Online Appendix to "Identification and Estimation of Online Price Competition with an Unknown Number of Firms"

Yonghong An* Michael R. Baye† Yingyao Hu‡ John Morgan§
Matt Shum¶

Abstract

This online Appendix contains Monte Carlo simulations as well as some technical details omitted from the main paper.

*Department of Economics, Texas A&M University, College Station, TX 77845.

†Department of Business Economics & Public Policy, Kelley School of Business, Indiana University, Bloomington, IN 47405.

‡Department of Economics, Johns Hopkins University, Baltimore, MD 21218.

§Haas School of Business and Department of Economics, University of California, Berkeley, CA 94720.

¶Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125.

1 Monte Carlo Evidence

We present results from a simulation study in this section to (1) demonstrate that our estimation procedure performs well in a controlled, small-sample environment, (2) illustrate that failing to account for the unobservability of the potential number of firms can lead to biased estimates of model parameters, and (3) address some additional issues that may arise in empirical applications where limited data are available.

Based on the designed "true" parameter values $\left(\theta^{True}\right)$, we construct a simulated dataset based on the underlying theoretical model as follows. For each simulated period, t, we randomly draw a number of firms for that period, $N_t \in \{2, 3, ..., 15\}$ from a discrete uniform distribution. (The upper bound of this distribution corresponds to the maximum number of listings we observed across all product-dates in the actual data.) Next, we make N_t Bernoulli draws with parameter $\alpha^*(N_t; \theta^{True})$ (defined in Proposition 1 of the main paper) to simulate whether each of these N_t firms list or not. Let $A_t \in \{1, 2, \cdots, 15\}$ denote the number of firms listing prices in simulated period t. For each of these A_t firms, we next draw a listed price from the distribution $F^*(p|N_t; \theta^{True})$ defined in Proposition 1 of the main paper. We repeat this process until we have retained exactly T simulated periods—the sample-size. Our simulations consider sample sizes of $T = 600$ and $T = 1200$.

In addition to examining the performance of our estimation procedure in small sample environments, we also use the simulation to address some additional issues. One such issue is whether controlling for unobservability in the number of firms "matters." As discussed in Section 2 of the main paper, several existing studies of online price dispersion simply assume that $N = A$ in estimation. To evaluate the degree to which controlling for unobservability of N matters, we compare the estimates obtained from our two-step approach with those where $N = A$ is assumed (henceforth, the "naïve" specification). In particular, we use MLE methods, i.e. equation (9) in the main paper to estimate the model parameters θ under the naïve specification.

Another issue addressed by the simulation is the impact of pooling observations with large numbers of participating firms into a single cell of the matrix used in estimation. More specifically, the first step of estimation requires the manipulation of a 14×14 matrix; however, in small sample environments there will be only a few observations where the number of listings is above some threshold, and this may lead to inaccuracy in the matrix decomposition. To examine this, we pool simulated observations for large A, e.g., we combine all the observations with $A \geq 11$ into a single bin to determine its effects on the resulting parameter estimates. While the results of Baye and Morgan (2009) provide a theoretical

Table 1: Monte Carlo Results

Params.	True value	T=600 Two-step estimate (1)	T=600 Two-step estimate (2)	T=600 "Naïve" estimate	T=1200 Two-step estimate (1)	T=1200 Two-step estimate (2)	T=1200 "Naïve" estimate
ϕ	5	5.25	5.42	3.24	5.35	5.29	3.35
		(0.74)	(0.66)	(1.41)	(0.63)	(0.59)	(0.97)
r	300	290.99	287.63	266.97	300.03	308.17	269.74
		(29.60)	(23.16)	(130.25)	(26.81)	(19.62)	(121.39)
m	120	107.04	114.45	135.48	106.31	122.79	136.96
		(5.74)	(12.54)	(35.21)	(6.40)	(13.01)	(37.01)
M	15	15.54	17.39	46.07	15.33	14.80	45.46
		(1.62)	(2.81)	(11.60)	(1.46)	(1.68)	(9.59)
S	10	13.02	8.31	22.01	12.80	12.03	20.88
		(1.07)	(0.99)	(7.75)	(1.41)	(0.72)	(8.79)
γ	0.1	0.09	0.12	0.20	0.09	0.10	0.19
		(0.01)	(0.03)	(0.06)	(0.01)	(0.01)	(0.05)

justification for pooling observations where the number of firms is above some threshold, the impact of pooling for a given threshold is an empirical issue.

More specifically, this approach imposes restrictions on the joint distribution between A and N when both A and N exceed 11, and these may impact the resulting estimates. Thus, we compare the estimated parameters when pooling the data $A \geq 11$ and $A \geq 8$, respectively, as a single matrix cell with the true parameter values used to produce the simulated data.

The simulation results are presented in Table 1, along with standard errors obtained via bootstrapping (with 200 resamples employed). Results in column (1) correspond to estimates where we pool observations consisting of $A \geq 11$ listings into a single bin. Results in column (2) offer estimates when we pool observations where $A \geq 8$ list prices. Finally, the columns labeled "Naïve estimate" correspond to estimates assuming $N = A$. As Table 1 reveals, regardless of the pooling procedure or whether the sample size is 600 or 1200, the parameters are precisely estimated using our two-step procedure, and very close to the true values. By contrast, when we (wrongly) assume the number actual and potential competitors are the same, the resulting "Naïve" estimates do not closely correspond to the true parameters. Specifically, this approach results in estimates that understate the true listing fee (ϕ) and overstate the number of shoppers (S), the conversion rate (γ) as well as the total number of loyals (M). Intuitively, treating the observed number of listings as the actual number of competing firms understates the degree of competition in the market, since, $A \leq N$. Nevertheless, it is an empirical question how the "naïve" approach affects the estimation of parameters and there is no theoretical guidance. On balance, these simulations suggest that failing to account for the unobservability of the potential number of firms can produce distorted estimates of model parameters.

2 Proof of Proposition 1

As in Baye and Morgan (2001), it is readily seen that equilibrium has the following two key properties: (1) A firm must be indifferent between listing its price at the clearinghouse or not; and (2) a firm must earn the same expected payoff from posting any price $p \in [p_0, r]$ at the clearinghouse.

A firm that eschews the comparison site earns profits of

$$\pi_0 = (r - m)\,\gamma \frac{M}{N} + (r - m)\,(1 - \alpha)^{N-1} \frac{\gamma S}{N} \tag{1}$$

A firm that advertises a price r on the site earns

$$\pi = (r - m)\,\gamma \frac{M}{N} + (r - m)\,(1 - \alpha)^{N-1} \gamma S - c\,(1 - \alpha)^{N-1} S - \phi$$

Since firms must be indifferent between listing or not, it then follows that $\pi = \pi_0$. We may use this equality to obtain a closed-form expression for α:

$$(r - m)\,\gamma \frac{M}{N} + (r - m)\,(1 - \alpha)^{N-1} \frac{\gamma S}{N} = (r - m)\,\gamma \frac{M}{N} + (r - m)\,(1 - \alpha)^{N-1} \gamma S - c\,(1 - \alpha)^{N-1} S - \phi$$

Simplifying, this reduces to

$$\phi = (1 - \alpha)^{N-1} S \left((r - m)\,\gamma \frac{N-1}{N} - c \right) \tag{2}$$

Or equivalently,

$$\begin{aligned} (1 - \alpha)^{N-1} &= \frac{\phi}{S\left((r - m)\,\gamma \frac{N-1}{N} - c\right)} \\ &= \frac{N\phi}{S\left((r - m)\,\gamma\,(N - 1) - Nc\right)} \end{aligned}$$

Hence, the equilibrium advertising propensity is:

$$\alpha^* = 1 - \left(\frac{\phi}{S\left((r - m)\,\gamma \frac{N-1}{N} - c\right)} \right)^{\frac{1}{N-1}} \tag{3}$$

The conditions on ϕ and c identified in Proposition 1 in the main paper imply that $\alpha^* \in (0,1)$.

Substituting for α^* in equation (1), we obtain equilibrium profits of :

$$\begin{aligned}\pi_0 &= (r-m)\gamma\frac{M}{N} + (r-m)\frac{\phi}{S\left((r-m)\gamma\frac{N-1}{N} - c\right)}\frac{\gamma S}{N} \\ &= (r-m)\gamma\frac{M}{N} + \frac{\phi}{N\left(1-\frac{c}{(r-m)\gamma}\right) - 1}\end{aligned}$$

It remains to determine the equilibrium distribution of listed prices. Recall that a firm listing a price p, earns expected profits of

$$\pi(p) = (p-m)\gamma\frac{M}{N} + (p-m)(1-\alpha F(p))^{N-1}\gamma S - c(1-\alpha F(p))^{N-1} S - \phi$$

Such a firm must be indifferent between charging p and not advertising at all, i.e. $\pi(p) = \pi_0$. It is convenient to express π_0 in terms of α for the moment. Hence, we have:

$$\begin{aligned}\pi(p) &= (p-m)\gamma\frac{M}{N} + (p-m)(1-\alpha F(p))^{N-1}\gamma S - c(1-\alpha F(p))^{N-1} S - \phi \\ &= (r-m)\gamma\frac{M}{N} + (r-m)(1-\alpha)^{N-1}\frac{\gamma S}{N} = \pi_0\end{aligned}$$

Solving this expression for $(1-\alpha F(p))^{N-1}$, we obtain

$$\begin{aligned}(1-\alpha F(p))^{N-1} &= \frac{(r-p)\gamma\frac{M}{N} + (r-m)(1-\alpha)^{N-1}\frac{\gamma S}{N} + \phi}{S((p-m)\gamma - c)} \\ &= \frac{(r-p)\gamma\frac{M}{N} + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1)-Nc}\phi}{S((p-m)\gamma - c)}\end{aligned}$$

which implies

$$F(p) = \frac{1}{\alpha}\left(1 - \left(\frac{(r-p)\gamma\frac{M}{N} + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1)-Nc}\phi}{S((p-m)\gamma - c)}\right)^{\frac{1}{N-1}}\right)$$

To verify that $F(p)$ is a well-defined atomless probability distribution, we will first show that $F(r) = 1$, or equivalently, $(1-\alpha F(r))^{N-1} = (1-\alpha)^{N-1}$. To see this, note that

$$\begin{aligned}(1-\alpha F(r))^{N-1} &= \frac{\frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1)-Nc}\phi}{S((r-m)\gamma - c)} \\ &= \frac{N\phi}{S((r-m)\gamma(N-1) - Nc)} \\ &= (1-\alpha)^{N-1}\end{aligned}$$

where α is defined in equation (3).

Next, we determine the lower support of the equilibrium listed price distribution; that is p_0, where $F(p_0) = 0$. Equivalently, p_0 satisfies $(1 - \alpha F(p_0))^{N-1} = 1$, or

$$\frac{(r - p_0)\gamma \frac{M}{N} + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1) - Nc}\phi}{S((p_0 - m)\gamma - c)} = 1$$

Cross-multiplying and collecting the p_0 terms:

$$\frac{\gamma r M}{N} + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1) - Nc}\phi + S\gamma m + Sc = p_0\left(S\gamma + \frac{M}{N}\gamma\right)$$

Solving for p_0 gives

$$p_0 = m + \frac{1}{\left(S\gamma + \frac{M}{N}\gamma\right)}\left(\gamma\frac{M}{N}(r-m) + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1) - Nc}\phi + Sc\right)$$

which exceeds m.

Finally, we verify that F is strictly increasing, or equivalently, that $(1 - \alpha F(p))^{N-1}$ is strictly decreasing in p. Recall that

$$(1 - \alpha F(p))^{N-1} = \frac{(r - p)\gamma \frac{M}{N} + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1) - Nc}\phi}{S((p - m)\gamma - c)}$$

and define $num \equiv (r-p)\gamma\frac{M}{N} + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1) - Nc}\phi > 0$ and $den \equiv S((p-m)\gamma - c) > 0$. Differentiating with respect to p reveals

$$\frac{\partial(1 - \alpha F(p))^{N-1}}{\partial p} = -\frac{\gamma\frac{M}{N}(den) + S\gamma(num)}{(den)^2} < 0.$$

One can enrich the model to allow some shoppers not to click on any offer, perhaps owing to finding the listed prices or products unacceptable. Likewise allowing conversions to decrease with the price can also be accommodated. This would result in firm demand per shopper being some function $I_{p=p\min}\delta(p)\gamma(p)$, where I is an indicator function and $\delta(p)$ represents the probability of not clicking as a (decreasing) function of p, and similarly for loyals. This formulation is isomorphic to a setting where consumers have downward sloping demand, as in Baye and Morgan (2001). The qualitative characteristics of the resulting distribution of prices and the probability of advertising on the comparison site are identical to the simpler model we study. Thus, we opt for the simpler specification for purposes of estimation.

References

Baye MR, Morgan J. 2001. Information gatekeepers on the internet and the competitiveness of homogeneous product markets. *American Economic Review* : 454–474.

Baye MR, Morgan J. 2009. Brand and price advertising in online markets. *Management Science* **55**: 1139–1151.