Web Appendix for "The Value of Reputation in an Online Freelance Marketplace"

Hema Yoganarasimhan [*]

1 Details of Second-Step Estimation

All the estimates from state transitions $(\theta_p, \theta_g, \theta_t)$, η_is, and non-parametric joint distributions of bid attributes are assumed to be known at this point, *i.e.*, consistent estimates of these parameters and distributions from Step 1 are used.

Throughout we estimate CCPs using weighted nested logits that are flexible functions of $\{x_{it}, \hat{\eta}_i, s_i\}$. When doing so, we treat η_is as if they were observed by substituting $\hat{\eta}_i$ in their lieu. However, because s_i is unobserved, at the each step of the EM algorithm, we weigh the nested logits with current estimates of ρs. We use flexible logits to estimate CCPs because commonly used non-parametric estimators are infeasible in our context because we have many continuous state variables and an extremely large state space. With a sufficiently large sample and an exhaustive set of covariates, CCP estimates from weighted flexible logits are similar to those obtained using non-parametric estimators (Arcidiacono & Miller 2011, Bajari et. al., 2010). In our flexible logit, we incorporate all the state variables, their higher order terms, and interactions.

Step 1: Let the number of unobserved types be 2, and the initial guess of population probabilities be $\pi^1 = \{\pi_1^1, \pi_2^1\} = \{0.5, 0.5\}$. Let $\{\alpha^1, \beta^1, \delta^1, \sigma_1\}$ be the first guess of the structural parameters. We start with estimates from the model without persistent unobserved heterogeneity and a random α_{b11}.

Step 2: Update the ρs as follows:

$$\rho_i^2(k|d_i, x_i; \alpha^1, \beta^1, \delta^1, \sigma^1, \pi^1) = \frac{\pi_k^1 \prod\limits_{t=1}^{T_i} \left[\Pr(d_{it}|x_{it}, \hat{\eta}_i, s_i = k, \alpha^1, \beta^1, \delta^1, \pi^1) \right]^{I(d_{it})}}{\sum\limits_{k=1}^{2} \pi_k^1 \prod\limits_{t=1}^{T_i} \left[\Pr(d_{it}|x_{it}, \hat{\eta}_i, s_i = k, \alpha^1, \beta^1, \delta^1, \sigma^1 \pi^1) \right]^{I(d_{it})}} \, \forall \, k \quad (1)$$

Then update population probabilities of types as:

$$\pi_k^2 = \frac{\sum\limits_{n=1}^{N} \rho_i^2(k|d_i, x_i; \alpha^1, \beta^1, \delta^1, \sigma^1, \pi^1)}{N} \, \forall \, k \quad (2)$$

Step 3: Update CCPs using weighted flexible logits using estimates of individual weights from the previous step $(\rho_i^2 s)$. Denote the CCP of canceling at this step as $p^2(2|x_{it}, \hat{\eta}_i, s_i = k, \rho_i^2)$ for type k.

---
[*]Please address all correspondence to: hemay@uw.edu.

Step 4: To set-up the weighted sum of log-likelihood from Equation (32) using choice probabilities (Equations 19-21), we need to first numerically evaluate $\int V'(x_{it+1}, \eta_i, s_i = k) f(x_{it+1}|1, x_{it}, \eta_i) dx_{it+1}$ for each of the wait options seen in the data, for each k. We do this as follows:

- Step a: For each i, t, such that $t \neq T$, take one step into the future by generating one complete draw of a future state, say $\tilde{x}_{it+1}$, as follows:
  - Make a draw of b_{it+1} (say $\tilde{b}_{it+1}$) from the Poisson distribution $h_p(b_{it+1}|z_{t+1}, \hat{\theta}_p, \hat{\eta}_i)$
  - Make $\tilde{b}_{it+1}$ draws of the three bid attributes, bid price, number of bidder ratings, and average bidder rating, from the joint distribution $\hat{\psi}_c$, where $i \in c$. Denote each of these draws as $\{\tilde{bp}_q, \tilde{bn}_q, \tilde{br}_q\}$, where $q \in \{1, ..., \tilde{b}_{it+1}\}$.
  - For each of these draws, generate $\tilde{gx}_{iq} = \{1, I(g_i = 1), I(g_i = 2), I(g_i = 3), m_i, I(n_i = 0), \ln(n_i + 1), r_i, \ln(\tilde{bp}_q + 1), I(\tilde{bn}_q = 0), \ln(\tilde{bn}_q + 1), \tilde{br}_q\}$. Now draw a seller region for each of the q draws using the MNL distribution $h_g(\tilde{bg}_q|\tilde{gx}_{iq}, \hat{\theta}_g)$.
  - Next, let $\tilde{tx}_{iq} = \{1, I(g_i = 1), I(g_i = 2), I(g_i = 3), m_i, I(n_i = 0), \ln(n_i + 1), r_i, a_i, c_i, l_i, \ln(\tilde{bp}_q + 1), I(\tilde{bn}_k = 0), \ln(\tilde{bn}_q + 1), \tilde{br}_q, \tilde{br}_q^2, I(\tilde{g}_q = 1), I(\tilde{g}_q = 2), I(\tilde{g}_q = 3)\}$. Now draw a buyer-seller past interaction indicator for each of the $\tilde{b}_{it+1}$ draws using the logit distribution $h_t(\tilde{bt}_q|\tilde{tx}_{iq}, \hat{\theta}_t)$.
- Step b: Calculate the CCPs of canceling at $t+1$ based on the projected state $\tilde{x}_{it+1}$ as $\tilde{p}^2(2|\tilde{x}_{it+1}, \hat{\eta}_i, s_i = k, \rho_i^2)$.
- Step c: Then, use $\left[\gamma - \ln(\tilde{p}^2(2|\tilde{x}_{it+1}, \hat{\eta}_i, s_i = k, \rho_i^2))\right]$ as the current estimate of $\int V'(x_{it+1}, \eta_i, s_i = k) f(x_{it+1}|1, x_{it}, \eta_i) dx_{it+1}$ (see Equation 23). To improve the efficiency of the results, we perform Steps a-c 100 times and take the average to obtain the current estimate of $\int V'(x_{it+1}, \eta_i, s_i = k) f(x_{it+1}|1, x_{it}, \eta_i) dx_{it+1}$.

Step 5: Plug the current estimates of $\int V'(x_{it+1}, \eta_i, s_i = k) f(x_{it+1}|1, x_{it}, \eta_i) dx_{it+1}$s into the choice probabilities (Equations 19-21) and update the structural parameters by maximizing:

$$(\alpha^2, \beta^2, \delta^2) = \arg\max_{\alpha, \beta, \delta} \sum_{n=1}^{N} \prod_{t=1}^{T_i} \sum_{k=1}^{S} \rho^2(k|d_i, x_i; \alpha^1, \beta^1, \delta^1, \pi^1) \cdot \ln\left[\Pr(d_{it}|x_{it}, \hat{\eta}_i, s_i = k)^{I(d_{it})}\right] \quad (3)$$

Using the new estimates $\{\alpha^2, \beta^2, \sigma^2, \delta^2\}$, go back to Step 2 and update the ρs. Continue and repeat Steps 2-5 till all the parameters, π_ks, and the weighted log-likelihood (in Step 5) converge. At which point, we have consistent estimates of CCPs, population probabilities of unobserved types π_ks, structural parameters $\{\alpha, \beta, \sigma\}$, and the discount factor δ.

2 Identification Proof for CCPs and State Transitions

We present the proof of identification for CCPs and state transitions.

Bids arrive independently. Let $\mathcal{A}$ be the number of observable bid attributes and $\mathcal{B}$ the maximum number of bids that an auction can receive. Let $\mathcal{T}$ be the maximum time periods for which an auction can remain active, and let $\mathcal{X}$ and $\mathcal{S}$ be the total number of observed and unobserved buyer/auction specific state variables (which remain constant during the auction). Given the structure of our problem (*i.e.*, waiting is the only

continuation decision), the number of possible sequences of decision and state variables is:

$$\text{Seq} = \mathcal{X} \sum_{b=0}^{\mathcal{B}} \frac{(b+\mathcal{T}-1)!}{b!(\mathcal{T}-1)!} \mathcal{A}^b - 1 \tag{4}$$

where $\frac{(b+\mathcal{T}-1)!}{b!(\mathcal{T}-1)!}$ is the number of ways to distribute b bids among $\mathcal{T}$ periods. Without placing any structure, the number of first step parameters to be estimated include the following number of CCPs, bid arrivals[1], joint distributions of bid attributes, and the initial distributions of unobserved types:

$$
\begin{aligned}
\text{CCP Par} &= \mathcal{X}\mathcal{S}\mathcal{T} \sum_{b=0}^{\mathcal{B}} (b+1)\mathcal{A}^b \\
\text{Initial dist. of types Par} &= \mathcal{X}\mathcal{S} - 1 \\
\text{Bid arrival Par} &= \mathcal{X}\mathcal{S}\mathcal{T} \\
\text{Joint bid dist. Par} &= \mathcal{X}(\mathcal{A} - 1)
\end{aligned}
$$

Hence the total number of parameters is:

$$\text{Total Par} = \mathcal{X}\mathcal{S}\mathcal{T} \sum_{b=0}^{\mathcal{B}} (b+1)\mathcal{A}^b + \mathcal{X}\mathcal{S} + \mathcal{X}\mathcal{S}\mathcal{T} + \mathcal{X}(\mathcal{A} - 1) - 1 \tag{5}$$

For identification, we need:

$$
\begin{aligned}
& \sum_{b=0}^{\mathcal{B}} \frac{(b+\mathcal{T}-1)!}{b!(\mathcal{T}-1)!} \mathcal{A}^b \geq \mathcal{S}\mathcal{T} \sum_{b=0}^{\mathcal{B}} (b+1)\mathcal{A}^b + \mathcal{S} + \mathcal{S}\mathcal{T} + \mathcal{A} - 1 \\
\Rightarrow\ & \sum_{b=0}^{\mathcal{B}} \left[\frac{(b+\mathcal{T}-1)!}{b!(\mathcal{T}-1)!} - \mathcal{S}\mathcal{T}(b+1) \right] \mathcal{A}^b \geq \mathcal{S} + \mathcal{S}\mathcal{T} + \mathcal{A} - 1 \tag{6}
\end{aligned}
$$

The left hand side terms have $\mathcal{A}^b$, where $\mathcal{A}$ is very large and $\mathcal{A}^b$ increases exponentially with b. Moreover, the terms within the bracket is positive for $\mathcal{T} \geq 3$ and sufficiently large bs. So as long as we have enough time periods and a large number of potential bids (B), we have more decision-state sequences than parameters. In our case, we have $\mathcal{T} = 14$ and $B = 149$, and even the mean number of bids received by buyers is around 10, so the CCPs and state transitions are non-parametrically identified.

Before we proceed though, we do need to ensure that a key corner case is identified. The corner case in these kinds of identification proofs is always the one where both time and state space dimension are lowest. In our case, since we use CCPs to calculate continuation values of waiting from period 1, we need all CCPs from second period forward to be identified. Hence, the corner case we check is $t = 2$ and one bid. When a buyer has one bid in the second period, there are only two ways for him to have reached that position – a) receive no bids in the first period and wait, and receive one bid in the second period, and b) receive one bid

---

[1] If bid arrivals were not independent, the number of parameters to be estimated for the bid arrival process would be $\mathcal{X}\mathcal{S}\mathcal{T}(\mathcal{B} - 1)$ instead of $\mathcal{X}\mathcal{S}\mathcal{T}$.

in the first period and wait, and receive no bids in the second period. Thus to calculate terminal CCPs for the second period in this state (one bid), we only have two equations. With two equations, we can identify CCPs for at most two unobserved types, but no more. Hence, this corner case establishes that, in this setting, we can only allow for two unobserved types.

Finally, note that even though the number of sequences is much greater than parameters, first period termination CCPs (CCPs for cancellation and bid choice) are not identified because there is only one path to reach a state in the first period and once a buyer terminates the auction, there are no paths out of it. Hence, these CCPs don't appear in more than one sequence and cannot be identified. However, we don't use first period CCPs in our analysis; only CCPs from the second period and forward are used in continuation value calculations. Hence the non-identification of first period termination CCPs is not a problem.[2]

3 Simulation Details: Aggregate Choice Probability Maps

Consider a focal bid j with fixed reputation attributes and price, say $\{bp_j, bn_j, br_j\}$. Simulate each auction i 100 times using the steps described below.

1. For each iteration of auction i, start with $t = 1$ and simulate each period as follows:

    - Draw the number of bid arrivals for this period using the Poisson estimates and $\hat{\eta}_i$. If the first bid for this iteration of the auction arrives in this period, treat it as the focal bid with reputation and price $\{bp_j, bn_j, br_j\}$. Then, draw geography and past interaction indicators for it using the logit models.

    - Then draw bid attributes for each non-focal bid that arrives in this period using the estimated KDEs and logit models.

    - If $t \neq T$, calculate the continuation value and derive the probability of picking the focal bid (if it exists), picking one of the other bids (if they exist), waiting, or canceling the auction using the closed form expressions for each unobserved type $s_i = \{1, 2\}$. Then weight the probabilities using ex post probabilities of belonging to the two unobserved types (*i.e.*, ρ_is from the last iteration of the EM algorithm).

    - Construct a CDF using the above probabilities, and draw a random number from 0 to 1 to simulate the outcome for this period.

2. Record the outcome for this period as follows:

    - If focal bid was chosen, record this, and terminate this iteration of auction i.

    - If another bid was chosen or if auction was canceled, record this, and terminate this iteration of auction i.

    Then, start next iteration for auction i if 100 iterations haven't been reached, else move on to the next auction.

    - If wait is chosen, proceed to the next time period for this iteration of auction i.

3. When 100 iterations for auction i are completed, calculate the probability of the focal bid winning auction i as = (number of times it was picked in the 100 iterations)/100.

---

[2]Unlike first period termination CCPs, first period wait CCPs are identified because they appear in many decision-state paths.

4. When all the auctions have been simulated and the average probability of being picked in each auction has been calculated, sum the probabilities, and divide by the total number of auctions to get the average success probability for the focal bid in the website.

4 Details of Backward Solution Method used in Counterfactuals

All the estimates from state transitions $(\theta_p, \theta_g, \theta_t)$, η_is, non-parametric joint distributions of bid attributes, and the structural parameters $(\alpha, \beta, \delta, \sigma, \pi)$ are treated as known at this point.

First, split the observed state variables at time t, x_{it}, into auction/buyer specific variables that time invariant for the duration of the auction, xa_i, and bidder specific state variables that are denoted as xb_{it}. Then based on the definition in the paper, write out the inclusive value of the bid nest for buyer i at time t as: $I_{it} = I(xb_{it}) = \ln\left[\sum_{q=1}^{B_{it}} e^{\frac{Y_{iq}(xb_{it})}{\sigma}}\right]$, where $Y_{ij}(xb_{it})$ is defined in Equation (6) and B_{it} is the total number of bids that i has at period t. Next, discretize the state variables, xa_i, I_{it}, and η_i by allowing them to vary from zero to a high enough value, in steps of Δ, such that: $xa_i \in \{0, xa_{max}\}$, $I_{it} \in \{0, I_{max}\}$, and $\eta_i \in \{0, \eta_{max}\}$. Using these definitions, specify the continuation value as a function of the time-invariant observables and the inclusive value at time t as $V'(t, ax_i, I_{it}, \eta_i, s_i)$. Now to generate the value functions at each combination, follow the steps given below:

Step 1: Start at the last time period T. Since waiting is not an option at T, there is no continuation value in this period. Hence, the conditional choice probability of cancellation at T is simply obtained from the following equation:

$$P(2|T, xa_i, I_{iT}, \eta_i, s_i) = \frac{1}{1 + e^{[W_{ib}(xa_i, s_i) + \sigma I_{iT}]}} \tag{7}$$

Now using all the values in the grid of a_i and I_{iT} (in $\{0, xa_{max}\}$ and $\{0, I_{max}\}$ in steps of Δ), populate the cancellation probability matrix. Note that at this point we can simply ignore η_i and s_i since we are just calculating the cancellation probabilities at different levels of a_i and I_{iT} at T.

Step 2: Next, use the matrix of cancellation probabilities at T to calculate the matrix of continuation values at each point in the grid of $\{xa_i, I_{iT}\}$ using the following equation:

$$V'(T, xa_i, I_{iT}, \eta_i, s_i) = \gamma - \ln(P(2|T, xa_i, I_{iT}, \eta_i, s_i)) \tag{8}$$

Step 3: Now move to period $t = T - 1$ and at each grid point in $\{a_i, I_{iT}, \eta_i\}$, take the following steps:
- Step 3a: Follow Step a under Step 4 in §1 of the Technical Appendix in order to simulate one future draw of bids for period $t + 1$. Based on the draw of future bids, obtain the new inclusive value I_{it+1} (using both current bids and simulated bids). Then look up the matrix of continuation values calculated in the previous step and obtain $V'(t + 1, xa_i, I_{it+1}, \eta_i, s_i)$
- Step 3b: Repeat Step 3a for a large number of times to obtain $\int V'(t+1, xa_i, I_{t+1}, \eta_i, s_i) f(x_{it+1}|d_{it}, x_{it}, \eta_i) dx_{it+1}$.
- Step 3c: Now plug back this integral into the choice-specific continuation value of waiting at t to obtain:

$$v(1, t, xa_i, I_{it}, \eta_i, s_i) = W_{iw}(xa_i) + \delta \int V'(t + 1, xa_i, I_{it+1}, \eta_i, s_i) f(x_{it+1}|1, x_{it}, \eta_i) dx_{it+1}$$

- Step 3d: Substitute this choice specific value function and the parameter estimates in Equation 20 to obtain the cancellation choice probability as:

$$P(2|t, xa_i, I_{it}, \eta_i, s_i) = \frac{1}{1 + e^{[v(1,t,xa_i,I_{it},\eta_i,s_i)]} + e^{[W_{ib}(xa_i,s_i)+\sigma I_{it}]}} \tag{9}$$

Performing Steps 3a-3d at all the grid points in $\{xa_i, I_{it}, \eta_i\}$ should give a matrix of cancellation probabilities at t. Then, as in Step 2, use the matrix of cancellation probabilities at t to calculate the matrix of continuation values at each point in the grid of $\{xa_i, I_{it}, \eta_i\}$ using the following equation:

$$V'(t, xa_i, I_{it}, \eta_i, s_i) = \gamma - \ln(P(2|t, xa_i, I_{it}, \eta_i, s_i)) \tag{10}$$

Step 4: Repeat Step 3 for all periods till the algorithm reaches $t = 1$ and obtain the continuation values $V'(t, xa_i, I_{it}, \eta_i, s_i) \,\forall\, 1 \leq t < T$.

Step 5: With the matrices of continuation values derived above and the estimated parameters, the system can now be simulated as many times as necessary to generate the counterfactual outcomes.

5   Additional Tables and Figures

Project Title	MaxBid	Project Description
Web Site Photo Gallery Assistance	20	Have code already for a website's photo gallery. And for the website, I just need gallery code merged into existing page.
Need Recommendations for AdSense placement in our website	50	Hi, We have placed AdSense Ads in our website, even though we are getting some revenue from AdSense, we are not getting return on our investment. So, we would like to try placing AdSense Ads in strategic places and need recommendations. We need recommendations with size of Ads, location of Ad placement and any other help that you can give us to increase our revenue through Ads. Thanks
Wordpress Theme Modification	80	Hi Guys, I got the "TheCorporation" theme from elegant designs. I have a PSD ready on how the theme needs to look like. I also have the original PSD files of the theme. I need you to edit the theme so that it will look exactly the same as in the PSD that it needs to look like.
Streaming Data Capture	100	I am looking for somone [sic] who can write me a program that can store streaming numerical data on a web page to an excel spreadsheet, or as csv or at least as a text file.

Table 1: Examples of project titles and descriptions

z_t	θ_p	std. error
Ind. $t = 1$	4.4866	0.0445
Ind. $t = 2$	2.6337	0.0447
Ind. $t = 3$	1.8599	0.0455
Ind. $t = 4$	1.3820	0.0468
Ind. $t = 5$	1.0787	0.0484
Ind. $t = 6$	0.8096	0.0500
Ind. $t = 7$	0.6770	0.0512
Ind. $t = 8$	0.4983	0.0518
Ind. $t = 9$	0.3739	0.0525
Ind. $t = 10$	0.3037	0.0550
Ind. $t = 11$	0.2351	0.0552
Ind. $t = 12$	0.1358	0.0571
Ind. $t = 13$	0.0513	0.0572

No. of auctions, auctions×timeperiods = 6076, 28817

Goodness of Fit Measures

RMSE, MAE, ME = 0.9258, 0.4249, 3.59×10^{-8}

Table 2: Estimates of θ_p from the fixed effects Poisson model. Standard errors are adjusted for clustering on auction identity. All the estimates, except the last two, are significant at the 1% level (the parameter coefficient for Ind. $t = 12$ is significant at the 10% level).

gx_{ij}	Region 1 θ_{g1}	Std. error	Region 2 θ_{g2}	Std. error	Region 3 θ_{g3}	Std. error
Buyer region = 1	0.2094	0.0426	−0.1800	0.0597	−0.1322	0.0561
Buyer region = 2	−0.0496	0.0288	0.2007	0.0384	−0.0042	0.0365
Buyer region = 3	0.0051	0.0185	0.0382	0.0242	0.0563	0.0226
MaxBid	−0.0006	0.0004	−0.0000	0.0005	0.0025	0.0005
Ind. no. of buyer ratings = 0	−0.3433	0.1531	0.7286	0.2137	−0.3328	0.1890
ln(no. of buyer ratings + 1)	0.0236	0.0068	0.0324	0.0087	−0.0260	0.0086
Buyer mean ratings	−0.0338	0.0154	0.0660	0.0215	−0.0276	0.0190
ln(bidprice + 1)	0.0654	0.0184	0.1120	0.0236	−0.0861	0.0229
Ind. no. of seller ratings = 0	−1.6200	0.0809	1.1569	0.1232	3.7824	0.1554
ln(no. of seller ratings + 1)	−0.0046	0.0067	−0.1334	0.0087	0.0376	0.0082
Seller mean rating	−0.1732	0.0085	0.1491	0.0098	0.4270	0.0159
Constant	3.0200	0.1789	−2.4263	0.2535	−3.5782	0.2504

No. of observations = 126927

Log likelihood = −145597.83

Table 3: Estimates of $\theta_{g1}, \theta_{g2}, \theta_{g3}$ from the Multinomial Logit model. Region 4 is the base outcome.

tx_{ij}	θ_t	Std. error
Buyer region = 1	-0.0041	0.1277
Buyer region = 2	0.0048	0.0813
Buyer region = 3	-0.0363	0.0616
MaxBid	-0.0109	0.0014
No. of uncanceled past auctions	0.0007	0.0001
Cancel ratio	-0.8858	0.1050
Buyer tenure on site (in days)	-0.0006	0.0000
Ind. no. of buyer ratings = 0	-1.2024	0.4578
ln(no. of buyer ratings + 1)	0.5432	0.0220
Buyer mean ratings	-0.0165	0.0513
Seller region = 1	0.1341	0.0638
Seller region = 2	0.3069	0.0777
Seller region = 3	0.0115	0.0258
ln(bidprice + 1)	0.2271	0.0589
I.(no. of seller ratings = 0) = 1	0.1668	0.9083
ln(no. of seller ratings + 1)	0.4639	0.0153
Seller mean rating	-0.1491	0.2145
Square of seller mean rating	0.0323	0.0129
Constant	-7.4925	1.0504
No. of observations = 126927		
Log likelihood = -9671.063		

Table 4: Estimate of θ_t from the Logit model. $bt_{ij} = 0$ is the base outcome.

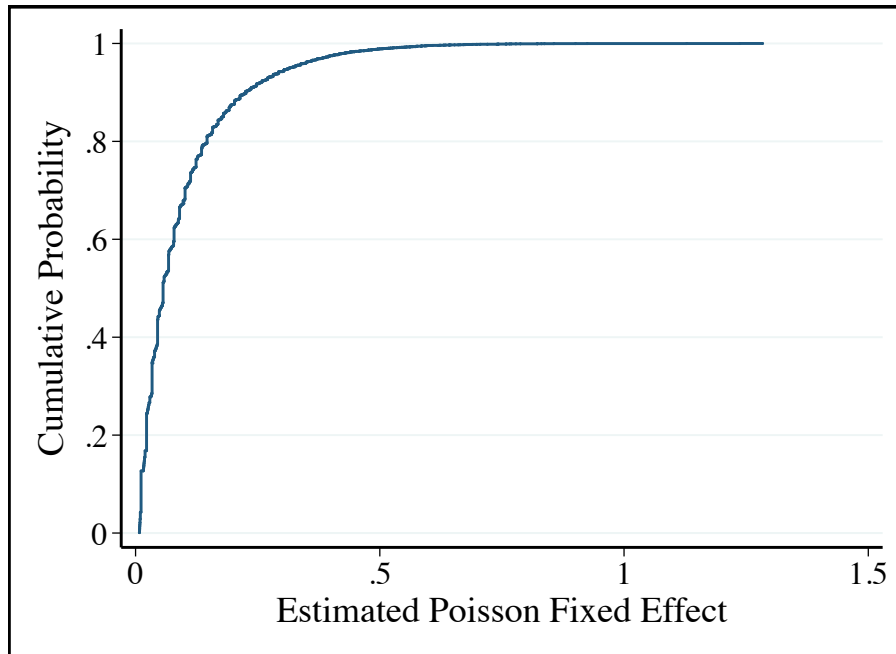Figure 1: CDF of estimated Poisson fixed effects ($\hat{\eta}_i$s) for the auctions in our sample. Summary statistics of the distribution: Mean, Std. Dev, 25th, 50th, 75th percentiles = 0.0952, 0.1068, 0.0246, 0.0563, 0.1238.