IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP5269

This paper appears in *Managing Modern Organizations Through Information Technology*, Proceedings of the 2005 Information Resources Management Association International Conference, edited by Mehdi Khosrow-Pour.

Shopbot Market Coverage

Gove Allen and Jianan Wu

A.B. Freeman School of Business, Tulane University, 7 McAlister Dr., New Orleans, LA 70118, USA, {gallen, jwu1@tulane.edu}

The increased availability of pricing data at shopbot websites has resulted in a proliferation of empirical researches on pricing behavior in electronic markets. The data obtained from these shopbots have been extensively used to make inferences on a variety of vendor pricing behaviors in many product markets. Some such studies have used multiple shopbots to increase market coverage (Kauffmann and Wood 2000, Clay et al. 2001). However, because the sampling process is not random, the effect that increasing coverage has on sample representativeness is unclear. As shopbots become popular data agents for making inferences on Internet pricing behavior, it is important to question: "To what degree do individual shopbots represent a market being studied?"

To answer the question rigorously, we present a large scale study in which we tracked 459 books offered for sale by 84 online vendors as reported by eight shopbots for a four month time period (between March and July of 2002, totaling 2.2 million price observations. The books could be categorized as "Bestsellers" (hardcover and paperback), "Classics," and "New Releases". Bestsellers were selected from the New York Times best seller list, classic books (whose text is now public domain, such as *A Tale of Two Cities* and *The Count of Monte Crsito*) were selected from litrix.com and new releases were selected from the "Coming Soon" area of randomhouse.com.

Since the selection of shopbots to adequately represent a market is the very question that this study examines, we employ a heuristic selection technique. This approach features two heuristics to compensate the lack of prior knowledge on this issue. Heuristic #1: The more shopbots, the better. Theoretically, if we assume that data reported at each shopbot represents an independent sample of the real market, then the union of the data reported at all shopbots should represent a sample with smaller sample variance. Ideally, every active shopbot that provides information about any of the products in the basket of products should be included. However, the dynamic nature of the shopbot arena makes attaining the ideal only remotely possible and the conclusive demonstration that the ideal has been achieved virtually impossible. Heuristic #2: The more diversification, the better. This set should include both shopbots that try to cover a wide array of products (better economy of scale) as well as those that focus on a narrow market (better niche knowledge). We included eight shopbots in our study: Addall, Bizrate, Dealtime, ISBN, MySimon, PriceGrabber, PriceScan, and Yahoo. This set of shopbots was identified through a web search using the Google search engine.

To collect the data over the four-month study period, we developed distributed internet data retrieval agent, named "Synchronous Electronic Commerce Research Engine and Toolkit" (SECRET) Agent which can collect both longitudinal and cross sectional data from publicly available web sites. Its distributed nature allows for very rapid parsing and storage of hypertext data. The configuration used for the current study consisted of two servers and 12 multithreaded nodes, processing more than 1,000 web pages per minute, archiving approximately 600 megabytes of data per hour. The agent produced a data set of over 2.2 million product/price observations in the online book market.

We examine the shopbots along two important dimensions: 1) How well does a shopbot represent *vendors' products* (vendor coverage)? Vendor coverage is selected because it has been a major concern on search engines (e.g., Bradlow and Schmittlein 2000). 2) How well does a shopbot represent *vendors' prices* (market representativeness)? Vendor price is selected because it has been a major data source for testing pricing behavior in e-commerce (e.g., Kauffmann and Wood 2000, Ellison and Ellison 2001, Baye 2001, Baylis and Perloff 2002),.

Table 1 shows the shopbots' representations of vendor prices and vendor products for the top 20 vendors in terms of number of products offered. On shopbots' representation of vendor products, several observations can be made. First, shopbots do not cover the vendors equally. For example, PriceScan covers the top 20 vendors with average coverage 70%, but Dealtime achieves an average coverage of the same vendors of only 16%. Second, even if the shopbots cover vendors equally well on average, they do not cover the same vendors. Of the top 20 vendors, Yahoo reports prices from seven and Bizrate reports prices from eight; however, of those, the two shopbots hold only four vendors in common. Third, it seems that certain affiliations between shopbots and vendors exist. Although amazon.com is clearly a dominant player in the market, it is ignored by both Yahoo and Pricegrabber.

Similar observations can be made on shopbots' market representativeness. First, some vendors consistently charge higher (or lower) prices than the average market prices. For example, doublediscount.com and AllDirect.com charge lower than average market price for over 90% of the books in our data set while premierbooks.com, powells.com and varsitybooks.com offer books at higher than average market price for over 90%. Second, vendors with higher awareness do not necessarily charge higher prices. For example, Barnes & Noble and Amazon offer 64% and 42% of the books above the average market price. On average, Barns & Noble seems to offer slightly higher price ($13.43) than Amazon ($12.41).

Although more research is needed to determine the extent to which these differences in market coverage affect the way in which shopbot data categorizes a market, it is clear that different shopbots represent the same market substantially differently. Future research should also consider what effect collecting data from multiple shopbots simultaneously has on market representativeness.

REFERENCES

Baye, M., J. Morgan. 2001. Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets. *American Economic Review*, 91 (3), 454-474

Baylis K., J. Perloff. 2002. Price Dispersion on the Internet: Good Firms and Bad Firms. Working Paper, UC Berkeley.

Bradlow, E., D. C. Schmittlein. 2000. The Little Engines That Could: Modeling the Performance of World Wide Web Search Engines. *Marketing Science*. 19 (1), 43-62.

Clay, K., R. Krishnan, E. Wolff. 2001. Prices And Price Dispersion On The Web: Evidence From The Online Book Industry. *The Journal of Industrial Economics*. 49 (4), 521-539.

Ellison, G., S. F. Ellison. 2001. Search, Obfuscation, and Price Elasticities on the Internet. Working Paper, MIT.

Kaufmann, R., C. Wood. 2000. Follow the Leader? Strategic Pricing in Ecommerce. *The Proceedings of the International Conference on Information Systems 2000*. Brisbane, Australia.

Table 1. Shopbots' Representations of Vendor Products and Vendor Prices (Top 20 Vendors)

Vendor	# of Books Offered	Shopbots' Representation of Vendor Products								Vendor Average Price	Vendor Books > Mean Price
		Addall	Bizrate	Deal-time	ISBN	My-Simon	Price-Grabber	Price-Scan	Yahoo		
Bn.com	446	348	256	160	275	353	265	356	408	$13.43	64%
alphacraze.com	443	238	223		212		225	442		$10.66	1%
wordsworth.com	442					309		441		$13.65	77%
fatbrain.com	439	439		126	329	439		333		$12.85	61%
1bookstreet.com	407	281	241	162		287	251	297	384	$14.64	86%
amazon.com	405	340	268	157	329	350		368		$12.41	42%
doublediscount.com	400		253			283		393		$11.67	3%
powells.com	392	239	177	113	202		198	282	373	$12.53	91%
textbookx.com	387	258	225	153		260	251	282	348	$12.07	6%
alldirect.com	379					263		273	318	$11.15	2%
booksamillion.com	374	334	273		81		19	369		$11.95	35%
ecampus.com	358	270		164	141	270	41	268	302	$11.44	3%
bookvariety.com	350			143		331		338		$13.86	59%
varsitybooks.com	350					324		349		$15.68	91%
a1books.com	348	288				298		347		$13.65	50%
faithpoint.com	340					340				$11.72	34%
alibris.com	339	328		52						$10.97	45%
premierbooks.com	331								331	$38.71	99%
page1book.com	314					314	167	277		$12.70	30%
indigo.ca	309	309								$14.29	71%
Shopbot Average Representation		48%	24%	16%	19%	58%	18%	70%	32%		
Shopbot % Books > Mean Price		47%	40%	49%	47%	45%	45%	44%	52%		

# of Books Offered is the number of books identified by at least one of the eight shopbots in the time window. For example, 446 books were offered by bn.com in the basket of 459 selected books in our study. **Shopbots' Representation of Vendor Products** is the number of the books identified by the shopbot at a vendor. For example, Addall represents 348 books offered at bn.com. However, for meaningful comparison on prices, we consider only those books (179 in total) each of which is identified at each of the top 20 vendors for at least one day during our sampling window in calculating the price related measures. **Vendor Average Representation** is the average price across products at a specific vendor. For example, $13.43 is the average price of all the 446 books offered at bn.com. **Vendor % Books > Mean Price** is the proportion of the books offered at prices higher than the average market prices at a specific vendor. For example, 64% of books offered at bn.com are priced above the average market prices. **Shopbot Average Representation** is the average proportion of the books identified by a shopbot. For example, 48% of the books offered are identified by Addall. **Shopbot % Books > Mean Price** is the weighted average of **Vendor Books > Mean Price** with **shopbots' Representation of Vendor Products** as weights. For example, 47% of the books identified by Addall is priced above the average market prices.