Pricing Externalities in Real-Time Bidding Markets

Joseph Reisinger[∗†]
joeraii@cs.utexas.edu

Michael Driscoll[†]
michael@metamarketsgroup.com

[∗]Department of Computer Science
University of Texas at Austin
Austin TX 78705

[†]Metamarkets Group
300 Brannan Suite 507
San Francisco CA 94107

Abstract

Online publishers rely on *real-time bidding* (RTB) markets to sell their remnant inventory and increasingly to command higher prices for "premium" content. However, content providers have lagged advertisers historically in the sophistication of their pricing models, as evidenced by the large increase in demand-side platforms without corresponding investment on the sell-side. Informed advertisers collect user-intent and demographic information in addition to the publisher context in order to score the relevance of impressions. The resulting differential impression pricing is only visible to publishers as a positive externality to revenue unless they collect the same audience targeting information. In this paper, we introduce a Bayesian hierarchical model of auction clearing price that can naturally account for the presence of *publisher-advertiser information asymmetry* and quantify its impact on price. Although the current model is simply exploratory, it suggests that richer models of dynamic floor pricing may improve publisher revenue.

1 Introduction

Real time bidding (RTB) markets have emerged as the preferred medium for ad networks and large advertisers to buy remnant inventory [3]. Individual publisher impressions are auctioned off in real time to participating advertisers, allowing them fine-grained control over audience targeting. In theory, publishers set floor prices in line with their view of the value of their inventory, and the degree of risk they must take on selling media to potentially unknown third-parties. Advertisers bid on the individual impressions, buying specific audience information, such as demographic, session history and intender status, from third-party data providers and demand-side platforms (DSPs), leading to information asymmetry between the demand- and supply-side. This presence of *informed bidders* amongst advertisers bidding on particular inventory causes *adverse selection* [6], with publishers raising floor prices across the board to avoid selling inventory at a perceived discount.

In this paper we explore the effects of informed bidders and information asymmetries between the supply- and demand-side in RTB markets and the resulting effects on empirical ad market microstructure. We posit that quantifying the effects of *information externalities* on informed bidding in aggregate can lead to more informed supply-side floor pricing, and hence increased publisher revenue, without the need for publishers to identify what impression level information is being specifically acted on by bidders. That is, the *presence* of differential pricing strategies, such as those employed by DSPs can be inferred directly from the bid price distribution.

Towards this end, we develop a mean-shift latent variable model in the context of linear regression to study publisher-advertiser information asymmetry, applying it to a large anonymized auction data set. The fundamental model assumption is that additional information available to a subset of informed advertisers, e.g. provided by DSPs, affects bid price additively, causing it to be overdispersed when compared to the baseline model. Hence, markets undergoing significant adverse selection due to information asymmetries will appear to publishers as additional clearing price dispersion. Although the underlying signals driving differential pricing may not be available on the supply side, publishers can still pool their auction data to estimate its economic impact.

In addition to the basic model, we discuss several extensions, including the potential to improve dynamic floor pricing mechanisms and produce more accurate estimates of marginal floor price. Ultimately, pooling sell-side data will help give publishers more fine-grained control over their inventory pricing and improve market efficiency.[1]

2 Mean-shift Mixture of Generalized Linear Models

We propose a simple generative model of auction clearing price p_i^a as a function of floor price p_i^f, publisher id x_i^{pid}, and a latent externality indicator z_i. Publishers set the floor price distributed over their inventory as a noisy signal of quality, forcing higher correlation between p^f and p^a. Advertiser willingness-to-pay is derived from a latent audience signal (unobserved information externality z_i) and site context.

Although the exact latent audience signal z_i cannot be reconstructed from the data, an aggregate estimate can be obtained by treating it as a latent variable in an overdispersed generalized linear model (GLM) framework,

$$\begin{array}{lcll}
\mathbf{w}|\Sigma_{\mathbf{w}} & \sim & \mathcal{N}(0, \Sigma_{\mathbf{w}}) & \text{(parameter weights)} \\
z_i|\mathbf{x}, \mu_0, \mu_1, \sigma_0, \sigma_1 & \sim & GMM(\cdot|\mathbf{x}, \{\mu_0 \leqslant \mu_1\}, \{\sigma_0 = \sigma_1\}) & \text{(latent group indicator)} \\
\boldsymbol{\alpha}|\sigma_\alpha & \sim & \mathcal{N}(0, \sigma_\alpha) & \text{(price mean-shift)} \\
p_i^a|\mathbf{x}_i, z_i, \mathbf{w}, \alpha & \sim & GLM(\cdot|\left[\begin{array}{c}\mathbf{x}_i \\ z_i\end{array}\right], \left[\begin{array}{c}\mathbf{w} \\ \boldsymbol{\alpha}\end{array}\right]) & \text{(regression)}
\end{array}$$

where $\mathbf{x}_i = \{p_i^f, x_i^{pid}\}$. This model combines a standard GLM with a two-component, equal variance Gaussian mixture model (GMM) indicating whether the auction price has been mean-shifted due to (unobserved) impression-level information. Such mean-shift mixture models are common in outlier detection, and can be used to model over-dispersion due to unobserved factors [2]. GLMs with latent factors can be fit using standard EM techniques; we adopt a Bayesian approach using Gibbs sampling [cf. 5].

Using this framework, we derive three particular models of p^a:

- **Externality-free** – Auction clearing price depends only on the observed variables,

  $$p_i^a = w_0 + w_{0,x_i^{pid}} + (w_1 + w_{1,x_i^{pid}})p_i^f + \epsilon_i$$

  where $w_{0,\cdot}$ are intercept parameters and $w_{1,\cdot}$ are slope parameters. This model captures the contribution of the floor price and publisher id to the prediction of the auction clearing price.

- **Aggregate Externalities** – Audience pricing externalities are assumed constant across all publishers, hence some percentage of each publishers inventory experiences a latent mean-shift:

  $$p_i^a = w_0 + w_{0,x_i^{pid}} + \alpha_0 z_i + (w_1 + w_{1,x_i^{pid}} + \alpha_1 z_i)p_i^f + \epsilon_i$$

  where α_0 and α_1 are additional intercept and slope parameters respectively for mean-shifted observations. This model captures additional price dispersion not accounted for in the baseline **externality-free** model, i.e. separating the systematic/floor-dependent portion of the bid from the additional unobserved audience segment signal.

- **Publisher-dependent Externalities** – Additional per-publisher coefficients are included to address *contextual* pricing externalities,

  $$p_i^a = w_0 + w_{0,x_i^{pid}} + (\alpha_0 + \alpha_{0,x_i^{pid}})z_i + (w_1 + w_{1,x_i^{pid}} + (\alpha_1 + \alpha_{1,x_i^{pid}})z_i)p_i^f + \epsilon_i$$

  This model captures per-publisher deltas on the **aggregate externalities** model.

[1]For example, empirically, floor prices in RTB markets may be set too high, reflecting unreasonably high yield expectations for remnant inventory given the underlying market mechanics [8].

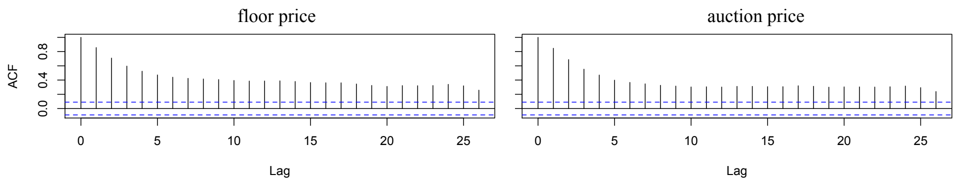

Figure 1: Hourly autocorrelation for $\mathbb{E}[p^f]$ and $\mathbb{E}[p^a]$ remains significant even for lags of a day or more, indicating time-of-day effects acting on auction pricing. We account for these effects before modeling other externalities by correcting for hourly residuals.

as well as variants that do not include publisher-specific effects (e.g. $w_{1,x_i^{pid}}$).

Examining the fits produced by each of these models allows us to determine the extent of differential pricing based on information externalities both across- and within publishers. If the model assigns $\mu_0 = \mu_1$ or $\alpha = 0$, then there is no additional price dispersion to be accounted for, and all price differentials must be due to publisher context. Furthermore, if $\alpha + w_0$ in the **aggregate externalities** model is less than w_0 in the **externality-free** model, then additional audience information has a *negative* impact on overall revenue, and vice-versa.

Finally, we note that this model is completely retrospective, and cannot be used to infer the existence of externalities acting on impression auctions before they take place; rather its utility lies in its use as a component for a differential floor-pricing model and demand-side weathervane.

3 Pricing Demand-Side Externalities

Dataset: 9M impression auctions from a real-time bidding market[2] with 5K publishers and 70 active advertisers[3] over a 2 week span during July 2010. Publisher ID and floor price are observed for all auctions; auction outcomes are observed in the form of the auction clearing price (second highest bid or floor price, whichever is higher). Since price data is typically closer to log-normal (no negative prices), we transform p^a and p^f into the log-domain. We also identify significant autocorrelations in both p^a and p^f on an hourly scale (Figure 1), indicating that time-of-day plays a role in price dispersion. To account for this potentially confounding effect, we fit a null model regressing on hour and adjust p^a and p^f according to the residuals.[4]

Table 1 provides statistics from the top publishers by volume from our auction sample, and highlights a wide variety in publisher strategies and inventory qualities. For example, all publishers except for 5 and 8 have dynamic floor pricing (nonzero floor price variance). Publisher 4 sets the highest floor price, maintaining inventory sell-thru of 10% and also has the highest divergence between floor and clearing price $\mathbb{E}(p_a - p_f)$. The publishers with the highest floor prices also have the highest correlation between p_a and p_f, indicating that advertisers are not willing to significantly overbid the floor. In general, publishers with low correlation between p_a and p_f are experiencing the most differential selection by informed bidders.

Table 2 summarizes the deviance[5] and most salient parameter coefficients for each of the models. The addition of the latent externality indicator z_i significantly improves the model fit, lending evidence for aggregate differential pricing based on unobserved information. However, the addition of per-publisher externality effects does not significantly reduce deviance beyond the aggregate model, indicating that advertisers may not be pursuing differential targeting based on publisher context. Rather, they may be primarily targeting cross-cutting demographics and user cookies. Figure 2 summarizes the contribution of each component of the **publisher-dependent externalities** model to the overall fit.

[2]Modified Vickrey (second-price) auction where the winning bidder pays the second highest price + $0.01.

[3]Winning at least one auction.

[4]The main results presented here do not depend on this correction.

[5]$D(y) = -2\left[\log(p(y|\hat{\theta}_0)) - \log(p(y|\hat{\theta}_s))\right]$, where θ_0 are the parameters of the inferred model and θ_s are the parameters of a model with perfect fit (one parameter per data point)

Publisher	n	$N(p_a > p_f)$	$\mathbb{E}[p_f]$	$\mathbb{E}[p_a]$	$\rho(p_a, p_f)$	$\mathbb{E}(p_a - p_f)$
0	1.6M	192K	125$\pm$ 75	151.3$\pm$ 96.8	0.72	45.1 $\pm$ 79.9
1	1.5M	61.5K	35$\pm$ 77	113.1$\pm$ 110.4	0.53	66.1 $\pm$ 89.3
2	219K	37.9K	568$\pm$ 504	462.9$\pm$ 258.1	0.82	150.2 $\pm$ 199.0
3	174K	11.6K	111$\pm$ 40	204.3$\pm$ 167.2	0.51	100.1 $\pm$161.1
4	138K	12.8K	734$\pm$ 396	632.1$\pm$ 254.8	0.87	151.0 $\pm$ 120.2
5	95K	35K	250$\pm$0	388$\pm$ 130.0	-	138.7 $\pm$ 130.0
8	44K	26K	0$\pm$ 0	129$\pm$ 169.9	-	129.3 $\pm$ 169.9

Table 1: Examples of publisher impression price distributions for 7 of the top 10 publishers by volume. n is total impressions, $N(p_a > p_f)$ is the number of successful (cleared) auctions, $\mathbb{E}[p_f]$ is the average floor price (CPM in cents), and $\mathbb{E}[p_a]$ is the average auction clearing price. $\rho(p_a, p_f)$ is the Spearman's correlation between the floor and clearing price and $\mathbb{E}(p_a - p_f)$ is the expected lift over the floor price. Errors are standard deviations.

Model	$D(y)$	$\mathbb{E}[w_0]$	$\mathbb{E}[\alpha_0]$	$\mathbb{E}[w_1]$	$\mathbb{E}[\alpha_1]$
Aggregate Effects Only					
Externality-free	248736	3.21$\pm$ 0.00	-	0.44$\pm$0.00	-
Aggregate Externalities	201443	2.57$\pm$0.00	1.32$\pm$0.01	0.39$\pm$ 0.00	-0.09$\pm$ 0.00
Publisher-Dependent Effects					
Externality-free	174768	0.96$\pm$0.02	-	0.87$\pm$0.00	-
Aggregate Externalities	110254	0.85$\pm$0.01	2.22$\pm$0.01	1.13$\pm$0.00	-0.40$\pm$0.00
Publisher-dependent Externalities	106122	2.79$\pm$0.02	1.92$\pm$ 0.01	0.31$\pm$ 0.00	-0.26$\pm$ 0.00

Table 2: Inferred model parameters and model deviance. The *Aggregate Effects Only* models do not include per-publisher coefficients for the latent externality indicator z_i, while the models under *Publisher-Dependent Effects* do include such coefficients. $D(y)$ is the model deviance; w_0 is the log-price intercept; α_0 is the intercept delta when $z_i = 1$ (i.e. in the presence of a pricing externality); w_1 is the log-price slope; and α_1 is the is the externality slope delta. In the publisher-dependent effects case, reported values for the slope and intercept parameters are averaged across all publishers.

Across all models, the base price $w_0 + \alpha_0 z_i$ is significantly higher in the presence of externalities, as $\alpha_0 > 0$. Furthermore, $w_0 + \alpha_0$ in the externality model is greater than w_0 in the baseline model, indicating that additional information has a positive effect on auction revenue, at least for auctions that result in a sale.

The slope coefficient $w_1 + \alpha_1 z_i$ is lower in the externality models, as $\alpha_1 < 0$. This result makes intuitive sense: in auctions where externalities are found to affect bid price, clearing price is less sensitive to floor price (i.e. slope is near 0). In other words, the floor price, or publisher context, is less important as a signal of quality when advertisers have specific information about the particular impression (e.g. *auto-intender*, or *recently bought shoes*).

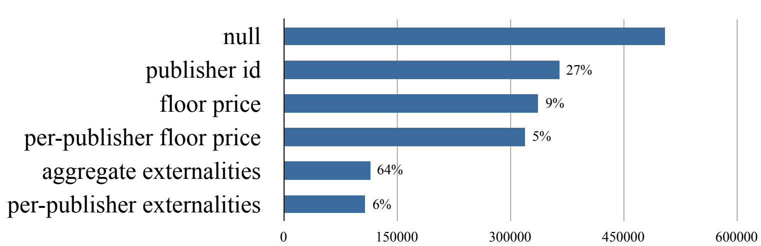

Figure 2: Residual deviance of linear model fit broken down over regression factors in the **publisher-dependent externalities** model. Publisher ID and latent mean-shift components induce the highest absolute reductions in deviance. Percentages show relative reduction in deviance with respect to the previous model.

4 Discussion

We have demonstrated the application of GLMs with a latent mean-shift parameter to quantifying the effects of information externalities and informed bidders on revenue in RTB markets. Such models are potentially useful to publishers interested in predicting the pricing dynamics of their remnant inventory.

4.1 Limitations

The main limitation of the proposed model is that it cannot predict *which* particular auctions are subject to information externalities, rather, it can only capture the aggregate effects on price dispersion retrospectively. However, publishers could potentially capture additional session and user features in order to make such predictions. Our model can then be used to gauge how much price dispersion is captured by such features.

Predictions from this model can be conflated with other causes of price dispersion, such as advertiser budgets fluctuating during the sample period, or seasonal effects on price.[6] In order to model demand-side externalities more accurately, it would be necessary to hold publisher, site context and the advertiser pool constant, observing price variation. However such controlled experiments are untenable in live markets.

4.2 Future work

Demand-side Modeling: Standard models of auctions assume bidders are endowed with their own private values over inventory and the auction clearing price is derived from the this set [cf. 4]. In this paper we have limited ourselves to modeling publisher effects, but could easily extend the analysis to include bidder preferences as well, bringing it more in line with traditional auction theory.

Supply-side Audience Targeting: There is significant market evidence for differential pricing based on audience targeting [cf. 8], and a natural consequent is for similar targeting to take place on the supply-side as well. Such dispersion due to dynamic floor pricing can be captured in our model.

Modeling Sell-through: Predicting sell-through (impressions sold) is also possible in the proposed framework, and is potentially interesting as unsold inventory may have undergone adverse selection due to information externalities.

Censored Models and Optimal Floor Pricing: In order to build models suitable for optimizing floor pricing it is necessary to have an estimate of what advertisers *would have* bid if a floor price were lower. Floor price can be treated as a dynamic *left-censoring*, where auction clearing price is not observed if it is below the floor price. Tobit regression can be used in place of linear regression in the presence of censored variables, and could potentially be used to reconstruct the full bid distribution [1]. Such models also allow straightforward temporal extensions [7].

Models of the full bid distribution would allow publishers to compute the *marginal floor price* and hence derive optimal floor pricing strategies. Theoretically, the optimal floor price is simply the second highest bid price (i.e. the market clearing price). However, in thin (demand-constrained) markets with few bidders and poor price-discovery, the floor price acts as a pseudo-bidder and can improve empirical supply-side revenue [9].

Pricing Risk: Finally, we envision extending our pricing models temporally in order to predict future spot market demand and volatility, key components in controlling publisher risk.

Acknowledgments

Thanks to Metamarkets[7] for making their RTB data available, David Soloff for helpful discussions on ad market microstructure and the anonymous reviewers for their detailed comments and for suggesting an interesting buy-side extension.

[6] We did not attempt to model temporal effects at time-scales longer than 24 hours as our sample period is too short to do do accurately.

[7] `http://www.metamarketsgroup.com`

References

[1] Siddhartha Chib. Bayes inference in the tobit censored regression model. *Journal of Econometrics*, 51(1-2):79–99, 1992.

[2] R. D. Cook and S. Weisberg. *Residuals and Influence in Regression*. Chapman and Hall, 1982.

[3] LLC DeSilva + Phillips. Getting real ad exchanges, RTB, and the future of online advertising. `http://goo.gl/EGXe`.

[4] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[5] Lauren A. Hannah, David M. Blei, and Warren B. Powell. Dirichlet process mixtures of generalized linear models. In *Proc. of Artificial Intelligence and Statistics (AISTATS-10)*, 2010.

[6] Larry Harris. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford Press, 2003.

[7] Lung-fei Lee. Estimation of dynamic and arch tobit models. *Journal of Econometrics*, 92(2):355–390, October 1999.

[8] Mike Nolet. Price floors, second price auctions and market dynamics. `http://goo.gl/PgGR`.

[9] Michael Ostrovsky and Michael Schwarz. Reserve prices in internet advertising auctions: A field experiment. Research papers, Stanford University, Graduate School of Business, 2009.