mathcal{U}_i$ over his beliefs on the number of bids he expects as well as their attributes. Thus, the inside integral in Equation (17) is the integral over bid attributes for all the bids for a given draw of number of bids (or q_i) and the outside integral is the summation over the distribution of the number of bids, q_i, where $\mathcal{G}(\cdot)$ is the distribution of bid attributes and $\mathcal{H}(\cdot)$ is the distribution of number of bids.

In our estimation, we obtain numerical estimates of $\mathcal{EU}_i$ for each buyer simulating from the estimated equilibrium distribution of bids derived in the paper and the Poisson bid arrival model shown in §E.1. Once we have $\mathcal{EU}_i$, we can obtain the entry probability of buyer i as:

$$P(enter|O_i,r_i,v_i) = \frac{e^{\mathcal{EU}_i}}{1+e^{\mathcal{EU}_i}} \tag{19}$$

E.3 Results of Poisson Model and Entry Probabilities

We find that buyers with longer tenure on the site, and those who have posted many successful auctions and canceled few auctions, are likely to get more bids. (Please see Table A6 in the Web Appendix for the parameter estimates of the Poisson model.) We also find that buyers from the Indian sub-continent and Eastern Europe attract more bids, followed by those from developed countries. Finally, High and Medium type auctions attract fewer bids than Low type auctions, possibly because the supply of sellers who can perform low type jobs is larger.

In Figure A2, we present box plots of entry probabilities of buyers for the three auction types. There are two points of note here. First, entry probability is decreasing with unobserved auction difficulty/type. The average entry probability for a Low type auction is 0.635, whereas it is 0.532 for a High type auction. This difference ($\approx 19\%$) reiterates the importance of accounting for auction specific unobservables. The discrepancy in entry probabilities stems from the difference in the number of bids received and the equilibrium distribution of bid prices across auction types – Low type auctions get many more bids and significantly cheaper ones, increasing the expected value of entry for buyers. Second, after accounting for the entry model, the a priori (before entry) population distribution of the Low, Medium, and High type jobs/auctions is found to be: 16.78%, 45.92%, and 37.29%, respectively. While these post-hoc findings have no impact on the estimates of seller costs, they can have significant implications for counterfactuals; see §6.7 for details.

F Step-by-Step Procedure for Counterfactual Simulations

F.1 Step 1: Solving for sellers' bidding strategy

First, we solve for sellers' bidding strategy given a set of auction attributes and number of bids, $\{A_i, v_i\}$. Since unobserved auction heterogeneity v_i is known during counterfactuals, we essentially have the problem of obtaining optimal bids given a set of state variables. Recall that the FOC of seller's optimization is:

$$\begin{aligned} b_{ji} &= \beta(A_i, X_{ji}, c_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot)) = \beta(O_i, q_i, r_i, v_i, X_{ji}, c_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot)) \\ &= \frac{c_{ji}}{1-r_i} - \left(\frac{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji} | O_i, r_i, q_i, v_i)}{\partial b_{ji}} \right)^{-1} \mathcal{S}_{ji}(X_{ji}, b_{ji} | O_i, r_i, q_i, v_i) \end{aligned} \tag{20}$$

Throughout this section, we use the expanded nomenclature $A_i = \{O_i, q_i, r_i\}$ to be explicit about the primitives that are modified in counterfactuals. To derive a seller's equilibrium bid under a counterfactual scenario, we need information on both the equilibrium distribution of bids, $\mathcal{G}(\cdot)$, and the buyers' decision, $\mathcal{P}(\cdot)$. With a structural interpretation of buyers' decisions, we can treat the estimates from the Nested logit model from §C.5 as primitives of buyer utilities. Since primitives of utility are unlikely to change under counterfactual scenarios, we can continue using them. However, equilibrium distribution of bids will change, and we need to estimate the new $\mathcal{G}^{new}(\cdot)$.

To obtain the new equilibrium distribution of bids and seller attributes, $\mathcal{G}^{new}(X_{-ji}, b_{-ji} | O_i, q_i, r_i, v_i)$, we start by assuming a distribution, $\mathcal{G}^1(X_{-ji}, b_{-ji} | O_i, q_i, r_i, v_i)$, in the first iteration. Then, for each seller j, in each auction i, we solve for the optimal bid b_{ji}^1 in iteration 1 based on the current estimate of seller attributes and bids, $\mathcal{G}^1(X_{-ji}, b_{-ji} | O_i, q_i, r_i, v_i)$, using Equation (20). Note that solving for the optimal bid is not straightforward because the FOC is an implicit function of b_{ji}. Thus, for each seller j in each auction i, we not only need to use a root-finding algorithm such as Newton-Raphson[4] to obtain the new b^1_{ji}, but we also need to numerically simulate the expected probabilities of winning and its derivatives at each step of the root finding algorithm. Next, with the new estimates of b^1_{ji}s and X_{ji}s, we update our estimate of the distribution of seller attributes and bids to $\mathcal{G}^2(X_{-ji},b_{-ji}|O_i,q_i,r_i,v_i)$. These in turn are used to generate the new estimates of bids, b^2_{ji}s. This process continues till the joint distributions of equilibrium bids and seller attributes converge.

This process is computationally intensive, because in order to reach the overall fixed-point of the system, we need to calculate the fixed-point of each agent using a root-finding algorithm at each iteration, which in turn requires numerical simulations at each of its iterations.[5] Since we need to derive the equilibrium bids for all auction-seller combinations observed in the data, this can take some time.

F.2 Step 2: Solve for buyers' entry decisions

Once we have estimates of optimal bids for each auction-seller combination, we solve for buyers' entry decisions. For each combination of auction attributes, $\{O_i,r_i,v_i\}$, simulate entry probability as follows:

- Step (a): Simulate a draw of number of bids, q_i^{new}, from the estimated Poisson bid arrival model. Next, simulate the seller-bid attributes for the q_i^{new} bids using the bid distribution $\mathcal{G}^{new}(X_{-ji},b_{-ji}|O_i,q_i^{new},r_i,v_i)$ estimated in Step 1. Then using the Nested Logit estimates, derive the Inclusive value from entry, $\mathcal{U}_i^{new}$, as specified in §E.2. This constitutes one realization of the auction.
- Step (b): Perform Step (a) a large number of times and average to derive the new expected Inclusive value $\mathcal{EU}_i^{new}$ for the given combination of auction attributes.
- Step (c): Derive the new entry probability $P^{new}(enter|O_i,r_i,v_i)$ using Equation (19).

Using these steps, derive the entry probability for each auction type.

F.3 Step 3: Combine buyer and seller decisions to obtain new system-level equilibrium

Start with the pre-entry population distribution of auction types $\{O_i,r_i,v_i\}$. Draw a large number of auctions from this distribution. For each draw, simulate: (a) buyer's entry decision using estimates from Step 2, (b) number of bid arrivals using estimates of Poisson bid arrival model (Table A6), (c) seller-bid attributes using the bid distributions estimated in Step 1, and (d) buyer's choice decision

[4] Since there exists a unique best-response for a given $G(\cdot)$ and $\mathcal{P}(\cdot)$, any root-finding algorithm will reach the unique optimal bid.

[5] It is possible to ease the computational burden by avoiding root-finding algorithms in the initial iterations. For example, the researcher may simply substitute the previous estimate of the bid in the right hand side of the FOC and obtain the new estimate of the bid. This requires only one set of numerical simulations of probabilities of winning and its derivative per iteration, as opposed to repeated simulations at each step of the root-finding algorithm. We found that employing this method for the first few steps and then switching to the full solution speeds up the convergence considerably without compromising convergence.

using estimates from the Nested logit model (Table A5). Keep track of the outcomes to calculate the new clearance rates and site revenues.

G Code and Monte-Carlo

We provide code for estimating costs in beauty contest auctions. Synthetic data for a set of beauty contest auctions with reputation and price as seller attributes with three unobserved types of auctions are generated for a set of parameter values. The code necessary to retrieve the cost distributions for this dataset are posted on the authors' website. The code comes with a detailed ReadMe file.

To demonstrate the performance of the estimator we employ a bootstrap procedure with 100 samples. Figure A3 shows the three true distribution of costs and Figure A4 shows the retrieved distribution of costs, and the standard errors.

H Additional Tables and Figures

Explanatory variables (X_{ji}^l)	Coefficient	Std. error
Buyer region = 1	-4.585×10^{-1}	6.472×10^{-1}
Buyer region = 2	-3.480×10^{-1}	2.889×10^{-1}
Buyer region = 3	-6.335×10^{-1}	7.785×10^{-1}
Number of bids	-8.369×10^{-2}	1.072×10^{-2}
Square of number of bids	6.262×10^{-4}	1.255×10^{-4}
Indicator for auction attachment	-6.495×10^{-1}	2.279×10^{-1}
Buyer's success ratio	3.485	4.861×10^{-1}
ln(Number of past auctions canceled by buyer)	-4.027×10^{-1}	1.983×10^{-1}
ln(Number of past auctions posted by buyer)	1.132	2.105×10^{-1}
ln(Buyer tenure in days + 1)	-2.543×10^{-1}	6.099×10^{-2}
Indicator for zero buyer ratings	-8.531	4.050
ln(No. of buyer ratings + 1)	-1.634×10^{-1}	1.449×10^{-1}
Buyer mean rating (centered)	-8.833×10^{-1}	4.082×10^{-1}
ln(No. of buyer ratings + 1) × Buyer mean rating (centered)	9.916×10^{-1}	3.045×10^{-1}
ln(No. of seller ratings + 1)	5.664×10^{-1}	5.742×10^{-2}
Seller mean rating (centered)	5.581×10^{-1}	1.365×10^{-1}
Square of seller mean rating (centered)	4.161×10^{-2}	1.782×10^{-2}
Seller mean rating (centered) × Buyer mean rating (centered)	3.320×10^{-2}	7.916×10^{-3}
Seller region = 1	-7.604×10^{-1}	2.301×10^{-1}
Seller region = 2	5.399×10^{-1}	2.129×10^{-1}
Seller region = 3	-5.493×10^{-1}	2.649×10^{-1}
Constant	-7.506	6.015×10^{-1}
No. of observations = 44274; Log likelihood = -868.526		

Table A3: Estimates of Logit model of indicator for buyer-seller past interactions.

Explanatory variables (X^g_{ji})	Region 1		Region 2		Region 3	
	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
Buyer region = 1	1.488×10^{-1}	7.576×10^{-2}	-1.281×10^{-1}	1.020×10^{-1}	-2.631×10^{-1}	1.010×10^{-1}
Buyer region = 2	-4.009×10^{-2}	5.262×10^{-2}	4.392×10^{-1}	6.805×10^{-2}	3.647×10^{-2}	6.639×10^{-2}
Buyer region = 3	-1.523×10^{-1}	9.585×10^{-2}	4.888×10^{-1}	1.372×10^{-1}	1.370×10^{-2}	1.194×10^{-1}
Number of bids	1.186×10^{-2}	1.880×10^{-3}	-8.098×10^{-3}	2.188×10^{-3}	-1.124×10^{-3}	2.294×10^{-3}
Square of number of bids	-9.520×10^{-5}	2.050×10^{-5}	7.450×10^{-5}	2.330×10^{-5}	2.950×10^{-5}	2.430×10^{-5}
Indicator for auction attachment	-9.026×10^{-2}	3.786×10^{-2}	-2.077×10^{-1}	4.597×10^{-2}	-6.432×10^{-2}	4.760×10^{-2}
ln(Deadline_days+1)	-1.181×10^{-2}	1.014×10^{-2}	2.320×10^{-3}	1.213×10^{-2}	2.815×10^{-2}	1.282×10^{-2}
Buyer's success ratio	-3.046×10^{-1}	1.118×10^{-1}	-4.237×10^{-1}	1.314×10^{-1}	-3.408×10^{-1}	1.420×10^{-1}
ln(Total number of past auctions of buyer)	1.544×10^{-1}	7.487×10^{-2}	2.841×10^{-1}	8.655×10^{-2}	7.499×10^{-2}	9.384×10^{-2}
ln(Number of past auctions canceled by buyer)	-1.037×10^{-1}	7.206×10^{-2}	-4.069×10^{-1}	8.300×10^{-2}	-1.377×10^{-1}	9.023×10^{-2}
ln(Sum of buyer ratings + 1)	5.931×10^{-1}	2.026×10^{-1}	1.058×10^{-1}	2.420×10^{-1}	3.968×10^{-1}	2.560×10^{-1}
ln(No. of buyer ratings + 1)	-7.187×10^{-1}	2.329×10^{-1}	-6.453×10^{-2}	2.784×10^{-1}	-4.384×10^{-1}	2.948×10^{-1}
Buyer mean rating (centered)	2.024×10^{-1}	1.169×10^{-1}	-2.197×10^{-1}	1.426×10^{-1}	-3.431×10^{-2}	1.451×10^{-1}
Square of buyer mean rating (centered)	3.197×10^{-2}	1.291×10^{-2}	-2.095×10^{-2}	1.557×10^{-2}	2.381×10^{-3}	1.600×10^{-2}
ln(No. of buyer ratings + 1) × Buyer mean rating (centered)	-7.772×10^{-2}	4.582×10^{-2}	7.884×10^{-2}	5.736×10^{-2}	-3.167×10^{-2}	5.693×10^{-2}
Indicator for zero seller ratings	-2.439×10^{-1}	2.197×10^{-1}	-9.696×10^{-2}	2.934×10^{-1}	-6.915×10^{-1}	3.087×10^{-1}
ln(Sum of seller ratings + 1)	-7.459×10^{-1}	1.818×10^{-1}	7.098×10^{-1}	2.165×10^{-1}	3.550×10^{-1}	2.193×10^{-1}
ln(No. of seller ratings + 1)	8.575×10^{-1}	2.036×10^{-1}	-9.223×10^{-1}	2.438×10^{-1}	-2.772×10^{-1}	2.449×10^{-1}
Seller mean rating (centered)	-1.675×10^{-1}	2.578×10^{-2}	1.331×10^{-1}	3.234×10^{-2}	4.111×10^{-1}	3.503×10^{-2}
Square of seller mean rating (centered)	-2.828×10^{-2}	5.560×10^{-3}	3.099×10^{-2}	6.982×10^{-3}	6.383×10^{-2}	7.349×10^{-3}
Buyer mean rating × Seller mean rating	-1.736×10^{-3}	6.589×10^{-4}	-9.303×10^{-4}	7.757×10^{-4}	3.151×10^{-4}	8.450×10^{-4}
Constant	1.433	4.844×10^{-1}	-1.090	5.767×10^{-1}	-1.572	6.016×10^{-1}
No. of observations = 44272; Log likelihood = -50928.8						

Table A4: Estimates of Multinomial Logit model of seller regions. Region 4 is the base outcome.

Explanatory variables (X_{ji}^{n})	Coefficients varying within the bids nest(θ_v)			
	Model N1		Model N2	
	Coefficient	Std. error	Coefficient	Std. error
price	-6.007×10^{-3}	4.608×10^{-4}	-6.246×10^{-3}	4.791×10^{-4}
price $\times$ Buyer's success ratio	2.532×10^{-4}	4.196×10^{-4}	2.633×10^{-4}	4.363×10^{-4}
price $\times$ ln(Deadline_days+1)	2.529×10^{-4}	1.303×10^{-4}	2.629×10^{-4}	1.354×10^{-4}
Indicator for zero seller ratings	-3.560×10^{-1}	8.132×10^{-2}	-3.702×10^{-1}	8.456×10^{-2}
ln(Number of seller ratings + 1)	1.637×10^{-1}	2.129×10^{-2}	1.702×10^{-1}	2.213×10^{-2}
ln(Number of seller ratings + 1) $\times$ Seller mean rating (centered)	2.065×10^{-1}	1.962×10^{-2}	2.147×10^{-1}	2.040×10^{-2}
Seller region = 1	-1.891×10^{-1}	6.627×10^{-2}	-1.966×10^{-1}	6.891×10^{-2}
Seller region = 2	2.556×10^{-1}	1.179×10^{-1}	2.657×10^{-1}	1.226×10^{-1}
Seller region = 3	3.093×10^{-1}	6.662×10^{-2}	3.216×10^{-1}	6.927×10^{-2}
Seller region = Buyer region	2.548×10^{-1}	1.136×10^{-1}	2.649×10^{-1}	1.181×10^{-1}
Indicator for no buyer-seller past interaction	-1.242	1.510×10^{-1}	-1.291	1.570×10^{-1}
	Coefficients common for bid nest (θ_w)			
ln(Deadline_days+1)	1.526×10^{-1}	5.316×10^{-2}	1.552×10^{-1}	3.400×10^{-2}
ln(No. of buyer ratings + 1)	-4.726×10^{-1}	1.101×10^{-1}	-4.624×10^{-1}	1.091×10^{-1}
ln(No. of auctions uncanceled by buyer)	1.392	1.322×10^{-1}	1.368	1.161×10^{-1}
ln(No. of auctions canceled by buyer)	-9.540×10^{-1}	8.495×10^{-2}	-9.496×10^{-1}	7.669×10^{-2}
Indicator that auction has attachment	1.048	1.124×10^{-1}	1.058	1.126×10^{-1}
Buyer region = 1	-1.930	5.135×10^{-1}	-1.927	5.131×10^{-1}
Buyer region = 2	4.286×10^{-1}	1.853×10^{-1}	4.157×10^{-1}	1.855×10^{-1}
Buyer region = 3	-2.733×10^{-1}	3.863×10^{-1}	-2.802×10^{-1}	3.846×10^{-1}
Indicator high type			1.567×10^{-1}	1.554×10^{-1}
Indicator medium type			-4.067×10^{-1}	1.275×10^{-1}
Constant	-5.327×10^{-1}	2.788×10^{-1}	-3.065×10^{-1}	2.250×10^{-1}
Nest correlation	3.710×10^{-1}	2.343×10^{-2}	3.858×10^{-1}	2.200×10^{-2}
No. of auctions, Log-likelihood	$4002, -2323.155$		$4002, -2312.002$	

Table A5: Nested Logit Estimates of Buyers' Equilibrium Allocation Rule.

Explanatory variables (η_i)	Coefficient	Std. error
ln(Sum of buyer ratings + 1)	-3.418×10^{-2}	2.084×10^{-2}
Buyer region = 1	3.179×10^{-1}	9.178×10^{-2}
Buyer region = 2	1.300×10^{-1}	6.112×10^{-2}
Buyer region = 3	3.458×10^{-1}	1.123×10^{-1}
ln(Buyer tenure in days + 1)	1.769×10^{-2}	9.403×10^{-3}
ln(Number of past auctions canceled by buyer)	-5.849×10^{-2}	2.437×10^{-2}
ln(Number of auctions uncanceled by buyer)	9.663×10^{-2}	3.545×10^{-2}
Indicator high type	-2.099×10^{-1}	7.118×10^{-2}
Indicator medium type	-2.612×10^{-1}	7.890×10^{-2}
Constant	2.430	8.783×10^{-2}
No. of observations = 4002; Log likelihood = -27557.068		

Table A6: Estimates of Poisson bid arrival model, θ_p.

Percentile	Low Type		Medium Type		High Type	
	Model C1	Model C2	Model C1	Model C2	Model C1	Model C2
10%	30.77	32.46	144.33	143.73	354.74	363.70
20%	80.33	82.77	195.46	195.45	361.69	364.92
30%	112.46	118.69	255.40	256.54	363.80	365.35
40%	153.12	153.58	280.48	280.53	364.71	365.51
50%	195.55	195.45	322.49	322.76	365.14	365.58
60%	246.61	254.95	353.36	360.50	365.37	365.61
70%	280.65	285.76	364.71	364.88	365.50	365.63
80%	361.46	362.38	365.49	365.53	365.57	365.64
90%	365.62	365.55	365.62	365.63	365.62	365.65

Table A7: Cost distributions without and with uncertainty in number of bids.

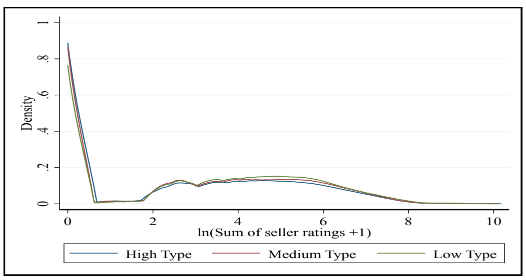

Figure A1: CDF of ln(Sum of seller ratings + 1) for the three auction types – Low, Medium, and High.

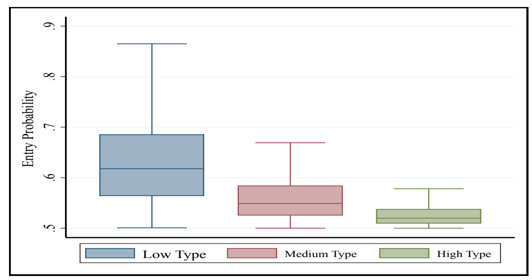

Figure A2: Distribution of buyer entry probability by unobserved auction type.

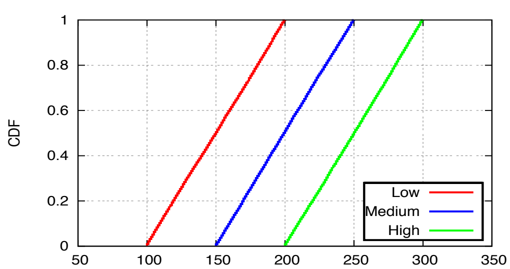

Figure A3: True cost distributions used to generate the synthetic data.

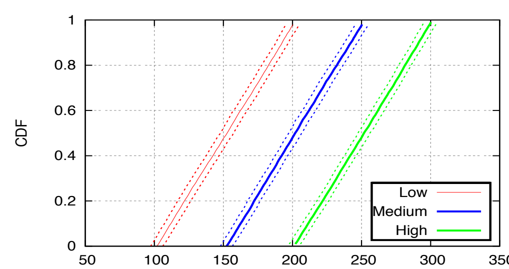

Figure A4: Estimated cost distributions and standard errors for the synthetic data.

References

P. Bajari, C. Benkard, and J. Levin. Estimating Dynamic Models of Imperfect Competition. *Econometrica*, 75(5):1331–1370, 2007.

A. Gray and A. Moore. Rapid Evaluation of Multiple Density Models. In *Artificial Intelligence and Statistics*, 2003.

A. Ihler. Kernel Density Estimation Toolbox for MATLAB, 2003. URL `http://www.ics.uci.edu/~ihler/code/kde.html`.

B. W. Silverman. *Density Estimation for Statistics and Sata Analysis*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1986. ISBN 9780412246203.