Comparison Shopping Agents and Online Price Dispersion: A Search Cost based Explanation

Bhavik K. Pathak

Indiana University South Bend, School of Business and Economics, South Bend, USA, bkpathak@iusb.edu

Abstract

Search costs and consumer heterogeneity are two important explanations for the price dispersion in the brick and mortar (B&M) markets. Comparison shopping agents (CSAs) provide a single click decision support for consumers' purchasing related decision problems and reduce their search costs by providing detail price dispersion related information. Contemporary researchers in IS observe that even with such negligible search costs, price dispersion still continues in the online markets. Consumer heterogeneity and retailer heterogeneity have been agreed upon as two primary explanations for online price dispersions. In this paper, popular CSAs are analyzed to check if they provide complete and accurate price dispersion information. It is shown that because of the selection bias and temporal delay in updating information, contemporary CSAs may not present complete and accurate price dispersion information. In order to reach to an optimal purchasing decision, consumers may have to rely on a sequential search across multiple CSAs or browse through various retailers. This research adds a search cost dimension to explain the continuance of price dispersion in the online markets.

1   Introduction

The advent of the Internet has provided sellers a cost effective platform to extend their reach beyond any geographical or temporal barriers. Moreover, because of the centralized inventory, online retailers have expanded their product portfolio and embraced long tail phenomenon. The shoppers' purchasing decision has become complex because of ever increasing number of options in terms of sellers and product varieties. Alternate business models of information intermediaries have emerged that provide detail information to shoppers at minimum level of efforts. Comparison shopping agents are such infomediaries that facilitate detail seller and product related information to shoppers. Researchers in information systems (IS), marketing, and economics have been intrigued by the role these CSAs play in reducing price dispersion in the online markets.

Analysis of ongoing price dispersion in online markets, even in the presence of CSAs, is an interesting problem and researches in marketing, economics, and information systems have studied this issue. Prior literature discusses consumer heterogeneity and search costs based justification for the existence of price dispersion in the brick and mortar markets [33], [34], [42]. The CSAs provide detail price and retailer comparison based on the product information supplied by the shoppers. This reduces the search costs associated with determining price quotes from various retailers and hence should remove the factors that tend to facilitate price dispersion in the market. However, researchers have found that just like their B&M counterparts, the online retail market also displays spatial and temporal price dispersions [2], [7], [10], [15], [35]. Much of the prior research acknowledges the role of CSAs in reducing search costs and provides non-search costs based explanations for the existence of price dispersion in online markets. Previous studies have provided consumer heterogeneity, seller heterogeneity, and price discrimination as potential factors that contribute to online price dispersion. The underlying presumption in these studies is that because of CSAs, the impact of search costs is negligible in online shopping.

This research studies the online price dispersion problem and analyzes the role of search costs in explaining online price dispersion. One click price comparison provided by CSAs can potentially remove search costs associated with comparison shopping. However, in this paper, I study whether comparison shopping information provided by CSAs is accurate and complete. If such information is incomplete and inaccurate then shoppers may have to conduct additional search costs. Thus, the objective of this research is to study and explain the search costs based justifications for online price dispersion. There are two major reasons that CSAs may provide incomplete and inaccurate price dispersion information: temporal delay and selection bias.

 In the past, CSAs mainly relied on the real time Web-scrapping to obtain pricing data from online retailers [27]. This data collection methodology has some major problems including query response time [27], data quality and merchant obfuscation [17], and merchant blocking [16]. Advances in information technology has created alternate means of sharing or collecting data (e.g. XML, FTP) and today's CSAs mainly rely on obtaining pricing data directly from the retailers. Although, due to lower menu costs, online sellers tend to make more SKU level price changes [1], [25] and sellers may provide price feed data to CSAs at discrete time periods rather than synchronize their price changes with the CSAs in real time. Thus, at any given point of time, CSAs may not present accurate price dispersion due to such temporal lag in updating price information.

Moreover, majority of CSAs provide free comparison shopping services to shoppers and primarily rely on listing and click-through fees from online retailers for their revenue. They charge various fees such as referral fees, merchant storefront fees, search result ordering fees, advertisement charges, revenue sharing programs and, listing fees to online retailers. Every CSA has its own pricing structure. For example, shopping.com's referral fees in cost-per-click (CPC) ranges from $0.05 to $0.95 depending on a product category. For some other CSAs, such as bizrate.com, retailers have to decide how much they want to pay for the CPC charges. The more they pay, the higher their listing appears in the search results. Typically, the selection and ordering of online retailers in the search results of CSA is based on competitive bids. For the same product, because of such selection bias, spatial price dispersion may differ from one CSA to another. Such economic incentive-based selection bias of CSAs may restrict the merchant participation in their programs. Hence, the merchant-supported revenue model of the CSAs may not present complete price dispersion to shoppers and may lead towards suboptimal purchasing decisions.

In the absence of complete and accurate price dispersion information, shoppers may pursue their search across multiple CSAs or even various online merchants. In this paper, these two phenomena of temporal delay and selection bias are analyzed to provide search cost based explanations for the online price dispersion. The results show a significant delay in price update at the online CSAs across different retailers and products. It clearly shows that participating sellers differ significantly from one CSA to another and hence selection bias is evident. Due to selection bias and temporal delay in update, CSAs may present incomplete and inaccurate price dispersion information and hence may not optimize shoppers' purchasing decisions without comprehensive search across multiple CSAs or online retailers. Prior researchers have provided consumer heterogeneity, retailer heterogeneity, obfuscation, and price discrimination based justifications for online price dispersion. One of the major contributions of this paper is to add search costs based factor in explaining online price dispersion.

In this paper, the phenomenon of temporal delay and selection bias and its impact on online price dispersion is analyzed. The rest of this paper goes as follows. Section 2 presents theory and hypotheses. Section 3 provides data collection approach. Results and analysis are provided in section 4 and section 5 concludes this paper with the discussion, research implications and limitations.

2   Theory and Hypotheses

Price dispersion has been of high concern for both online as well as B&M retailers. In the B&M case, it has been shown that the price dispersion can exist when firms have perfect information about buyers' reservation prices and demand functions [33]. Prior researchers discuss that the information acquisition is costly and hence it has an impact on market equilibrium. It has been shown that consumers differ in their costs of information acquisition. In such cases, firms, as discriminating monopolists, exercise price discrimination. However, the scope of such price discrimination is significantly restricted in a competitive market with perfect information [34].

The online market is considered as a competitive market with a very high degree of information transparency. The impact of practicing price dispersion on the online retailers may be significantly higher given the emergence of the CSAs, where consumers can compare prices across multiple retailers just by a few clicks. Prior researchers in various fields have extensively studied CSAs on various issues including price dispersion [5], [6], [8], [10], [14], [15], consumer behavior [11], [22], [32], [38], [42], information search costs [3], [24], [41]-[43], and recommender systems [20], [23], [45]. The focus of this research is on price dispersion. A brief survey about CSAs has been published by Pan *et al.* [29]. They observe that the majority of studies in this area have recorded substantial price dispersion in the online markets and in general online price dispersion is not smaller than the traditional markets. Moreover, they conclude that over a period of time, the online price dispersion has declined, but continues to be substantial. They suggest that multichannel retailers generally charge higher prices and they are an important source for price dispersion in online markets. Other researchers also show the existence of the price dispersion even after controlling for the shipping costs in the online markets [4].

Although the emergence of CSAs has made it easy for the consumers to obtain price quotes from various merchants, price dispersion still exists in the online markets. The law of one price does not prevail in the online markets. Prior studies have provided heterogeneity based reasons for the continuing price dispersion in the online markets. Retailer heterogeneity has been described as one of the major explanations for the price dispersion in online markets [10]. Retailers can be differentiated based on shopping experience, shipping, return policy, and other related services. Online tools and services such as Amazon.com's recommender systems and prime shipping (i.e. a subscription based service that provide free two-day delivery) create switching costs for consumers and hence contribute towards online price dispersion. Other major explanation for price dispersion is the consumer heterogeneity. Consumers may differ in terms of their brand loyalty [10] and the knowledge of prevailing pricing in the market [34], [42]. Other researchers [13] show that when consumers are brand sensitive, retailers may adopt asymmetric mix pricing strategy wherein they have higher prices on average for majority of products but lower prices on some of the products. Smith [36] concludes that unlike B&M retailers where consumers make their purchasing decisions based on proximity of retailers, online retailers are more likely to be affected by brand name recall. Some prior studies have addressed the issue of service heterogeneity. Varian [42] predicts that online retailers will fall into two main categories: great service and higher prices, average service and lower prices. Service heterogeneity based price dispersion has also been empirically validated in prior studies [30]. Price formats (e.g. EDLP Vs Hi/Lo) have also been identified as one of the potential sources for price dispersion [12]. Search costs based explanation has been discussed as a major source of price dispersion in B&M markets. While the majority of these studies have focused on non-search cost based explanations for price dispersion, there is very little, if any, work done in explaining the search cost-led online price dispersion.

CSAs provide an easy access for comparing product and price information across multiple retailers. Consumers may make an optimal purchasing decision based on this information only if the information provided on these CSAs is complete and accurate. If the price dispersion information provided by CSAs is incomplete and inaccurate, consumers may have to perform an additional search in order to make informed and optimal purchasing decisions. These additional search costs need to be accounted for and can be one of the potential sources of online price dispersion. There are two specific characteristics of CSAs that creates such possibility of additional search costs.

2.1   Temporal Delay in Updating Price Information

Prior research in economics and marketing field has focused both on spatial and temporal price dispersion. Hal Varian was among the first few who described and studied temporal price dispersion. Temporal price dispersion is briefly characterized as follows.

"In a market exhibiting temporal price dispersion, we would see each store varying its price over time. At any moment, a cross section of the market would exhibit price dispersion; but because of the intentional fluctuations in price, consumers cannot learn by experience about stores that consistently have low prices, and hence price dispersion may be expected to persist [42]."

Temporal price dispersion has also been studied in the context of price rigidity. Price rigidity is the proposition that some prices change slowly in response to the market dynamics. Levy [25] discusses that price rigidity varies from market to market and this variation depends on the magnitude of menu costs. Menu costs are price adjustment costs. In the case of traditional retail markets, where menu costs are quite high, retailers have fewer incentives to adjust their prices in accordance with the market dynamics. However, menu costs for online retailers are very low and they may change their prices more frequently in response to the market dynamics [25]. Price change decisions also depend on the market structure and prices are adjusted more frequently in markets with a large number of competitors. In online markets, especially in computer and electronics products, generally one observes a very large number of competitors and hence it is reasonable to estimate frequent price changes in the online markets. Further, these changes are irregular, random, or unpredictable and hence both consumers and competitors cannot learn about the stores that consistently sell at the lower prices [7], [42]. Temporal price dispersion is also described as "hit and run pricing" and it has been shown that online retailers employ these kinds of pricing strategies [6].

One of the major implications of the temporal price dispersion, hit and run pricing, lower menu costs or price rigidity is that level of prices are unpredictable at any instant of time. CSAs present the price comparison results for the consumer's product search query and in order to maximize consumer utilities, it is extremely important for the CSAs to provide the most accurate price information. In order to provide such accuracy, ideally CSAs should query all online sellers at the time of consumer's search requests. In the past, majority of CSAs used to collect price data from the websites of the online retailers in real time. For each user query, CSAs would initiate a search across multiple merchants to download and parse web pages and finally present results to the user [27], [36]. Although this approach was popular because of its merchant independence, it has some major limitations including data quality and merchant blocking problem [40] and obfuscation and bait-and-switch strategies by merchants [17], [36]. Majority of CSAs collect data directly from retailers by using Web interfaces, Web services, APIs, or XML [18]. This merchant-dependent approach has two major issues. First, it brings merchant-bias in CSA-based search results [26]. Second, and more importantly, data may get collected or updated at discrete intervals which may lead towards temporal delay in updating price. This leads to the first hypothesis.

Hypothesis 1: There is a substantial temporal delay between online price adjustments by retailers and updating of such price adjustments on CSAs.

It is important to realize that due to this temporal delay, CSAs may present inaccurate price information to consumers which may result in suboptimal purchase decisions by consumers. The probability of such suboptimal decisions increases with the increase in temporal delay. Once consumers realize that there is a temporal delay in updating the information and CSAs may represent inaccurate price information, they may decide to visit each online store to validate the price displayed by CSAs. Most of the contemporary CSA research assume zero or negligible search costs. Because of temporal delay in price update, consumers may have to incur sequential search across multiple CSAs.

2.2 Selection Bias

Just like traditional retail markets, online markets also present substantial spatial price dispersion [2], [6], [10], [15], [35]. Although there are many different theoretical rationales for price dispersion, rationales based on search costs and heterogeneity of consumers are dominant in the academic literature [33], [34], [42]. The search costs based theory suggests that if there is a positive marginal cost for getting price information from each retailer then equilibrium price dispersion may exist. In the traditional retail market these search costs can be interpreted as the costs of making a trip to the retailer or calling a retailer [33]. However, this theory assumes that consumers are homogeneous. Theory based on consumer heterogeneity assumes different types of consumers based on their awareness of price information. These consumers have been broadly classified into informed and uninformed consumers [34], [42]. They assume that informed consumers have the knowledge of the entire distribution of offered prices and uninformed consumers know nothing about such distribution [42]. Here, the emphasis is on the knowledge of the entire price distribution and in the absence of CSAs, it was assumed that this entire price distribution could have been obtained by 'listing services' or 'information clearinghouse'. Most of the research in the online price dispersion uses the model based on heterogeneity of consumers and assume negligible or zero search costs [7]. It is assumed that CSAs provide a single click access to the complete price dispersion information.

The previous generation of the CSAs used to collect data directly from retailers by using Web scraping methods. However, now CSAs obtain data directly from retailers [18]. The revenue model of the majority of CSAs is merchant-driven. Majority of CSAs have their price list for various product categories and charge substantial fees to list a vendor for a specific product category. They reflect revenue driven selection bias and hence limits the coverage of retailers in their data [31]. CSAs have different criteria for selecting merchants (online retailers). Some CSAs list merchant prices only if a merchant provides XML price feeds. Some CSAs restrict the number of merchants whose price will be listed for each product category. Some CSAs demand high fees for listing and even ask for CPC charges from merchants. Not all the merchants can satisfy these criteria and hence it is not necessary that all merchants can participate in all CSAs. Retailers may take an optimal decision based on various criteria such as CPC charges and reach of CSAs, and participate in one or only few of the CSAs. For example, staples.com does not participate in pricegrabber.com's merchant program. Also, some stores do not participate at all. For example,

radioshack.com does not participate in the merchant programs of any of the six major CSAs that have been analyzed in this research. Such limited coverage with major exclusions may result in substantially different price dispersion data between two different CSAs. Hence,

Hypothesis 2: For the same product, price dispersions presented by two different CSAs differ significantly.

3   Data Collection and Methodology

Each of the hypotheses in the prior section requires a different set of data. While the first hypothesis addresses the inaccuracy in the information related to price dispersion because of a temporal delay in updating pricing information, the second hypothesis focuses on the incompleteness of the price dispersion related information due to selection bias among CSAs.

3.1   Temporal Delay

In order to calculate the temporal delay in updating price information on a CSA, two specific time-based references need to be determined:

1.   Price adjustment time (PAT): This is a reference time at which the price of a product is changed by a retailer.

2.   Price update time (PUT): This is a reference time at which the price of a product is updated by a CSA.

The temporal delay is the difference between PAT and PUT. It is important to realize that PUT may differ from CSA to CSA for the same merchant-product combination as each CSA may use a different data extraction technology or have a different data update procedure. Thus, in order to investigate the first hypothesis, it is important to carry out a comprehensive analysis using temporal delay data from the multiple CSAs. In this research, PUTs of six popular CSAs have been measured. The popularity of CSAs was measured by using Alexa.com rankings. It is a standard source that provides website rankings based on browsing behavior data.PUTs on CSAs always lag PATs on the merchants' websites. Hence, the measurement of PUTs requires a timely identification of PATs on merchants' websites.

Given the negligible menu costs, online price adjustments are frequent, random, and SKU-specific. It is practically challenging to identify PATs in such scenario. It can be measured only if all product prices on a merchant's website are continuously tracked. However, many online merchants and majority of click and mortar (C&M) merchants have their weekly sales circulars. In these circulars, they advertise their forthcoming price changes for selective products that go on for a sale on a specific day, typically Sunday. Such price adjustment generally remains valid for at least a week. Moreover, some merchants release their weekly sales circulars much in advance. For example, Staples.com releases its sales circular on Thursday for the upcoming week starting from Sunday. These weekly circulars provide PATs for a selective set of products. Thus, latest by early Sunday morning, one can determine a set of products whose PATs are on Sunday. Typical online C&Ms follow a regular schedule to adjust their prices based on these weekly sales circulars. For example, Staples change their prices based on its sales circular at 6:00 am EST on Staples.com. Based on primary exploratory study, it was identified that by 8:00 am Sunday, all major C&M retailers change their prices based on an advertised sales circular for the week starting on Sunday. It is important to note here that not all products advertised in sales circular require price adjustments. Merchants even advertise those products whose prices remain unchanged from one week to another. These products are discarded from this study. Thus by Sunday, the information about the merchants and their list of product along with their prices (both for the previous week and the upcoming week) are captured. For all these products, for the sake of simplicity, 8:00 am Sunday is considered as PAT for this study. Appendix A provides a list of products, corresponding retailers and prices (previous week and upcoming week).

Prices for all of these product-merchant combinations were tracked twice a day (9:00 am and 9:00 pm) on six CSAs for the next four days. This provided the second set of time-based reference point at which the prices are updated on various CSAs. As the data are captured only twice a day, it does not reflect the exact PUT, but it provides realistic estimated time duration at which the price changes are updated on CSAs. Measuring the exact PUT for all product-merchant combination requires continuous monitoring of these CSAs, which may not be possible realistically. Moreover, prices are not tracked after four days as for a typical price adjustment which may remain valid only up to a week, four days passes a half-way milestone and any delay beyond this is measured simply as five days and more. As the sample products for this research vary from one week to another, practically monitoring prices of all products for a prolonged time period was logistically challenging. For this research, the sales circular data have been collected from three merchants. Merchants had been selected based on the availability of their sales circulars well before the PAT of 8:00 am Sunday. This makes it easy to observe and record both current and forthcoming prices of various products. Not all products from these sales circulars were selected. Certain products such as retailer-branded products, generic products, bundled promotions, and products without specific model number related information were excluded from the sample as obtaining price quotes for such items through six CSAs was not possible. For

each product-retailer-CSA combination, the difference between PAT and PUT was calculated to determine the temporal delay in updating price information.

3.2 Selection Bias

In order to evaluate the completeness of price dispersion based hypothesis, one needs to determine a complete list of merchants selling a specific item. This may eventually require searching all potential merchants selling sample items. Such exploratory study may not be fruitful as it may leave many potential merchants out because of the long-tail phenomenon in electronic commerce. An alternate approach to show the incompleteness of price dispersion information on a CSA is to compare the price dispersion information on two different CSAs and evaluate the differences. If such difference is significantly different than it suggests the selection bias of individual CSAs. In this study, two popular CSAs, Shopping.com and Cnet.com are selected and the price dispersion information for a sample of products is compared.

Products were selected from Amazon.com's top 100 electronics category. Out of these 100 products, digital products such as software and computer games were discarded as well as other products which had bundle promotions. Certain other products such as battery chargers and generic items were not selected as well. In the absence of any specific product identifier, it was difficult to find price quotes of such products on CSAs. Our final list of products consisted of 41 items. Each of these items was searched on two different CSAs and detail information about their prices was collected. Aggregated information from this data is presented in Appendix B. Various measures were used to compare the price dispersion information from different CSAs. In appendix B, *count* represents the number of merchants selling an item. *Max*, *Min*, and *Avg* are the maximum, minimum, and average price of an item respectively. *Stdev* is the standard deviation of various price quotes. CV stands for the coefficient of variation and it is a ratio of standard deviation and mean of the price quotes. Coefficient of variation is a standard unit less measure for price dispersion and widely used in this field [5], [28], [35], [37].

4 Results and Analysis

Temporal delays in updating prices are calculated by subtracting PUTs and PATs. Overall results for temporal delay are provided in Appendix A. In order to analyze the selection bias of CSAs, price dispersion information on two CSAs is evaluated by using standard measures such as CV, price range, maximum and minimum price, average price, and standard deviations. Results for this comparison are provided in Appendix B.

The meaning of substantial temporal delay in hypothesis 1 may have different significance based on consumer preferences, degree of price changes, or product characteristics. Given the nature of data transfer between merchants and CSAs, it is practically challenging to have real time data update on CSAs. However, a delay of more than 24 hours is substantial enough to misrepresent the price information to a potentially large audience. For the purpose of this research, it is assumed that any delay of more than one day is substantial. It is important to note here that due to lower online menu costs, retailers may make random and frequent changes in price at the SKU level [1], [25] and the combination of long tail and virtually global market reach [9] may create a substantial market for a given SKU on any given day. Hence, a delay of one day in updating price is substantial in today's fast changing marketplace. The null hypothesis suggests that the difference between PAT and PUT is less than a day and the alternative hypothesis claim that the delay is more than a day. It can be seen from Appendix A that for some products the temporal delay is clearly more than a day. For example, the difference between PUT on shopping.com and PAT on Staples.com for a 19" CRT monitor is two days.

One of the interesting observations from this data suggests that the temporal delay is a multi-dimensional phenomenon. Temporal delays in update vary across various retailers, CSAs, products, and even time periods. For example, as shown in figure 1, the temporal delays for two different retailers not only vary for individual retailers but also vary across different CSAs. In order to test hypothesis 1, a conservative measure of the minimum temporal delay is used to compare it with the notion of substantial temporal delay (i.e. one day). The minimum temporal delay for a merchant-product pair is the shortest time period required to update the price information on any of the six CSAs. This information is provided in the last column of Appendix A. For example, for *all in one printer (product 2 in appendix B)*, the PUT on pricegrabber.com is three days against five days by other CSAs. Hence, according to this conservative measure minimum time to update has been considered as three days. The results for the t-test are shown in table 1.

Table 1: Hypothesis 1 results

Mean	Standard Deviation	t-statistic	p Value
2.46	1.62	4.78	0.00003

As can be seen from table 1, hypothesis 1 is strongly supported. Thus, the difference between PAT on a merchant's website and PUT on a CSA is substantial. Comparison of these results across different CSAs is provided in Table 2.

Table 2: Comparison of temporal delay across various CSAs

	CNET	Shopping.com	MySimon	Bizrate	Nextag	PG
Mean	2.54	3.39	2.76	2.82	2.96	2.96
Standard Deviation	1.69	1.17	1.8	1.89	1.72	1.74
t-Statistic	4.81	10.86	5.15	5.1	5.46	5.37
p Value	0	0	0	0	0	0

As can be seen from Table 2, some CSAs, such as shopping.com, take on an average 3.39 days to update the price information. While other CSAs, such as Cnet.com take on an average 2.53 days to update price changes. Regardless of products, retailers, or time periods, these delays are substantial across all CSAs. If merchants start changing prices more frequently, then a CSA is more vulnerable to inaccurate information and in the absence of a sequential search or multiple click-through, consumers may make suboptimal purchase decisions.
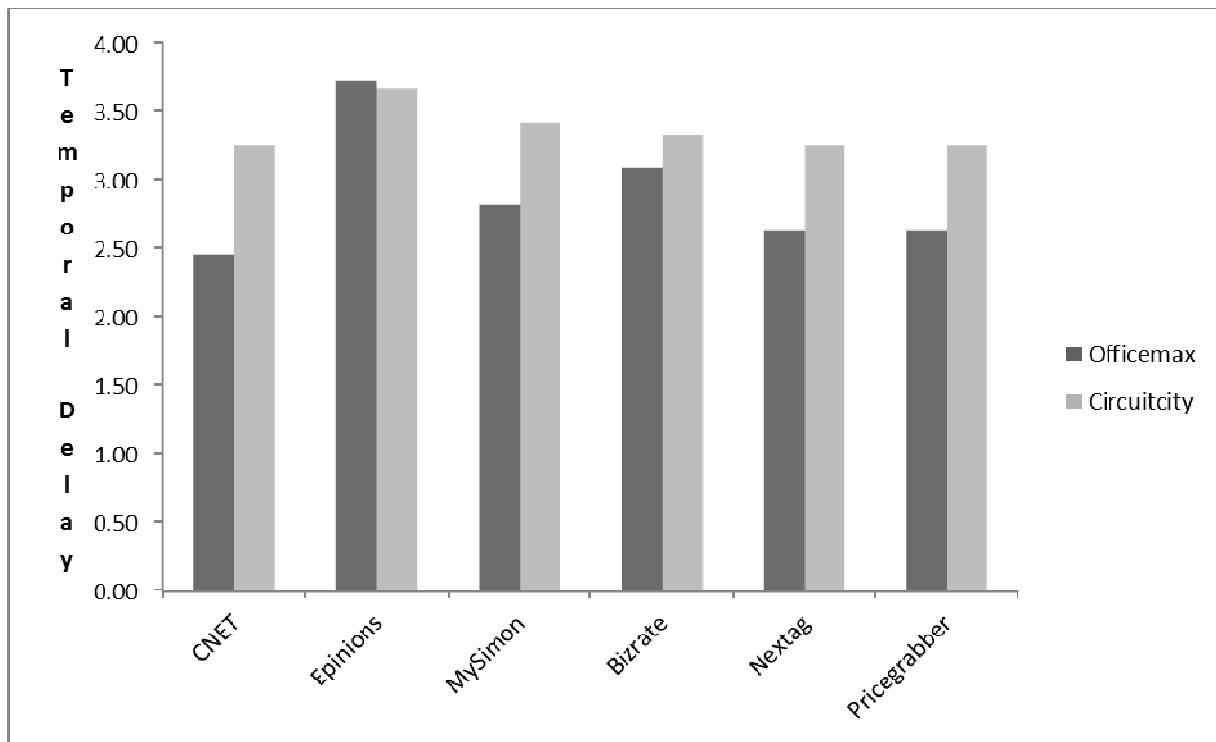

Figure 1: Comparison of temporal delays in price update across different CSAs

To compare the price dispersion between two CSAs, standard measures of price dispersion have been used such as range, standard deviation, mean, number of sellers, and coefficient of variation. As can be seen from Appendix B, although minimum prices are very similar between two CSAs, other price dispersion parameters such as standard deviation and CV differ. Various descriptive statistics in Appendix B help in understanding the difference between the price dispersion information presented by two CSAs. Table 3 provides the summary of average price dispersion parameters. Figure 2 (a) and Figure 2 (b) provide comparison of this information between these CSAs. Table 3 provides the aggregated price dispersion measures.

Table 3: Comparison of average price dispersion parameters from two CSAs

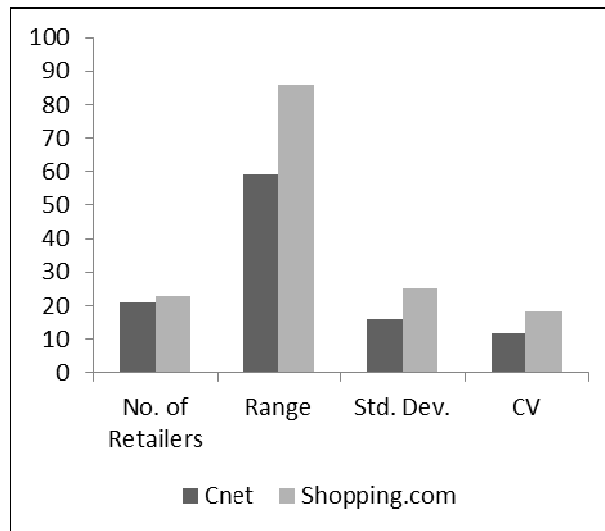| CSAs	No. Of Retailers	Max. Price	Min. Price	Standard Deviation	Mean Price	CV
Cnet	21	213	154	16	179	11.9
Shopping.com	23	238	152	25	185	18.2

Figure 2 (a): Comparison of average price dispersion parameters between two CSAs
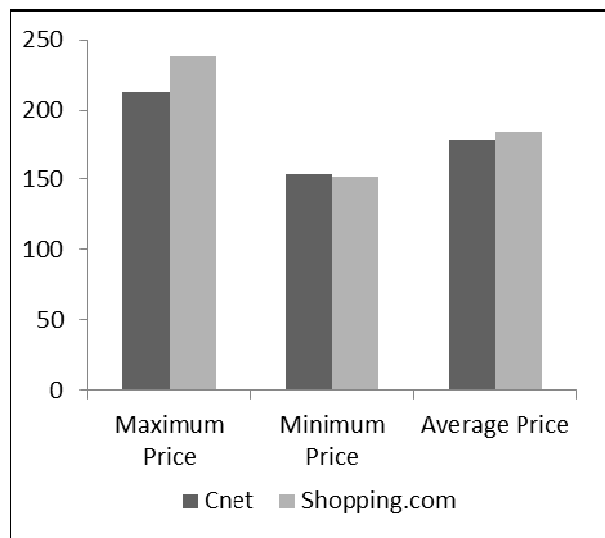

Figure 2 (b): Comparison of average price dispersion parameters between two CSAs

One of the most important measures of price dispersion is coefficient of variation. It has been widely used for comparing price dispersion across two different populations [7], [35], [37]. To test hypothesis 2, CVs of price dispersion from two CSAs are compared. The null hypothesis is that the difference of CV between two CSAs is zero. The t-statistic for a two-tailed test is -2.55 and p-level is 0.01.

Hence, hypothesis 2 is strongly supported. The price dispersions between two CSAs differ significantly and one CSA does not provide complete spatial price dispersion. In the absence of such incomplete price dispersion information, consumers may make suboptimal purchase decisions if they rely on a singular CSA. It is important to note here that the consumer's purchase decision may not be based on the minimum price alone. Consumers may rely on multiple factors such as price, retailer reputation, service, and shipping and return policies etc… and hence they may need complete information from potentially all sellers selling a specific product. Consumers may have to incur additional search costs in order to obtain complete price dispersion related information which may include visits to multiple CSAs or conducting a sequential search across all potential sellers of the product.

5   Discussions and Implications

Just like its B&M counterpart, the online retail market also displays spatial price dispersion. Consumer heterogeneity and search costs are the two primary reasons for the existence of the price dispersions in the online markets. Contemporary IS research shows that in the online markets, the law of single price does not prevail even with emergence of the CSAs. Consumer and retailer heterogeneities are provided as two primary explanations for this continuing price dispersion in the online markets. The fundamental contribution of this research is to provide a search-cost based explanation for online price dispersion.

The CSAs may not present complete price dispersion information because of their merchant-sponsored revenue model. CSAs selection criteria for merchants are predominantly revenue-based. Moreover, many merchants do not participate in the CSA-based referral programs. In this research, by comparing two different CSAs, it has been shown that the price dispersion information presented by an individual CSA significantly differs from the other CSA. Consumers may not make an optimal purchasing decision by relying on such incomplete price dispersion information from a single CSA and may have to search on multiple CSAs or even merchant websites in order to get comprehensive price dispersion information. Likewise, the CSAs may not present accurate price dispersion information because of their data collection methodology. Majority of contemporary CSAs receive pricing data from merchants at discrete intervals. Given lower menu costs, online merchants tend to change prices more frequently. However, it may not be feasible to update these SKU specific price adjustments to all CSAs in real time and hence at any given instance of time, the CSAs may present incorrect prices. Such realization may require consumers to click through multiple merchant websites to validate the pricing information.

Researchers have been intrigued by the ongoing existence of price dispersion in the online retail market. Prior research has provided retailer heterogeneity, consumer heterogeneity, and obfuscation based justifications for online price dispersion. This paper studies online price dispersion related information from various CSAs to evaluate the quality of information in terms of completeness and accuracy of price quotes from multiple retailers. The results clearly show that due to merchant-supported models of current CSAs and price quote data collections at discrete time intervals, the current generation of CSAs does not provide complete and accurate price dispersion. More importantly, the completeness and accuracy of the price dispersion varies from one CSA to another. Even if the user visits multiple CSAs, she may not be able to obtain complete and accurate price dispersion information. Under such conditions, users may have to incur additional search costs in order to make optimal purchasing decisions. Hence, the emergence of CSAs may not negate the search requirement and search costs should be considered as one of the potential explanations for the price dispersion in the online markets.

With regard to temporal delay, the results from appendix B show that temporal delays not only vary from one CSA to another but also differ from one retailer to another. More importantly, even for the same retailer, temporal delays vary from one product to another. For example, for all-in-one printers, the temporal delay for Circuitcity is 3 days for Epinions.com. For other CSAs, the temporal delay in updating price for all in one printer is 2 days. However, for digital cameras, the temporal delay for Circuitcity is at least 5 days for all CSAs. Thus, the price updates at different CSAs for different products and retailers happen at different times. In such cases, it is difficult for the user to predict any pattern and unless she visits individual retailer websites, it is difficult to ascertain the accuracy of a given price quote. Thus, sequential search across various retailers' websites is imperative in order to obtain accurate price dispersion information.

For selection bias, the coverage of retailers varies significantly from one CSA to another. Moreover, this coverage differs from one product to another. For example, as shown in appendix A, while Cnet.com covers 31 retailers for a product 3 (wireless desktop), shopping.com covers only 8 retailers. However, for a product 4 (digital camera), Shoppimg.com covers 49 retailers versus 34 in the case of Cnet.com. Likewise, minimum price may also differ from one CSA to another. For example, for the first product in the sample (digital camera), the minimum price at CNet is $219. However, the minimum price at Shopping.com for the same camera is $180. Such evidences clearly show that incomplete coverage by CSAs may provide suboptimal purchasing related decision support and hence may encourage additional search.

This research contributes towards our knowledge of the CSAs in multiple ways. First, while current literature explains the existence of online price dispersion by various phenomena including consumer heterogeneity [13], retailer heterogeneity [10], service heterogeneity [30], [31], pricing strategies [17], and retailer obfuscation [12], this research suggests that even with the presence of the CSAs, search costs may still persist in the online retail market because of selection bias and temporal delay. Thus this research provides a search cost based explanation for the existence of online price dispersion. Second, prior researchers have provided various suggestions for the improvement of CSAs including providing comprehensive product, price, promotion, and retailer specific search solutions [19], personalization [26], multi-parameter ordered list [46], and bundle product solutions [20], [44]. This research suggests that the next generation of CSAs should focus on the basics first – that is to reduce the search costs by providing accurate and complete price dispersion information. Third, researchers have developed models to address certain limitations of CSAs such as selection bias and temporal delay. For example, it has been suggested that CSAs should use learning algorithms to conduct a selective search across a limited set of retailers in real time to assure optimal comparison shopping results with limited coverage [27]. This research suggests that a combined approach of selective coverage and real time data extraction may optimize purchase decision making for the user.

This research has a several limitations. On the one hand these limitations restrict the generalization of the results of this study for different products and markets. On the other hand they provide excellent future research opportunities. First, majority of CSAs in our sample have merchant-supported revenue models. Some other CSAs like Bingshopping.com or Google products based on alternate revenue models are excluded from the research. Further research is required to compare the selection bias and temporal delay related issues among CSAs with different revenue models. Second, the results of this study are based on a limited set of data. In particular, only two CSAs are compared for the selection bias issue. Likewise, only three retailers are studied for the temporal delay related issue.

While six CSAs are analyzed for evaluating temporal delay, only two CSAs are studied for selection bias. The primary reason for this limited set of data is because of logistical reasons. For example, with regard to selection bias, obtaining price quotes for 41 products from on an average 22 retailers and two CSAs constitute 1760 price quotes in a limited time frame. As the selection of CSAs in this study is based on Alexa.com rankings and in general majority of popular CSAs operate similarly, it is reasonable to believe that the results of this study can be extended to online comparison shopping field in general. However, an extensive research is required to study the broader impacts. It is also interesting to determine the link between the prices shown on the CSAs, price dispersion, and its impact on consumers' buyer behavior. One of the potential future research areas is to conduct a lab study to analyze such relationships.

Third, the product sample for comparing selection bias mostly includes popular products. It remains to be seen how selection bias influences the price dispersion information for obscure items. Recent research [21] suggests that different product categories may respond differently in the online markets. The focus of this paper is on a single product category and it can be extended to analyze the impact of temporal delay and selection bias in different product categories. Fourth, this research does not measure or quantify the search costs and hence does not directly measure the impact of selection bias and temporal delay on search costs. Measuring such search costs is challenging given that it may require an exploratory study to check the availability, price, tax, and shipping related information on a large number of sellers. However, one can compare the differential impact of selection bias and temporal delay on search costs. As many retailers remain uncovered by various CSAs, one has to conduct a sequential search across individual retailers as well as multiple CSAs in order to get the complete price dispersion in the case of selection bias related issue. While, search cost related problems related to temporal delay may require less search costs because the user can click through hyperlinks and visit a limited set of retailers for validation purpose. Future research in this field should analyze the differential impact of these two phenomena on search costs and analyze the consumer product choice strategy under such situation in line with [39].

References

[1]   M. Arbatskaya and M. R. Baye, Are prices 'sticky' online? Market structure effects and asymmetric responses to cost shocks in online mortgage markets, International Journal of Industrial Organization, vol. 22, no. 10, pp. 1443-1462, 2004.

[2]   J. P. Bailey, Intemediation and electornic markets: Aggregation and pricing in Internet commerce, Ph.D. Thesis, Technology, Management and Policy, Massachusetts Institute of Technology, Cambridge, MA, 1998.

[3]   J. Y. Bakos, Reducing buyer search costs: Implications for electronic marketplaces, Management Science, vol. 43, no. 12, pp. 1676-1692, 1997.

[4]   M. Baye, J. Morgan and, P. Scholten, Persistent price dispersion in online markets, in The New Economy and Beyond: Past, Present and Future (D. W. Jansen, ed.). New York: Edward Elgar Publishing, pp. 122-143.

[5]   M. Baye and J. Morgan, Temporal price dispersion: Evidence from an online consumer electronics market, Journal of Interactive Marketing, vol.18, no. 4, pp. 101-115, 2004.

[6]   M. Baye and J. Morgan, Price dispersion in the lab and on the internet: Theory and evidence, The Rand Journal of Economics, vol. 35, no. 3, pp. 449-466, 2004.

[7]   M. Baye and J. Morgan, Price dispersion in the small and in the large: Evidence from an internet price comparison site, The Journal of Industrial Economics, vol. 52, no. 4, pp. 463-496, 2004.

[8]   K. Baylis and J. M. Perloff, Price dispersion on the internet: Good firms and bad firms, Review of Industrial Organization, vol. 21, no. 3, pp. 305-324, 2002.

[9]   Brynjolfsson, Erik, Yu J. Hu, and Michael D. Smith, From Niches to riches: The anatomy of the long tail, Sloan Management Review, vol. 47, no. 4, pp. 67-71, 2006.

[10] E. Brynjolfsson and M. D. Smith, Frictionless commerce? A comparison of internet and conventional retailers, Management Science, vol. 46, no. 4, pp. 563-585, 2000.

[11] E. Brynjolfsson and M. D. Smith, The great equalizer? An empirical analysis of consumer choice behavior at an internet shopbot, Carnegie Mellon University, Tepper School of Business, 2007.

[12] R. K. Chellappa, R. G. Sin, and S. Siddarth, Price formats as a source of price dispersion: A study of online and offline prices in the domestic U.S. airline markets, Information Systems Research, vol. 22, no.1, pp. 269-288, 2011.

[13] P. Chen and L. Hitt, Measuring switching costs and the determinants of customer retention in internet-enabled businesses: A study of the online brokerage industry, Information Systems Research, vol. 13, no. 3, pp. 255-274, 2002.

[14] K. Clay, R. Krishnan, and E. Wolff, Prices and price dispersion on the web: Evidence from the online book industry, The Journal of Industrial Economics, vol. 49, no. 4, pp. 521-539, 2001.

[15] E. Clemons, I. Hann, and L. Hitt,Price, Dispersion and differentiation in online travel: An empirical investigation, Management Science, vol. 48, no. 4, pp. 534-549, 2002.

[16] K. Crowston and I. MacInnes, The effects of market-enabling internet agents on competition and prices, Journal of Electronic Commerce Research, vol. 2, no. 1, pp. 1-22, 2001.

[17] G. Ellison and S. Ellison, Search, obfuscation, and price elasticities on the internet, Econometrica, vol. 77, no. 2, pp. 427-452, 2009.

[18] M. Fasli, Shopbots: A syntactice present, a semantic future, IEEE Internet Computing, vol. 10, no. 6, pp. 69-75, 2006.

[19] R. Garfinkel, R. Gopal, B. Pathak, and F. Yin, Shopbot 2.0: Integrating recommendations and promotions with comparison shopping, Decision Support Systems, vol. 46, no. 1, pp. 61-69, 2008.

[20] R. Garfinkel, R. Gopal, A. Tripathi, and F. Yin, Design of a shopbot and recommender system for bundle purchases, Decision Support Systems, vol. 42, no. 3, pp. 1974-1986, 2006.

[21] M. Goldmanis, A. Hortacsu, and C. Syerson, E-commerce and the market structure of the retail industries, The Economic Journal, vol. 120, no. 545, pp. 651-682, 2010.

[22] G. Haubl, B. Dellaert, K. Murray, and V. Trifts, Buyer behavior in personalized shopping environments, in Designing Personalized User Experiences in Ecommerce (C-M. Karat, J. O. Blom, & J. Karat, eds.). Norwell, MA, USA: Kluwer Academic Publishers, pp. 207-229.

[23] G. Haubl and K. Murray, Preference construction and persistence in digital marketplaces: The role of electronic recommendation agents, Journal of Consumer Psychology, vol. 13, no. 2, pp. 75-91, 2003.

[24] M. Janssen and J. L. Moraga-Gonzalez, Strategic pricing, consumer search, and the number of firms, The Review of Economic Studies, vol. 71, no. 249, pp. 1089-1119, 2004.

[25] D. Levy, M. Bergen, S. Dutta, and R. Venable, The magnitude of menu costs: Direct evidence from large U.S. supermarket chains, The Quarterly Journal of Economics, vol. 112, no. 3, pp. 791-825, 1997.

[26] A. Menczer, A. Monge, and W. Street, Adaptive assistants for customized e-shopping, IEEE Intelligent Systems, vol. 17, no. 6, pp. 12-19, 2002.

[27] A. Montgomery, K. Hosanagar, R. Krishnan, and K. Clay, Designing a better shopbot, Management Science, vol. 50, no. 2, pp. 189-206, 2004.

[28] J. Morgan, H. Orzen, and M. Sefton, An experimental study of price dispersion, Games and Economic Behavior, vol. 54, no.1, pp. 134-158, 2006.

[29] X. Pan, B. Ratchford, and V. Shankar, Price dispersion on the internet: A review and directions for future research, Journal of Interactive Marketing, vol. 18, no. 4, pp. 116-135, 2004.

[30] X. Pan, B. Ratchford, and V. Shankar, Can price dispersion in online markets be explained by differences in e-tailer service qualit?, Journal of the Academy of Marketing Science, vol. 30, no. 4, pp. 433-445, 2002.

[31] B. Pathak, A survey of the comparison shopping agent-based decisions support systems, Journal of Electronic Commerce Research, vol. 11, no. 3, pp. 177-192, 2010.

[32] P. Pendersen, Behavioral effects of using software agents for product and merchant, International Journal of Electronic Commerce, vol. 5, no. 1, pp. 125-141, 2000.

[33] J. Reiganum, A simple model of equilibrium price dispersion, Journal of Political Economy, vol. 87, no. 4, pp. 851-858, 1979.

[34] S. Salop and J. Stiglitz, Bargains and ripoffs: A model of monopolistically competitive price dispersion, Review of Economic Studies, vol. 44, no. 3, pp. 493-510, 1977.

[35] P. Scholten and A. Smith, Price dispersion then and now: Evidence from retail and e-tail markets, Advances in Applied Microeconomics, vol. 11, no. 11, pp. 63-88, 2002.

[36] M. Smith, The impact of shopbots on electronic markets, Journal of the Academy of Marketing Science, vol. 30, no. 4, pp. 446-454, 2002.

[37] A. Sorensen, Equilibirium price dispersion in retail market of prescription drugs, Journal of Political Economy, vol. 108, no. 4, pp. 833-862, 2000.

[38] S. Sproule and N. Archer, A buyer behavior framework for the development and design of software agents in e-commerce, Internet Research: Electronic Networking Applications and Policy, vol. 10, no. 5, pp. 396-405, 2000.

[39] B. Su, Consumer e-tailer choice strategies at on-line shopping comparison sites, International Journal of Electronic Commerce, vol. 11, no. 3, pp. 135-159, 2007.

[40] P. Todd and I. Benbasat, The use of information in decision making: An experimental investigation of the impact of computer-based decision aids, MIS Quarterly, vol. 16, no. 3, pp. 373-394, 1992.

[41] V. Trifts and G. Haubl, Information availability and consumer preference: Can online retailers benefit from providing access to competitor price information?, Journal of Consumer Psychology, vol. 13, no. 2, pp. 149-159, 2003.

[42] H. Varian, A model of sales, American Economic Review, vol. 70, no. 4, pp. 651-659, 1980.

[43] R. Waldeck, Search and price competition, Journal of Economic Behavior & Organization, vol. 66, no. 2, pp. 347-357, 2006.

[44] Y. Wan and G. Peng, What's next for shopbots?, Computer, vol. 43, no. 5 pp. 20-26, 2010.

[45] B. Xiao and I. Benbasat, E-commerce product recommendation agents: Use characteristics, and impact, MIS Quarterly, vol. 31, no. 1, pp. 137-209, 2006.

[46] Y. Xu and H. Kim, Order effect and vendor inspection in online comparison shopping, Journal of Retailing, vol. 84, no. 4, pp. 477-486, 2008.

Appendix A: Comparison of Price Dispersion Information

No.	Products	Cnet						Shopping.com					
		Count	Max	Min	Stdev	Avg	CV	Count	Max	Min	Stdev	Avg	CV
1	Digital camera	7	320	219	37	269	13.78	15	350	180	49	249	19.51
2	MP3 player	25	300	275	9	288	2.97	27	324	275	12	295	3.91
3	Wireless dekstop	31	61	28	8	44	18.58	8	52	30	9	41	23.07
4	Digital camera	34	400	292	38	349	10.86	49	499	283	45	355	12.73
5	Router	34	150	50	16	64	25.63	40	85	53	6	64	9.98
6	Cordless phone	7	138	80	19	115	16.69	28	180	60	25	111	22.18
7	Memory	22	105	70	11	84	13.04	20	183	70	29	94	31.24
8	MP3 player	6	76	67	3	70	3.96	9	70	59	4	67	6.73
9	Digital camera	33	350	235	32	297	10.8	55	410	220	44	296	15.01
10	Memory	3	130	123	4	128	3.19	7	219	112	39	146	26.41
11	Memory	32	180	84	19	104	18.12	31	231	84	36	111	32.21
12	MP3 player	17	200	190	4	196	1.94	17	200	190	3	197	1.68
13	MP3 player	4	204	190	6	196	2.88	9	265	174	29	211	13.78
14	Memory	10	110	72	13	89	14.38	9	230	72	49	103	47.33
15	MP3 player	12	129	80	16	101	15.39	19	139	80	19	105	17.71
16	MP3 player	14	150	140	4	146	2.53	17	150	143	3	148	1.89
17	Digital camera	30	1600	1248	100	1429	7.01	45	1600	1250	107	1445	7.43
18	MP3 player	4	195	182	6	192	3.38	7	303	182	44	231	19.04
19	Digital camera	39	361	258	26	291	8.88	54	400	245	35	296	11.88
20	Memory	23	80	61	5	68	7.38	18	80	60	5	68	7.66
21	Hard drive	11	179	128	14	142	9.74	4	179	136	18	155	11.86
22	Wireless dekstop	18	41	25	4	36	11.29	15	50	28	5	37	13.44
23	MP3 player	11	500	374	48	436	10.98	12	500	360	49	428	11.43
24	XM Radio	27	349	204	43	258	16.78	28	350	204	46	257	18.01
25	Memory	36	104	47	12	60	20.32	35	111	46	14	61	22.63
26	Memory	29	140	88	12	102	11.75	34	140	88	15	106	14.59
27	Memory	19	100	65	9	83	10.66	6	100	65	13	80	16.48
28	MP3 player	16	120	40	24	70	34.56	19	120	51	21	77	26.76
29	Memory	25	130	34	20	50	40.61	14	130	38	25	55	45.26
30	MP3 player	4	195	189	3	193	1.54	10	303	174	45	231	19.32
31	Memory	38	70	37	7	49	13.94	34	100	37	15	52	28.11
32	MP3 player	4	195	189	3	193	1.54	15	200	182	5	196	2.49
34	Memory	2	30	27	2	29	6.61	11	60	26	13	35	35.64
35	Digital camera	45	240	150	16	182	8.97	51	230	162	15	180	8.25
36	Printer	34	111	45	14	91	15.27	35	139	49	15	97	15.56
37	Wireless dekstop	18	59	38	6	46	12.42	12	53	29	6	46	14.08
38	MP3 player	17	200	182	5	196	2.32	17	200	182	4	197	2.28
39	Printer	51	164	60	19	135	13.8	49	169	104	12	142	8.7
40	MP3 player	23	350	324	8	342	2.23	17	350	324	8	342	2.32
41	Memory	37	89	41	11	52	20.5	32	120	40	20	56	36.14
	Average	21	213	154	16	179	11.9	23	238	152	25	185	18.2

Appendix B: Comparsion of Temporal Delay in Price Update

Product	Previous Price	New Price	Price Change Date	Retailer	Temporal delay (Number of days)*						
					CNET	Epinions	MySimon	Bizrate	Nextag	PG	Minimum
CRT Monitor	199	184	3-Apr	Staples	1	2	1	1	NA	NA	1
All in One Printer	130	110	3-Apr	Officemax	5	5	5	5	5	3	3
All in One Printer	130	115	3-Apr	Officemax	5	5	5	5	5	5	5
All in One Printer	130	100	3-Apr	Officemax	5	5	5	5	5	5	5
All in One Printer	400	360	10-Apr	Officemax	1	3	1	1	1	1	1
All in One Printer	300	260	10-Apr	Officemax	1	3	1	1	1	1	1
All in One Printer	180	160	10-Apr	Officemax	1	3	1	1	1	1	1
All in One Printer	200	180	10-Apr	Officemax	1	3	1	1	1	1	1
All in One Printer	300	260	10-Apr	Circuitcity	2	3	2	2	2	2	2
All in One Printer	400	360	10-Apr	Circuitcity	2	3	2	2	2	2	2
Cordless Phone	129	119	3-Apr	Staples	1	2	1	1	NA**	NA	1
Digital Camera	349	229	3-Apr	Staples	1	2	1	1	NA	NA	1
Digital Camera	400	380	17-Apr	Circuitcity	5	5	5	5	5	5	5
Laser Printer	600	500	3-Apr	Officemax	3	4	5	6	4	5	3
Laser Printer	500	425	3-Apr	Officemax	3	4	5	6	4	5	3
Laser Printer	700	560	10-Apr	Officemax	1	3	1	2	1	1	1
Laser Printer	400	300	10-Apr	Officemax	1	3	1	1	1	1	1
Memory	55	45	10-Apr	Circuitcity	2	3	2	2	2	2	2
Memory	60	45	17-Apr	Circuitcity	2	2	3	3	2	2	2
Memory	50	45	17-Apr	Circuitcity	2	2	3	2	2	2	2
Memory	105	85	17-Apr	Circuitcity	5	5	5	5	5	5	5
Memory	70	50	17-Apr	Circuitcity	5	5	5	5	5	5	5
Memory	45	40	17-Apr	Circuitcity	5	5	5	5	5	5	5
Memory	140	120	17-Apr	Circuitcity	5	5	5	5	5	5	5
MP3 Player	249	229	3-Apr	Staples	1	2	1	1	NA	NA	1
MP3 Player	150	140	10-Apr	Circuitcity	2	3	2	2	2	2	2
PDA	150	130	10-Apr	Circuitcity	2	3	2	2	2	2	2
Wireless desktop	99	79	3-Apr	Staples	1	2	1	1	NA	NA	1

* Temporal delay of 1, 2, 3, and 4 means prices are updated on the first, second, third, or fourth day respectively. Temporal delay of 5 suggests that the price update takes more than four days.
** During data collection, Staples was not participating in the referral programs of Pricegrabber.com and Nextag.com for a certain product categories.