Explaining Price Dispersion in the Manhattan Beer Market:
Information Acquisition, Geographic Distribution and Searching under the Influence

Vandana Ramachandran
University of Utah
vandana@business.utah.edu

Arun Sundararajan
NYU Stern School of Business
arun@stern.nyu.edu

1. Overview and Summary

The term *price dispersion* is commonly used to refer to firms in the same market selling identical goods for different prices at the same time. Since the early conjecture by Bakos (1997) that electronic markets are likely to feature lower price dispersion on account of lower search costs, a number of researchers have studied how information technology alters the observed variation of prices across sellers, in both traditional and online marketplaces.

We model the role of geographic concentration, online versus offline information acquisition, heterogeneous search costs and repeat purchasing in explaining price dispersion for homogeneous goods, and examine the drivers of measured price dispersion in the market for beer served at New York City bars. We formulate a *random bar crawler* model of search by bar patrons, a variant on the random surfer model of Brin and Page (1998), and use it to generate hypotheses about the equilibrium price dispersion of different brands of beer across bars dispersed in densely populated urban areas. We use panel data comprising price information in 2009 for over 4,200 beers sold in over 600 bars in New York City across neighborhoods with differing income levels and varying mixes of regular patrons/tourists. We combine this with information about the geographic concentration of bars and proximity to public transportation obtained from Google Maps, product quality information from beeradvocate.com, merchant desirability information based on reviews from Citysearch.com, urbanspoon.com and insiderpages.com, and product cost information from the New York State Liquor Authority.

Our preliminary evidence suggests that the significant price dispersion that exists in this market persists even after controlling for differences in a variety of bar characteristics and variation in the intensity of local competition as measured by neighborhood bar concentration. Our analysis has also uncovered patterns of price dispersion across products in different neighborhoods which is consistent with a theory in which in addition to simple random bar crawling, heterogeneous benefits are realized from the acquisition of price information, consistent with an economic model in which a fraction of buyers are "one-shot" consumers while others are "regulars" or repeat purchasers

2. Background and Motivation

Our study is motivated in part by the fact that information technologies affect price dispersion and its analysis in a number of different ways[1]. The model of Bakos (1997) predicts that the intermediation facilitated by electronic markets lowers the cost of buyers in acquiring information about both seller prices and product offerings, which in turn will lower price dispersion. There is a wide variety of evidence that supports this theory to differing degrees for different homogenous products. In the market for airline tickets in well-defined markets, for example, price dispersion, while substantial in the mid-1990's (Clemons, Hann and Hitt, 2002) has dropped since, and there is evidence that IT has led to more transparent electronic markets in this industry (Granados, Gupta and Kauffman, 2007). On the other hand, price dispersion for books, CDs and software, initially measured as substantial in the 1990's (Brynjolfsson and Smith, 2000) has persisted ten years later (Ghose, Ipeirotis and Sundararajan, 2009) in part because of observable and measurable heterogeneity in the reliability of the sellers of these products and of the quality of the trade processes (fulfillment, delivery, packaging, customer support) associated with the "purchase". The latter findings are relevant to our study because they suggest that price dispersion for a seeming homogeneous product (such as beer) may be mitigated by controlling for seller fixed effects, a finding supported partially in the past for the gasoline market as well (Lewis, 2008).

A different impact of IT on price dispersion that has received less attention in the literature and that we aim to contribute towards is the effect that online providers price and product quality information may have on price dispersion in offline markets. In our offline market, there is currently a wealth of web-based information about the quality of different merchants (bars and other drinking establishments) which is bound to affect the variation of prices for homogeneous products sold by heterogeneous merchants. A third impact that IT has had on the analysis of price dispersion is in the wealth of data that it has made available to the educated econometrician. Our study is made possible by the Internet-based availability of beer prices[2] as well as our ability to quantify the intensity of competition and the local concentration of sellers using the Google Maps API.

3. Data and Model

We gather data about beer prices in New York City from *beermenus.com*, a web site that lists the menus, prices and ABV ratings of beers sold in bars, pubs and other drinking establishments

[1] It is also motivated in part by a deep personal interest by one of the authors in this local industry coupled with extensive prior consumption of its products and services.

[2] Thus eliminating the need for expensive (although possibly fulfilling) hand-collection of pricing data.

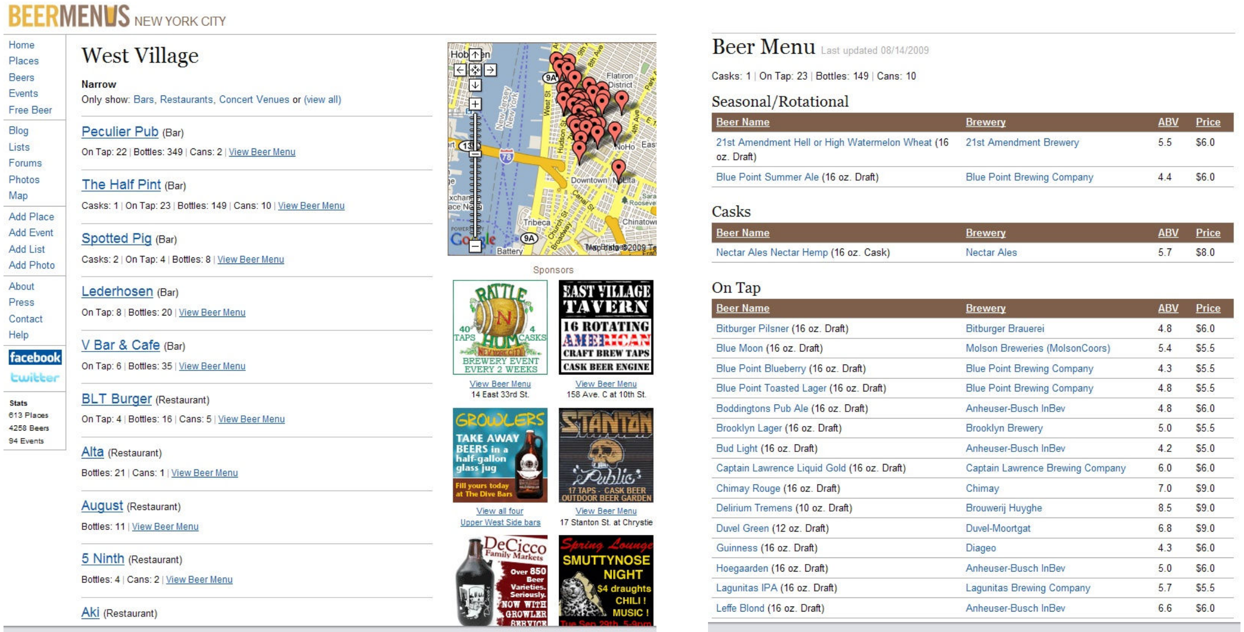

Figure 1: Illustrates the beermenus.com web site. The screenshot on the left lists a subset of the bars in a particular Manhattan neighborhood, while the screenshot on the right illustrates how the price and product information are displayed to a user.

(henceforth referred to as "bars" for brevity). We use a Perl spider and custom-built parser to traverse the site on a weekly basis. Our data collection began in March 2009 and is ongoing. The site also provides information about the street address (geographic location) of each bar. We use the Google Maps API to determine the latitude and longitude of each bar, and to compute the walking distances between bars in the same neighborhood. We use the Google Local Search API to get aggregate rating information for bars, as well as custom Perl scripts to spider *insiderpages.com*, *Citysearch.com* and *urbanspoon.com* for further bar rating information. We supplement this with product information gathered from *beeradvocate.com*.

Our empirical analysis is based on two models. The first builds on Sorenson (2001) and bases the eventual price dispersion on a consumer utility specification for consumption that has beer-specific, bar-specific and idiosyncratic components. Specifically, the indirect utility for consumer *i* from consuming beer *k* at bar *j* is given by

$$u_{ijk} = \theta_j - \alpha p_{jk} + \varepsilon_{ijk},$$

where θ_j is a bar fixed effect which captures the average "ambience" of the establishment or other bar-specific characteristics that may make it able to charge a higher price, p_{jk} is the price of beer *k* at bar *j* and ε_{ijk} represents preference aspects relating to bar *j* and beer *k* that are idiosyncratic[3] to individual *i*.

[3] For example, an explicit preference for 16oz. Stella Artois draft beer at the V-Bar.

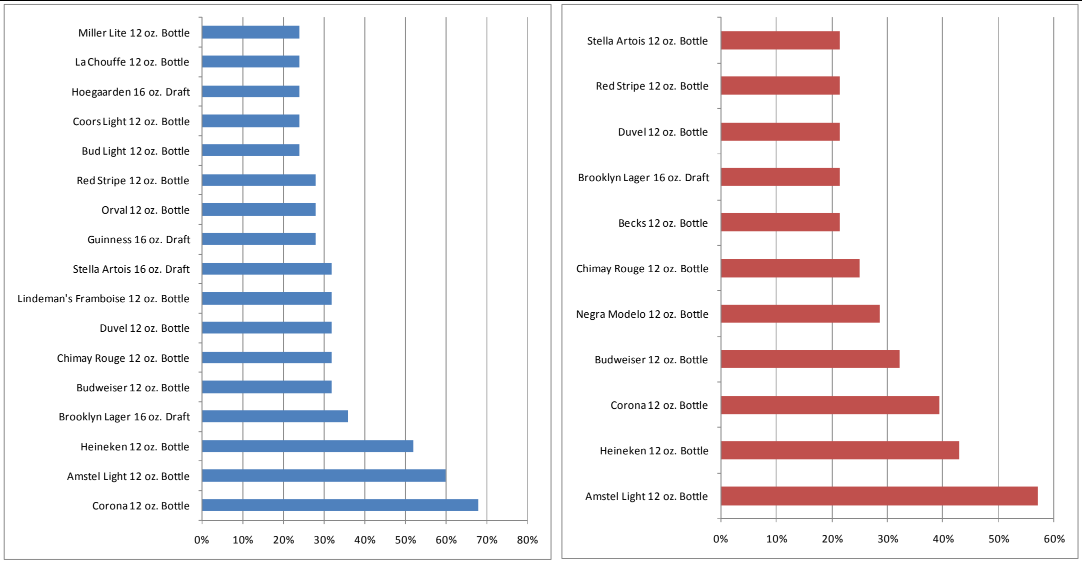

Figure 2: Illustrates the most popular beers (measured as a percentage of bars that they are available at) in two zip codes (10012, blue on the left, and 10014, red on the right).

We also assume that search is "all or nothing". Consumers either price-search exhaustively or select the bar that best matches their non-price preferences, expecting to pay the average price for the beer of their choice. The latter consumers, who we term "one-shot" consumers[4], will maximize the expected utility obtained absent a price search. The former consumers compute the maximum expected indirect utility from price-shopping, and are those whose indirect utility (amplified by an inflation factor alc_{ik} which represents the number of times the consumer expects to use the information they have acquired via this exhaustive search) exceeds their cost of exhaustive search τ_{ik}.

The second model, which is still under development, is a dynamic model of search by uninformed consumers, a model we call the "random bar crawler" model. A consumer starts by choosing a bar at random, where the probability of starting at any given bar is proportionate to the bar's exogenously specified quality (determined, for example, from some composite of its user-generated and expert ratings). Upon arriving at the bar, the consumer determines its price menu and true "ambience" by direct examination. The consumer then chooses whether to consume a beer at this bar or explore a second bar. If the choice is the latter, the probability that the consumer chooses a specific next bar in the neighborhood of the current bar is inversely proportionate to the distance of the former bar from the latter. When reaching the next bar, the consumer repeats this process. This process describes a Markov chain in which each bar location

[4] Or "one pint" consumers, perhaps.

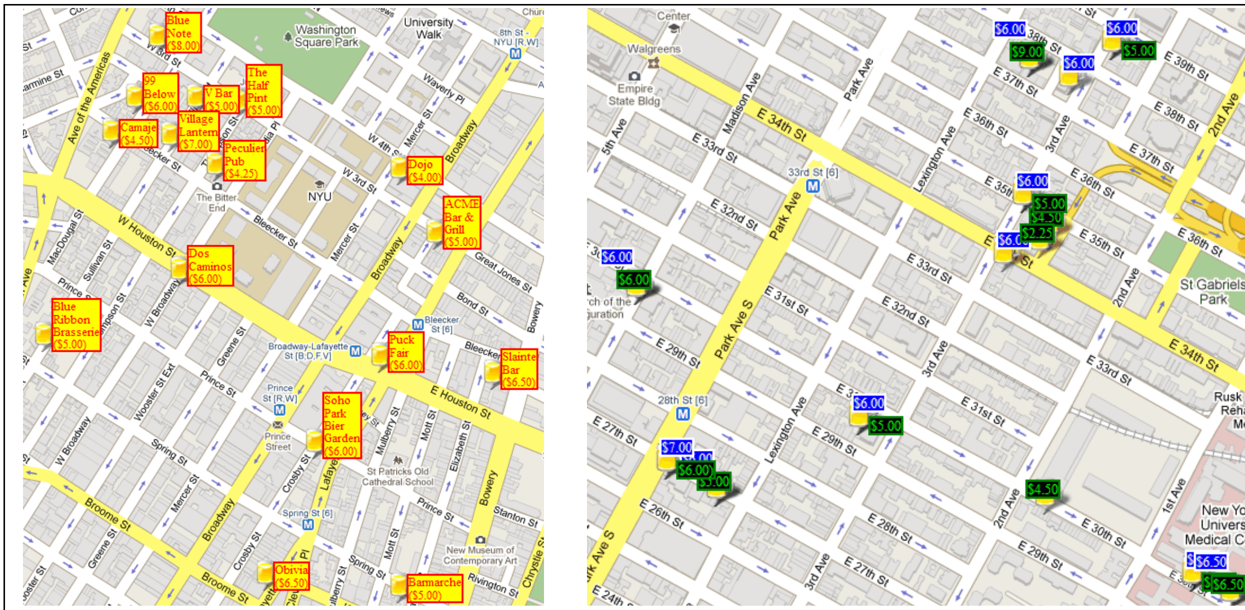

Figure 3: Illustrates (a, on the left) price dispersion in a specific zip code for Corona 12oz. bottles, and (b, on the right for zip code 10016) how the price dispersion for different beers at a sample of 14 locations, 10 of which sell both beers (blue labels are for Amstel Light 12oz. bottle, green labels are for Budweiser 12oz. bottle) is not completely explained by seller-specific characteristics.

is a state, and the transition probabilities are based on inter-bar distances[5]. We can then compute the steady-state probabilities of being at the "states" (bars) in this Markov chain. These steady-state probabilities, analogous to PageRank values, represent a measure of the pricing power of the bar in question as a function of both its geographic location and inherent quality.

4. Preliminary Results and Research Agenda

Our salient preliminary results are summarized below.

Heterogeneous preferences across neighborhoods: This is a pattern that emerges from our data, that there is considerable variation in the "popular" beers across different areas of Manhattan, indicative of heterogeneity in consumer tastes. The bar charts in Figure 2 illustrate this by charting the distribution of availability of specific beers in two candidate zip codes in our data set. As is evident from the figures, the different zip codes have fairly different distributions of popular beers. This could depend in part by the mix of tourists and regulars.

---

[5] A more complete and realistic model would recognize the fact that transition probabilities from bar to bar are a function of the number of beers consumed at the point of transition (and after a particular consumption level some states may become sinks), but we have not yet been able to formulate an analytically tractable version of this model.

Price dispersion is persistent even after accounting for bar quality: Figure 3a illustrates the price distribution[6] for a sample beer in a subset of bars in a candidate zip code (10012) from our data set. As is evident from the map on the left, there is considerable price dispersion across bars even for a very popular homogenous beer. Figure 3b illustrates the price dispersion for two popular beers in a different zip code, illustrating that bar (seller) fixed effects do not wash out all the variation in prices across bars. This is another pattern that emerges very clearly from our data.

Concentration and price dispersion: We have defined a measure of competition at both the bar level and the individual beer level based on our Google Maps data about geographic concentration. The simple measure of concentration counts the number of bars that are within a reasonable walking distance of each bar (for example, a bar with 85 other bars within walking distance is termed an "85-proof" bar). The measure of product concentration counts, for each beer sold at a bar, the number of other bars within walking distance that also sell that beer. Neither measure of concentration fully accounts for the observed price dispersion. This analysis is quite preliminary and we are working on refining these estimates.

Our further research agenda between now and December 2009 will focus on completing the development of our analytical model of random crawling, computing the relative market power of each of the bars in our sample using this model, and assessing whether this more refined measure of concentration can account for the price dispersion we observe in our data. We are also working on completing a more robust estimation of our model with one-shot and regular customers using our entire weekly panel, perhaps extending it to include a more sophisticated analytical model of consumer search. We hope to be able to report on our findings at the 20th anniversary WISE in Arizona.

(References available on request)

[6]A different kind of "bar chart".