

ONLINE PRICE DISPERSION, AN INTERNATIONAL COMPARISON

Bounie D.⁵⁹, Eang B.⁶⁰, Sirbu M.⁶¹ and Waelbroeck P.⁶²

6.1 Introduction

Most studies on online markets find significant and persistent price dispersion regardless of the product categories (Baye, Morgan and Scholten, 2004; Pan et al., 2004). By contrast, in theory, the price of an homogenous product offered by several sellers should quickly converge to the best offer. The fact that menu costs are much smaller online than the cost of printing new catalogs strengthens the theoretical argument for price convergence. Further, the availability of price comparison sites considerably lowers search costs online.

Several theoretical explanations have been proposed to solve this apparent puzzle. First, Varian (1980) has argued that search costs related to geographical barriers lead firms to post different prices. Secondly, Baye and Morgan (2001) explain how price dispersion can emerge due to search costs when where firms compete on price with homogeneous goods. These models posit that consumers purchase at the lowest posted price and that online price dispersion decreases with the number of sellers. Thirdly, Pan et al. (2004) have suggested that non-price competition such as reputation and obfuscation strategies can lead to price dispersion.

The empirical literature on online price dispersion has looked at these factors. Claye et al., (2002), Brynjolfsson and Smith (2000) and Brynjolfsson et al., (2010) have shown that search costs can explain price dispersion even when internet users search through comparison sites. Researchers have

⁵⁹ Department of Economics and Social Sciences, Telecom ParisTech.

⁶⁰ Department of Economics and Social Sciences, Telecom ParisTech.

⁶¹ Department of Engineering and Public Policy and Tepper School of Business, Carnegie Mellon University.

⁶² Department of Economics and Social Sciences, Telecom ParisTech.

found in online auctions such as eBay (Resnick and Zeckhauser, 2002; Bajari and Hortacsu, 2004; Houser and Wooders, 2006) that reputation increases sales probability and the price at which an item is sold. The reputation effect seems more difficult to prove in other contexts; Ghose et al. (2009) looked at the impact of vendor reputation on markets for both new and used goods on Amazon Marketplace (U.S.) and found an inconsistent effect of reputation on prices.

In addition to reputation and search costs, we believe that two additional dimensions are important to analyze: the time dimension and the maturity of the market. First, price dispersion could persist over time due to seller strategy for instance if they do not change or, conversely, if they substantially change their prices. Moreover, price dispersion could exist due to the entry of new sellers at different prices on the market. Secondly, in less mature markets, large sellers could have more market power and discourage entry which in turn reduces price dispersion. The opposite could be true in mature online markets.

To study the impact of these factors on price dispersion, we have collected data on all *new* books, CDs and DVDs entering the daily TOP 100 list in the US, the UK and France Amazon Marketplace platforms from March 2006 to November 2006. Unlike many shopbots, the Amazon Marketplace by default lists vendors from lowest total cost (price plus shipping) to high. This dataset has several interesting features. First, the search costs are low on the platform. Secondly, the time dimension of the data set can be used to study pricing strategies over time (how often vendors change their prices and the characteristics of entry of new sellers). Thirdly, using the cross variation in the characteristics of the sellers, we construct three measures of reputation that we use to estimate a reputation premium across markets. Fourth, we propose a method for inferring at what price a transaction takes place (search-based theories posit that transactions take place at the lowest prices) and find that transactions are not always at the lowest price. Finally, by comparing three platforms in different countries, we can analyze how price dispersion is sensitive to the maturity of the market.

This study enriches the growing empirical literature on the nature of competition in online markets that it is interested in topics such as pricing theories (Acquisti and Varian, 2005; Ghose and

Sundararajan, 2006), market structure and functioning (Ellison and Ellison, 2005), cultural variety and business models (Ghose and Gu, 2006), online communities and infomediaries (Caillaud and Jullien, 2003), etc. First, using posted prices, we show that vendors make few modifications to their prices over time (once every 10 to 20 days), and those variations that they do make are relatively small (on average 15 cents). This challenges explanations of price dispersion based on some form of mixed or random pricing strategy by vendors (Baye and Morgan, 2001). Secondly, we compare posted prices to transaction prices and we show that transaction prices do not necessarily correspond to the lowest posted price but rather are centered around the median price. This result confirms the findings of Brynjolfsson et al. (2010) and challenges findings by Ghose and Yao (2007). Thirdly, we show that online price dispersion increases with the number of sellers mainly because new sellers price their product significantly lower than the current average price. This intuitive result challenges studies by Spulber (1995), Claye et al., (2002) and Baye, Morgan and Scholten (2004) that find that price dispersion decreases with the number of sellers.⁶³ Fourth, we show that there is a significant reputation premium of about 10 to 15%. A well established seller can charge a price 10 to 20% higher than the average seller. Finally, our estimation results support the idea that price dispersion is sensitive to the maturity of the market. Consistent with our finding that price dispersion increases with the number of sellers, less mature markets, such as France, with fewer sellers exhibit lower price dispersion. We also find that large sellers account for a greater proportion of transactions in France than in the U.S.

The article is organized as follows. In Section 2 we describe our dataset. Section 3 examines posted price dispersion while Section 4 analyzes transaction price dispersion. Section 5 presents our conclusions.

6.2 Data Set and Definitions

Our data set consists of posted offers compiled from Amazon Marketplaces (Amazon.com, Amazon.uk and Amazon.fr) for CDs, DVDs and book markets. Amazon Marketplace is Amazon's

⁶³ The last study uses a controversial measure of price dispersion based on the difference between the two lowest posted prices that statistically decreases with the number of sellers and posted prices.

online marketplace that allows sellers to offer their new items alongside Amazon's offerings. Amazon merchants and marketplace sellers are independent⁶⁴.

The data are gathered daily using automated Perl scripts. We study each item (a CD, a DVD and a book) entering in the TOP 100 on the Amazon websites and analyze the posted offers for each item on the respective Amazon Marketplaces. The data were collected over 240 days between March 2006 and November 2006. We follow each item entering the TOP 100 any time in the first 22 weeks, from its first appearance until the end of the study interval, even if the item later passes out of the TOP 100. As a result of this procedure, the number of different items we follow in each category depends only on the rate of turnover in the top 100, and is not in any way representative of the market size. Each item is described by:

- Characteristics of the item i (CD, DVD and Book): ID (ISBN, etc.), title, type of the product (ex. hardcover, etc.), etc.
- Characteristics of a posted offer k of an item i at time t (price, sales rank, Amazon price, state of the good (new or used), etc.) on the Amazon marketplace.
- Characteristics of a seller j who posted an offer k of an item i at time t (rating score, price, etc.) on the marketplace platform.

The main goal of our study is to analyze price dispersion for new goods on the Amazon marketplaces; the study of used goods in the Amazon marketplace is outside the scope of this paper.

There are several data issues to deal with when we analyze posted price dispersion. First, some sellers post prices above \$3 million in the US DVD market. This is obviously a bug. As a general rule, we dropped items for which the price is greater than twice the median price of all items sold on the market in our dataset. We provide in Table 1 an overview of the number of offers available in the different markets. Secondly, some items that are selling way above the Amazon price on the marketplace platform are in fact unique items (signed by the author) or collectible items (such as first edition books), even for new products. We cannot construct an automated rule to identify these items.

⁶⁴ Some Amazon.com subsidiaries may operate on Marketplace.

Therefore, some prices above the average prices will be driven by product heterogeneity rather than by price dispersion. Thirdly, some items are offered at one cent on the book and DVD markets. Including these items would artificially increase price dispersion if these offers are bogus, especially if the measure of price dispersion is based on the minimum price. We decided to drop these offers as well (see Table 1) that represent a very small portion of all offers. Fourth, about 10% of offers in some categories are the same product by the same seller on the same day and at the same or different prices. We treat these as though they were a single offer (at the lowest of the prices) with multiple copies in stock.

Table 1 shows that the local platforms (US, UK and FR) have different characteristics. The US online book, CD and DVD markets are the largest in terms of the number of sellers, followed respectively by the UK and France. The turnover in the Top 100, and thus the number of distinct items followed, is considerably larger in France than in the U.S. and the U.K. This is due to the fact that the online market share of Amazon is smaller in France than in the other countries; furthermore, Amazon Marketplace itself is less developed in France, as there are generally fewer sellers per item than in the U.K. or the U.S. It appears that Amazon is used more frequently for items other than bestsellers, resulting in high volatility of the TOP 100 list. We defined "large sellers" arbitrarily as the top 10 in each category by range of products offered. These large sellers sometimes show higher ratings and, except for French DVDs, always more ratings than other sellers. Large sellers offer, on average, lower prices than other sellers for books in the UK, CDs in the U.K. and France and DVDs in the U.S.

		Book				CD				DVD		
Markets	US	UK	FR	US	UK	FR	US	UK	FR	US	UK	FR
Number of observations	4,925,495	3,309,467	3,486,379	3,170,116	3,653,848	4,591,715	4,130,331	2,239,062	1,259,779			
Number of one cent offers	148	2,362	10	0	1	1	2,253	19	0			
Number of offers greater than twice the median price	230,461	246,092	85,416	99,767	62,53	113,049	39,456	167,184	31,673			
Number of items	819	972	2,505	563	758	3,029	685	790	1,383			
Number of sellers	19,057	6,189	1,17	4,925	3,548	691	8,208	5,392	1,128			
Daily average number of sellers per item⁶⁵	33.1	19.4	7.9	35.3	28.3	9.8	35.6	17.2	5.5			
Daily average number of large sellers	7.7	7.7	5.1	8.1	7.4	5.0	8.0	6.5	2.2			
Daily average number of other sellers	25.5	11.7	2.9	27.2	20.8	4.8	27.5	10.7	3.2			
Average rating of sellers	4.5	4.6	4.5	4.6	4.7	4.5	4.5	4.6	4.5			
Average rating of large sellers	4.6	4.7	4.5	4.6	4.7	4.5	4.7	4.7	4.5			
Average rating of other sellers	4.5	4.6	4.5	4.6	4.7	4.5	4.5	4.6	4.5			
Average number of ratings	27,348	38,392	8,843	25,891	31,156	34,779	24,078	38,776	7,701			
Average number of ratings for large sellers	69,413	82,77	10,773	74,592	75,755	63,477	74,717	43,382	4,297			
Average number of ratings for other sellers	15,635	12,114	2,376	12,543	16,424	4,901	10,244	37,234	8,588			
Average product price (in local currency)	16.66	9.49	11.72	14.95	11.08	14.71	27.37	20.70	18.71			
Average product price for large sellers	16.68	8.60	11.79	15.14	10.97	14.57	26.98	20.62	18.76			
Average product price for other sellers	16.64	10.18	11.66	14.98	11.55	14.67	27.61	20.53	17.94			

Table 1 - Dataset Overview

There was surprisingly little overlap between countries in the items appearing in the TOP 100 and thus entering our dataset, even between the U.S. and the U.K. where there is not a language barrier (Table 2).

⁶⁵ To get daily averages, we first compute a daily average across sellers for each day and each distinct product. Because we have an unbalanced panel over time, we compute a time average for each product and then an average over products, weighting each product equally. We do the same separately for large and small subsets of sellers. Because large and small sellers may not have offers for equal numbers of days, the separate results for number of vendors may not total to the average number of all vendors. Similar problems affect the computation of average prices.

	US/UK	US/FR	UK/FR	UK/US/FR
Book	30	23	15	3
	3.7%	2.8%	1.5%	0.1%
CD	71	108	190	47
	12.6%	19.2%	25.1%	1.1%
DVD	146	74	158	52
	10.6%	10.8%	20.0%	1.8%

Table 2 – Overlap between TOP 100 Goods across Countries

6.3 Posted Price Dispersion – Time Dimension

To study price dispersion over time, we construct a balanced panel in order to avoid overweighting those items that enter the data collection earlier vs. later. For our balanced panel we keep all those items that were in our dataset beginning Monday of the second week of observation and follow only these items until the end of the sample (some items have entered or dropped out of the top 100 during the first week of the period; for instance, in France, 201 different books had appeared in the top 100 by the time we fixed our sample, of which 199 remained). We have 3 dimensions to any given item in our dataset. The first dimension is the product dimension: there are different books with different ISBN, CDs with different ASIN, etc. The second dimension is the time dimension since we follow each product over time. The third dimension is related to the number of sellers of a particular item on the Amazon marketplace. The meaning of price dispersion varies according to the dimension that we focus on. In the next section we consider how price dispersion across sellers varies over time.

For each item category (book, CD, DVD) and market (US, UK, FR), we construct several useful statistics. For the average price in a market, we first compute the daily (time) average price for each pair (item, seller) and then average over sellers and then over items (in all cases prices are in local currency: dollars, pounds and euros respectively). We use a similar procedure for the standard deviation and average minimum of prices and for average and standard deviation of number of sellers. We compute the percentiles of prices using the week as the time unit. Table 3 summarizes the descriptive statistics for prices posted by sellers on the marketplace and the last 4 lines represent the total number of offers, sellers and items over the whole period and the (daily) average number of sellers per item.

We can make several interesting international comparisons. The U.S. Marketplace generally has the largest number of sellers followed by the UK and then France. We also observe that most sellers in Amazon Marketplace offer only a single item for sale. The book markets in the U.S. and the U.K. have the most sellers, whereas in France it is the DVD market; for all goods, France has many fewer sellers than in the U.S. or the U.K. The median number of days that a seller offers an item online is relatively short for all goods in the US—two to three weeks—but is more than four months for books in France. Finally, we note that while the majority of offers in the U.S. and the U.K. are from "full line" sellers that offer virtually all of the items covered in the dataset for sale, very few French sellers offer a full catalog. (An examination of the full distribution shows that with the exception of Amazon itself, a typical "large" seller in France offers no more than about half the catalog.)

Looking at prices (all prices are in local currency), we note that CDs and DVDs are far more expensive in the UK, and the U.S. has the lowest prices for these goods. Book prices are comparable across the three countries.

France has a law limiting discounting of books to no more than 5% off of the publisher's list price. As a result, we find that Amazon does not offer a significant discount compared to other sellers, and the average minimum price offer tends to be closer to the average price offer for books in France than for other goods or books in other countries. We also note that the smallest coefficient of variation is in the French DVD and book markets (both .184)—with the two lowest average numbers of sellers, and the largest is in the French CD market (.280).

	Book			CD			DVD		
Markets	US	UK	FR	US	UK	FR	US	UK	FR
Number of observations	1,176,823	615,97	445,37	959,27	932,44	440,25	1,003,551	378,65	128,47
Number of one cent offers	0	53	10	0	0	0	164	3	0
Number of offers greater than twice the median price	52,57	47,876	29,615	24,252	11,699	9,163	11,918	24,472	2,066
Number of items	98	107	199	115	122	177	98	99	97
Total unique sellers	9,961	2,666	323	2,725	1,764	310	4,176	1815	347
Average number of sellers	51.1	24.2	9.5	40.6	34.2	11.2	44.5	17.9	6.3
Standard deviation of number of sellers	29.7	6.1	3.9	8.2	10.9	5.8	11.6	5.6	3.9
Percent of sellers offering only one item	73	77	65	59	62	43	60	69	46
Median total days seller has item on offer	9	24	132	23	44	77	12	15	40
% of offers accounted for by sellers offering >75% of catalog	47	69	16	70	52	26	70	57	10
Average Amazon Price	13.58	7.01	10.79	12.72	9.02	15.38	24.66	16.85	19.34
Average price	15.09	8.03	10.93	15.14	11.07	16.39	26.58	21.34	20.06
Standard deviation of prices	3.60	2.20	2.01	3.43	3.23	4.58	5.06	5.43	3.63
Average minimum price	9.51	5.03	8.94	9.60	6.22	11.15	18.60	13.96	16.67
Median price	14.73	7.76	10.45	14.84	10.87	15.72	26.25	20.77	19.70
75th percentile	16.91	9.28	12.11	16.31	13.24	19.58	28.53	24.79	22.41

Table 3 – Descriptive Statistics of Balanced Sample

To summarize, we observe that first the average number of sellers per item is the largest in the US, then in the UK and then in France. Secondly, the number of book offers in France is greater than the number of offers for CDs and DVDs, while the US secondary markets for the three product categories are fairly similar in terms of offers (the CD market is smaller in terms of the number of sellers). Thirdly, the Amazon marketplace for CDs is very active in the UK. Finally, the average Amazon price seems fairly competitive, as it is lower than the average price posted by the other sellers on the marketplace platform.

6.3.1 Distribution of Price Changes on Amazon Marketplace

6.3.1.1 Individual Seller Price Variability over Time

The first source of price dispersion is the individual price variability arising from the fact that a seller might increase or decrease its price over time. In Table 4 we compute the number of times

sellers increase/decrease the price of its items or leave them unchanged from one day to another.⁶⁶ We report the results of our computation separately for Amazon and for the other sellers.

		Book			CD			DVD		
		US	UK	FR	US	UK	FR	US	UK	FR
Amazon	Flat	96.62	96.93	99.25	92.43	90.51	90.06	91.57	87.02	92.54
	Decrease	1.37	1.47	0.34	3.62	4.61	4.56	4.18	6.46	3.52
	Increase	2.01	1.60	0.41	3.95	4.88	5.38	4.25	6.52	3.94
Sellers	Flat	93.51	91.49	98.81	91.63	87.83	86.35	89.92	87.81	95.00
	Decrease	3.88	5.05	0.71	5.00	6.96	8.13	6.12	7.29	3.67
	Increase	2.61	3.46	0.48	3.37	5.21	5.52	3.96	4.90	1.33

Table 4 – Distribution of Amazon Price Changes (in %)

The distribution of price change is fairly symmetric for Amazon. Price remains the same 90% of the time. If we assume that the time between price changes is exponentially distributed, these results suggest a mean time between price changes from a low of every 8 days for UK DVDs, to more than 100 days for books in France. Across product markets, prices are far more stable for books than for CDs or DVDs. The distribution of price changes is more asymmetric, with more price decreases than price increases (thus prices decrease over time on the platform). Given the relatively infrequent price changes, particularly for books, individual pricing strategies cannot be the main source of persistent price dispersion. Instead, persistent price dispersion is found even when sellers do not adjust their price very often.

To test random pricing by the sellers, we computed for each product the distribution of prices across all sellers and all days and broke the distribution into four quartiles. We then ask how often the seller of a given item changes prices so as to move from one quartile to another over the 269 days of observations. We found that, the vast majority of the time, sellers do not change their prices at all or sufficiently to alter their quartile rank. Offering prices in both the highest and lowest quartiles over the observation period generally happened less than 6% of the time (Table 5). We conclude from these data that sellers do not practice random pricing.

⁶⁶ In some cases, we have some missing observations and we use the last valid observation in our computations.

	Book			CD			DVD		
	US	UK	FR	US	UK	FR	US	UK	FR
Percent of sellers x items where seller's price remains in same quartile for entire observation period	85.9	74.9	70.7	79.4	68.6	58.4	75.5	71.1	66.8
% where seller's price appears for at least one day in the highest and in lowest quartile	1.5	3.3	2.4	2.2	3.6	5.7	3.6	6.0	4.8

Table 5 – Frequency at which Sellers Price Ranking Changes Significantly

To further check whether sellers propose large price variations, we compute the standard deviation of their prices over time (weeks) for each item. We then average over the items and plot the histogram of standard deviations over the different items in each market. The results are plotted in Figure 1 for the US book market and in the Appendix for the other markets. We find that the average of the standard deviation of price changes ((mean) sd_Prix, in Figure 1) is about 15 cents in the US book marketplace, which strongly challenges the assumption of random pricing by seller.

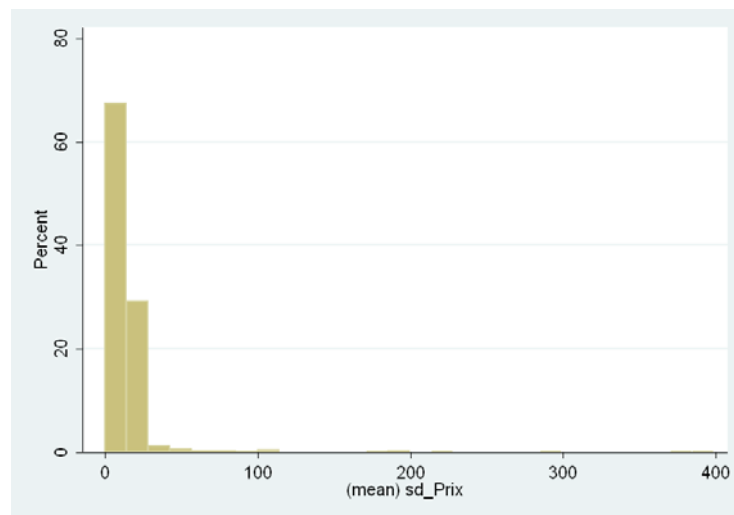


Figure 1 – Distribution of Price Changes (standard deviation, in cents)

6.3.1.2 Price Dispersion over Time

We also have found that price dispersion increases over time. That is, the longer an item is available for sale in our balanced sample, the greater the daily price dispersion. In searching for an explanation, we noted that the number of sellers for an item also tends to increase over time.

We investigated how the number of sellers for an item changes over time using a fixed effects panel regression of the number of sellers per item per day versus elapsed days since an item first appears in the Top 100 and thus enters our dataset. In general we find a positive relationship. For example, after 200 days on the market, on average the number of sellers for a book in the U.S. will have increased by roughly 14 sellers whereas the overall mean is 50 sellers (Table 6).

	Book			CD			DVD		
Markets	US	UK	FR	US	UK	FR	US	UK	FR
Dependent variable	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers	Number of Sellers
Elapsed days	0.06972***	0.01052***	0.00462***	0.04035***	0.03331***	0.00114***	0.05770***	0.01350***	0.00271***
	[0.00069]	[0.00027]	[0.00011]	[0.00056]	[0.00039]	[0.00015]	[0.00065]	[0.00035]	[0.00021]
Constant	42.55418***	22.83876***	9.05909***	36.45592***	30.98434***	11.11373***	36.85958***	17.04289***	7.42871***
	[0.10505]	[0.04027]	[0.01622]	[0.07631]	[0.05478]	[0.02038]	[0.09913]	[0.05090]	[0.03061]
R-squared	0.2998	0.05272	0.03438	0.17769	0.20992	-0.00296	0.25261	0.06151	0.00381
N	23,445	26,352	47,468	23,345	26,712	39,202	23,086	20,959	18,869

* p<0.05, ** p<0.01, *** p<0.001

Table 6 – Number of Sellers vs. Number of Days in the Market (Fixed Effects)

To determine the relative importance of time versus number of sellers on price dispersion, we regressed the coefficient of variation (expressed as a percentage) on both time and number of sellers. In Table 7 we again use a fixed effects panel regression to show that price dispersion is positively influenced by number of sellers and by elapsed time, with number of sellers having, in general, the larger effect. For example, if the number of sellers is larger by one standard deviation above the average (e.g. 29 sellers for books in the U.S. or 4 sellers for DVDs in France – Cf. Table 3) we find the coefficient of variation increases by 9 and 4.8 percent respectively. Previous empirical studies that have fixed the number of sellers investigated have missed this phenomenon.

	Book			CD			DVD		
Markets	US	UK	FR	US	UK	FR	US	UK	FR
Dependent variable	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)	Coefficient of Variation (%)
Elapsed days	-0.00270***	0.01977***	0.00521***	-0.00090**	0.02033***	0.00807***	-0.00392***	0.02003***	0.00684***
	[0.00040]	[0.00037]	[0.00027]	[0.00030]	[0.00033]	[0.00042]	[0.00031]	[0.00069]	[0.00073]
Number of Sellers	0.31173***	0.54646***	0.42187***	0.26538***	0.20389***	0.48292***	0.19560***	0.37063***	1.19991***
	[0.00323]	[0.00853]	[0.01166]	[0.00381]	[0.00494]	[0.01527]	[0.00331]	[0.01406]	[0.02541]
Constant	8.58682***	13.42176***	14.97511***	12.19362***	20.07548***	22.20086***	12.56333***	19.26073***	11.48760***
	[0.14756]	[0.20261]	[0.11394]	[0.14865]	[0.16035]	[0.18352]	[0.13474]	[0.26852]	[0.22584]
R-squared	0.34448	0.25348	0.03884	0.19345	0.27219	0.03041	0.14327	0.08345	0.10735
N	23,362	26,324	46,69	2,3105	26,619	38,527	22,729	20,508	17,866

* p<0.05, ** p<0.01, *** p<0.001

Table 7 – Coefficient of Variation (%) vs. Elapsed Days and Number of Sellers (Fixed Effects)

We hypothesize that this increase in price dispersion with number of sellers (Table 7) comes about because new sellers enter the market with offers at substantially lower prices than the median. We analyzed the price offer of sellers entering the market after our first day of observation, and compare their offer price to the median of all sellers for that item on the previous day. We find that, with the exception of the French book market, new sellers are more likely to enter the market at prices below the median, and, again with the exception of France, typically with a double digit percent discount from the prior day's median price (Table 8).

	Book			CD			DVD		
Market	US	UK	FR	US	UK	FR	US	UK	FR
Probability new entrant's price is below prior day's median price	0.79	0.71	0.47	0.73	0.67	0.57	0.68	0.71	0.72
mean percent difference from median for new sellers	-22	-16	-3	-18	-14	0.4	-14	-11	-8

Table 8 -- New Sellers Probability of Discount from Median and Mean Percent Discount

6.4 Price, Seller and Item Characteristics

We begin with the definition of a transaction on Amazon Marketplace. Next we compare transaction prices with posted prices discussed in the previous section. Finally we assess the effect of a seller's reputation and item characteristics on the prices of items that he or she sells using a simple hedonic regression.

6.4.1 Definition of a Transaction

We compute transactions on the Amazon marketplace as follows. We observe posted offers by sellers for a particular item at a given time. Among the posted offers, some will lead to transactions. A posted offer by a seller j at date t will be considered as a transaction in two cases: if the posted offer is no longer available after t , meaning that the seller does not propose it at any time after t ; or if the posted offer disappeared after t and reappeared some day after $t+2$.⁶⁷

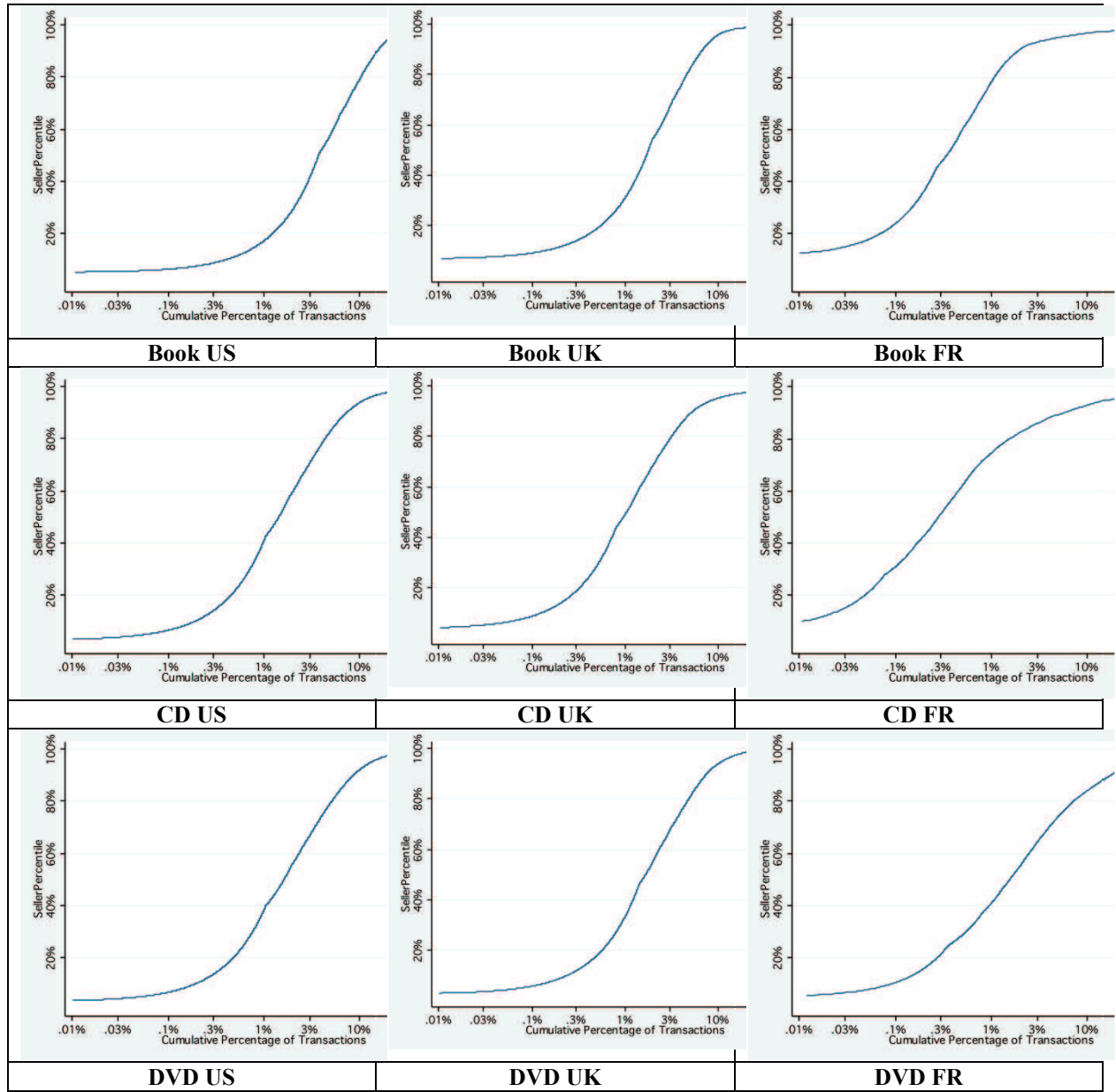
An illustration of the distribution of transactions is given in Figure 2 for every market. The figures confirm that a small percentage of sellers account for the vast majority of the transactions in all markets. Given the large number of sellers on Amazon marketplace, even 1% of sellers can mean anywhere from 6 to 190 firms, so no one firm may have significant market power.

This method has several shortcomings since it can overestimate or underestimate sales. First, the method can overestimate sales since a seller can drop his offer from the catalogue without necessarily selling it (due to pricing concerns for instance) or, conversely, the seller can cancel his offer on the Amazon marketplace because he sold the item by way of an alternative channel (physical, other website, etc.). However, in our analysis, the disappearance of this offer will be considered as a sale from the Amazon marketplace. Second, the method can underestimate sales: as long as an offer posted by a seller is available on the platform, we consider it as a posted price and not as a transaction. Now, a seller for instance can sell the item at date t and propose another offer at date $t+1$. Similarly, a seller may have multiple offers (at the same or different prices) for the same item posted on Amazon

⁶⁷ To control for possible bugs during the updating of the script, we eliminated from the database dates for which more than 7% of the items were not updated due to technical difficulties.

Marketplace. Yet, in the last two cases, as long as at least one posted offer is available, we will not record it as a transaction, thus underestimating transactions. Nevertheless, we have shown that large sellers account for most transactions so that the extent of such an underestimation is limited.

Figure 2 Distribution of Transactions over Sellers (log scale)



6.4.2 Transaction and Posted Prices: A Comparison

Table 9 (constructed in the same way as Table 3) reports descriptive statistics of transaction prices as defined above. We can compute the standard deviation of transaction prices for each item over the period and compare the results with the standard deviation of posted prices. Table 8 shows that transaction prices, as we measure them, are on average higher than Amazon's price for the same good; this challenges studies that assume that transactions are concentrated around the lowest posted price offers such as Baye *et al.* (2004). Note the small spread between average minimum and average maximum price for French books, due to the French law which limits book discounts.

	Book			CD			DVD		
Market	US	UK	FR	US	UK	FR	US	UK	FR
Average price	15.13	9.47	11.81	13.86	11.35	14.86	25.30	20.08	17.51
Average Amazon price	14.71	8.21	11.49	12.70	9.10	13.57	24.43	16.52	16.98
Average Median price	14.99	9.45	11.78	13.75	11.31	14.74	25.10	19.98	17.37
Average Minimum price	13.82	8.87	11.64	12.48	10.07	13.57	23.16	18.27	16.15
Average Maximum price	16.90	10.20	12.06	15.94	12.80	16.48	28.38	22.19	19.19
Number of transactions	150,202	94,06	89,447	129,111	124,998	130,534	177,055	108,508	41,954
Number of sellers	11,803	3,596	626	3,292	2,386	452	4,885	3,43	721
Number of items	806	923	2,356	561	706	2,854	667	760	1,308

Table 9 – Transaction Price Dispersion

In the next part, we analyze the determinants of transaction prices by looking at the reputation premium of a seller and at item characteristics.

6.5 Determinants of Transaction Prices: Seller Characteristics and Time

We now explain transaction prices (p_{ij}) of item i by seller j observed on the marketplace using the reputation of the sellers and characteristics of the items. There are several variables that capture the reputation of a seller. As in Ghose *et al.* (2009), we use both the average rating of a seller and the number of ratings that he or she received in the past. These serve as proxies for both reputation and size of operations (large sellers can better internalize fixed costs and offer better discounts). We

consider an additional reputational variable that we hypothesize is related to good will: the size of the seller. We considered two measures of size: the number of different items offered by the seller in a particular market, and the total transactions by the seller. We find these two measures to be highly correlated (typically > 0.9) and in the following use only the catalog size. Finally, we control for the time since release date ("time on market") of an item to detect any trend in prices over time.

We note that Amazon Marketplace does not provide ratings for Amazon itself, and ratings are missing for many other sellers as well—for Amazon.fr we are missing ratings for as many as 40% of the transactions. The shorter lifetime of the Amazon.fr marketplace compared to Amazon US probably accounts for fewer vendors having been rated. The distribution of missing values across vendor size appears to mirror the distribution for all transactions, so we believe that these missing values do not bias our regressions.

To estimate the effects of these variables, we run regressions with fixed item effects using the full unbalanced dataset but limited to 22 weeks because of gaps in the data collection after that date which could have introduced spurious transactions. Results are summarized in Table 10.

		Book			CD			DVD	
Market	US	UK	FR	US	UK	FR	US	UK	FR
Dependant variable	Price	Price	Price	Price	Price	Price	Price	Price	Price
Rating	28.4***	-5.41	4.21	-2.48	5.3	0.337	8.467	23.8	-3.59
Number of ratings	0.139*	-0.084	-0.156	0.421***	-0.884**	-6.35***	0.166	-3.19***	-16.0**
Catalog	0.516***	0.156***	-0.011***	0.838***	0.501***	0.0551***	1.15***	0.484***	0.0550***
Time on Market (TOM)	-0.215***	-0.170***	0.130***	0.0922**	-0.353***	-0.0188	-0.498***	0.129	1.66***
TOM ²	0.00007***	-0.00002	-0.00001	0.00004**	-0.00002	0.00005**	-0.00009	0.00011	-0.00030***
Fixed Effects by item	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1010	883***	956***	1070***	1090	1170***	2220	1430***	794***
R ² (within)	0.101	0.0434	0.0023	0.178	0.102	0.0111	0.162	0.0381	0.00861
N	149350	71191	48954	128189	105513	82762	176926	101067	29287
Number of items	792	753	1295	536	643	1726	663	707	834
Reputation Premium (in local currency) for one SD change in catalog	1.70	0.56	-0.08	1.70	1.17	0.47	0.41	1.36	0.23

Table 10 – Estimation Results

First, we find that traditional measures of reputation such as rating are not significant, except in the U.S. book market. Likewise, number of ratings is ambiguous sometimes increasing and sometimes decreasing the price premium when it is significant at all. Secondly, catalog and number of transactions are strongly correlated and better capture seller reputation: an increase of one standard deviation in catalog size is associated with price increases from .41 to 1.70 in local currency. Thirdly, the French markets differ from the UK and the US in several respects. As noted earlier, French law limits discounting for books; thus French booksellers with a large catalog are not undercut by smaller sellers as in the other markets. In addition we note that the R^2 values are smaller in all the French markets, suggesting that none of our variables have explanatory power in France. By contrast, in the U.S. CD market we explain 18% of the variance in price (ratings alone, without catalog size, explain no more than 1%). Fourthly, time since release date (ToM, measured in days) appears to be associated with declining prices for books (except in France, which is a special case because of the law), while having an ambiguous effect in the CD and DVD markets. We note that in several cases, the transaction date precedes the official release date, producing a negative value for ToM. In other cases, 40 year old classics appear in the top 100, producing ToM values over 15000. Obviously the variance of ToM across transactions will be greatest for new releases.

6.6 Conclusions

In this paper we examined comparable online marketplaces for new cultural goods in three countries over an eight-month period in 2006. We look at the structure of the online marketplace across countries and goods, and examine dispersion among posted and transaction prices.

Our first observation is that online marketplaces differ significantly across countries. The U.S., U.K. and France show widely differing roles for large online merchants, who carry an extensive catalog: there are many such merchants in the U.S., somewhat fewer in the U.K., and very few in France. In all countries, however, there is significant participation by small merchants who sell but a single good from within our sample set. These sellers appear and disappear rapidly; while the daily

number of sellers per good is modest (averaging 5-50 depending upon the market), more than 19,000 unique vendors appeared in the U.S. book market over our 8 month observation period.

When we examine how firms price, we uncover several findings. First, most sellers make infrequent changes to their posted prices. Moreover, changes that they do make are typically small. Thus over an eight month period changes in a vendor's posted price will result in a change of quartile ranking less than 1/3 of the time (with the exception of French CDs). In other words, vendors who enter the market with a high price will keep it high, while low price vendors tend to keep their price low over time. Even Amazon, the leader, makes relatively few day-to-day price changes, though it makes as many up as down. These findings contradict the hypothesis of random pricing put forward by other authors.

We also examine price dispersion, using coefficient of variation of posted prices as our measure, and find significant dynamics related to time and number of sellers. In general, the number of sellers for an item increases over time (even as its sales rank may drop). For example, after 200 days, the number of sellers of books in the U.S. will have increased by almost 25%. And price dispersion increases with both time and number of sellers. Thus an increase of one standard deviation in the number of sellers leads to as much as a 9% increase in the coefficient of variation. We show that this increase in dispersion is the result of new sellers entering the market at substantially lower prices than the previous day's median price. Previous studies that have fixed the number of sellers have missed this important effect on price dispersion. Finally, our results provide no support for traditional measures of reputation such as ratings and number of ratings and support new measures such as size of catalog and number of transactions.

To conclude, our results confirm that there are other dimensions than price competition in these markets. Reputation seems to be an important factor that questions two key assumptions of the search models: the symmetry of the sellers and homogenous goods. First, we have shown that most transactions are carried out by large sellers while a large number of small sellers realize at most two transactions. Second, it is hard to disentangle the product/seller couple which leads to product

differentiation among sellers with different degrees of reputation. As a consequence, we do not observe many transaction prices at the bottom of the price distribution, *i.e.* the best offers do not attract all the buyers. Further research needs to better characterize competition in these markets as well as seek other factors that create persistent price dispersion. The promising directions are the dynamics of entry and exit in these marketplaces (time dimension) and reputation. Since this initial attempt to compare posted prices with transaction prices shows significant discrepancies with theoretical predictions, future work should focus more on actual transactions.

6.7 References

- Acquisti, A. and Varian H., 2005. Conditioning Prices on Purchase History. *Marketing Science*, 24(3), 1-15.
- Bajari, P. and Hortacsu A., 2004. Economic Insights from Internet Auctions. *Journal of Economic Literature*, 42(2), 457-86.
- Baye, M. R., and Morgan, J., 2001. Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets. *American Economic Review*, 91(3), 454-474.
- Baye, M. R., Morgan, J., and Sholten, P., 2004. Price Dispersion in The Small And in The Large: Evidence From An Internet Price Comparison Site. *Journal of Industrial Economics*, 52(4), 463-496.
- Brynjolfsson, E. and Smith, M., 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science*, 46(4), 563-585.
- Brynjolfsson, Erik, A A. Dick, and M. D. Smith, (2010), "A Nearly Perfect Market? Differentiation vs. Price in Consumer Choice," *Quantitative Marketing and Economics*, 8(1), 1-33.
- Caillaud, B. and Jullien B., 2003. Chicken & Egg: Competition among Intermediation Service Providers. *RAND Journal of Economics*, 34(2), 309-28.

Clay, K., Krishnan, R., and Wolff, E., 2002. Prices and Price Dispersion on the Web: Evidence from the Online Book Industry. *Journal of Industrial Economics*, 49(4), 521-539

Ellison, G., and Ellison, S., 2005. Lessons about Markets from the Internet. *Journal of Economic Perspectives*, 19(2), 139-158

Ghose, A., and O. Y., 2007. Goodbye Price Dispersion? New Evidence from Transaction Prices in Electronic Markets. Third Research Symposium on Statistical Challenges in E-Commerce Research (SCECR), University of Connecticut, May.

Ghose, A., and Gu, B., 2006. Search Costs, Demand Structure and Long Tail in Electronic Markets: Theory and Evidence. Working Papers 06-19, NET Institute.

Ghose, A., and Sundararajan, A., 2006. Evaluating Pricing Strategy Using eCommerce Data: Evidence and Estimation Challenges. *Statistical Science*, 21(2).

Ghose, A., Ipeirotis, P. G., and Sundararajan, A., 2009. The Dimensions of Reputation in Electronic Markets. *Working Paper*.

Houser D. and Wooders J., 2006. Reputation in Auctions: Theory, and Evidence from eBay. *Journal of Economics and Management Strategy*, 15, 353-369.

Pan, X., Ratchford, B. T., and Shankar, V., 2004. Price Dispersion on the Internet: A Review and Directions for Future Research. *Journal of Interactive Marketing*, 18(4), 116-135.

Resnick, P. and Zeckhauser R., 2002. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System, in Baye M.R. (ed.), *The Economics of the Internet and E-Commerce*, Advances in Applied Microeconomics, Amsterdam, Elsevier Science, 11: 127-157.

Spulber D., 1995. Bertrand Competition when Rivals' Costs are Unknown. *Journal of Industrial Economics*, 43, 1-12.

Varian, H., 1980. A Model of Sales. *The American Economic Review*. 70, 651-659.

6.8 Appendix

Histograms of Price Changes in Cents (Standard Deviation Averaged over all Sellers, across
Items)

