Seller Strategies and Price Dynamics on Amazon Marketplace

Ryan Wang*

University of Chicago

May 24, 2012
Undergraduate Thesis
Advisor: Ali Hortacsu

Abstract

This paper studies seller strategies and price dynamics on the Amazon Marketplace for new books using price listings collected over a six month period. The setting naturally leads to heterogeneity in seller strategies which I seek to document. In particular, I identify and characterize distinct groups sellers among multiple dimensions of pricing strategy. In addition, I find that prices vary nontrivially over time, with a 30% difference between the highest and lowest price listed for the median textbook. Finally, I investigate to what extent Amazon's pricing behavior influences the prices listed by third-party sellers.

*Thanks to Ali Hortacsu for his general guidance. I also thank Victor Lima, Sam Peltzman, and Grace Tsiang for their thoughtful comments.

1. Introduction

The study of price dynamics is a classical topic for economists. One advantage of the Internet, among other things, is the ease with which we can collect data amenable to this topic. In particular the widespread availability of detailed pricing information allows us to empirically consider price dynamics from both the buyer's and seller's perspective.

This paper studies seller strategies and price dynamics in a relatively self-contained setting where buyers and sellers have access to a sorted list of prices for new books through Amazon Marketplace (AM). Price listings for multiple sellers across multiple books are collected over a six month period. I find that prices for the same book vary nontrivially over time, with a 33% difference between the highest and lowest price listed for the median book. Of practical interest, particularly to fellow students, this suggests gains to timing purchases of expensive textbooks, perhaps through the use of price tracking[1] or prediction websites. Using tools from cluster analysis, I identify differences in seller strategies and describe some interesting characteristics of their behavior. Finally, I find that Amazon's pricing behavior accounts for some of the variability of the lowest price through its influence on third-party sellers' behavior.

There is a copious literature that investigates price dispersion in online markets (e.g. Brynjolfsson and Smith (2000), Clay et al. (2001)), which is surprisingly persistent even in clearinghouse settings where buyers and sellers have access to a sorted list of prices for identical products (Baye et al. (2004a)).[2] Seller strategy is recognized as a key source of this dispersion, and yet there is little empirical work that documents individual seller behavior. The work that does exist has identified a rich variety of ways in which sellers avoid the Bertrand outcome of

[1]For example, the website camelcamelcamel.com provides price histories for products offered by Amazon and other large retailers.

[2]Note that price dispersion is not simply an online phenomenon. There is a rich literature dating back to Stigler (1961) that investigates price dispersion long before the advent of the Internet, and has strongly influenced current work. Other seminal papers include Varian (1980) and Borenstein and Rose (1994)

all-out price competition. Among other things, sellers engage in practices that frustrate consumer search (Ellison and Ellison (2009)), vary less salient prices such as shipping charges (Brown et al. (2010)) and conduct targeted pricing experiments (Einav et al. (2011)). Although I do not identify previously unseen behaviors, my paper is complementary to these studies in that I provide a comprehensive documentation of the strategies that appear within a single, relatively self-contained setting.

My paper also differs from studies of price dispersion in that I focus on the variability of prices over time as opposed to cross-sectional price dispersion, which is characterized by differences between competitor prices for the same product at a given point in time. Baye et al. (2004b) study temporal price dispersion, but focus on variability in the identity of the lowest-priced firm because of its implications for models of cross-sectional dispersion. Studies of retail price variation (Hosken and Reiffen (2004), Nakamura (2008)) are more closely related to my paper in that they are interested in the sources of price fluctuation. However, these studies are interested in broad variation across products, firms and locations whereas I focus on variation generated by sellers' pricing decisions. Note that the two literatures are linked by the observation that individual firms frequently offer sales or temporary discounts.

An important feature of AM is the wide range of seller types that it accomodates, from one-time sellers to large independent bookstores.[3] This naturally leads to heterogeneity in pricing strategies, which I seek to document. Studies that explicitly control for differences among sellers in other contexts (Lewis (2008) studies gasoline, Haynes and Thompson (2008) studies digital cameras) find that their results are difficult to reconcile with any single existing model. Thus I focus on characterizing regularities in seller behavior rather than testing a specific model. My paper is closest in content to Ellison and Snyder (2011), who also study a clearinghouse setting but focus on the factors that drive dif-

[3]AM accounts for about 30% of units sold through Amazon (see 2010 Annual Meeting of Shareholders). Other large retailers, including Barnes & Noble, Best Buy, Sears and Walmart have launched their own similarly designed platforms.

ferent sellers' repricing decisions.

The rest of the paper is organized as follows. Section 2 describes the data collected and features of the market setting. Section 3 outlines a guiding model. Section 4 documents interesting characteristics of seller strategies. Section 5 investigates seller strategies as a source of price variability. Section 6 concludes.

2. Data

The data consist of price listings gathered from AM between August 30, 2011 and March 5, 2012. I restrict my attention to the marketplace for new books, as opposed to used books, where there are minimal unobserved differences in product quality to account for. Price listings were gathered for three separate types of books: (1) those appearing on a New York Times Best Sellers List[4] between 5/24/2008 and 7/30/2011 (2) a selection of books that appeared on University of Chicago course syllabi during the 2010-2011 school year and (3) books that appeared on BookRags.com, a website that provides study guides and summaries for middle-school and high-school level literature. The first group covers a range of popular books, while the latter two groups cover a range of school books from textbooks to literature. Books are identified by ISBN.

Each listing contains the price offered, some seller information, and an optional text description. Displayed seller information includes the seller's username, overall rating, number of ratings and location. Listings are identified by the seller's username and the book's ISBN. By default the price list for each ISBN is sorted in increasing order by price plus shipping cost. The default shipping cost is $3.99 as of April 2012.

It is free to list on AM but a commission is charged in the event of a sale. For books, the commission is 15% of the sale price, plus a $1.35 closing fee, plus

[4]Lists are released weekly. They are further subdivided into categories, which include Hardcover Fiction, Hardcover Nonfiction, Trade Fiction Paperback, Mass Market Paperback, Paperback Nonfiction, Hardcover Advice, Paperback Advice, Picture Books, Children's Chapter, Children's Paperback, Children's Series, Hardcover Graphic, Paperback Graphic, Manga, Hardcover Business, Paperback Business and Hardcover Political.

a \$.99 closing fee, which can be waived by paying a monthly subscription fee. As of April 2012 a subscription costs \$39.99 per month. Amazon also provides a shipping credit, \$3.99 by default, which is added onto the price paid by the buyer. Thus a seller receives the sale price plus shipping credit minus commission. I work with the price plus default shipping cost since this is the price that most buyers would face. Figure 1 below provides a screenshot of the price list for Barro's Macroeconomics textbook. The features just discussed are highlighted by boxes.

Some sellers participate in a program, "Fulfillment By Amazon" (FBA), that allows a listing to be displayed with zero default shipping cost. Such listings are prominently labelled and are eligible for Amazon's various shipping programs, even when bundled with other Amazon products. The program allows sellers to send inventory to Amazon, which then handles fulfillment in the event of an order.[5] Sellers are charged for storage by volume, a per-unit handling fee by weight, and a fixed fee. Amazon's own price listings are treated as FBA on the price list. The existence of the FBA label is useful because it can be confidently assumed that such listings are genuine, since sellers must pay for them and Amazon holds the inventory. Later I use FBA as a benchmark for the results from clustering.

Although I do not have data on actual transactions, I suspect that the sellers with the lowest prices capture a majority of the sales for two reasons. First, AM is reached by clicking a link from the product page, which provides product information and advertises Amazon's own price.[6] This suggests that the type of buyers that reach AM are more price-sensitive. A quality-sensitive buyer could be highly confident of receiving a high-quality product within the stated shipping range by purchasing from Amazon directly. Second, empirical evidence from price comparison websites suggests a discontinuity in demand in such

[5]Note that FBA is a listing-specific characteristic as opposed to a seller-specific one. That is, a seller may have both FBA and non-FBA listings.

[6]This front page display is called the "buy box" in other product categories third-party sellers can capture the buy box through a combination of high reputation and low price.

settings. Baye et al. (2009) estimate that a seller can gain 60% more clicks by offering the lowest price. Note that the price comparison website they study redirects buyers to the seller's independent website whereas, buyers on AM can purchase from a third-party seller without leaving Amazon's website.

The data contain a number of listings with nonsensical or unreliable prices because it is free to list. I ignore outliers by removing listings with price greater than twice the median listing price for the same ISBN on the same day. I also remove listings with price greater than twice the list price. There are a number of listings where third-party prices have converged to $0.25 or below, not including shipping. It is impossible to study pricing strategies in such cases since prices have nearly reached a minimum. Thus I remove ISBNs where the lowest price is less than $0.25 more than half of the time. Finally, I remove ISBNs for which I suspect the majority of listings may be unreliable. First, I remove ISBNs where Amazon never lists a price. Second, I remove ISBNs with less than two sellers on average. Lastly, I remove ISBNs whose average Best Sellers Rank is above 1,000,000. A high rank corresponds to lower volume. Amazon does not provide much detail regarding how ranks are calculated, but I suspect that ISBNs with such high ranks are only sold rarely. Note that the median rank is about 75,000, so the ISBNs removed are relatively low volume.

Because listings are observed only semi-daily, all averages are weighted by the number of days elapsed since the last observation. Standard deviations are also weighted. Thus,

$$\bar{x} = \frac{\sum_t n_t x_t}{\sum_t n_t}$$

and

$$s^2 = \frac{\sum_t n_t (x_t - \bar{x})}{(\sum_t n_t) - 1}$$

where t indexes the observation, n_t indicates the number of days elapsed since observation $t-1$ and x indicates the variable of interest. For example, $t = 1$ may correspond to 8/30/2011 while $t = 2$ may correspond to 9/3/2011. This method of weighting is equivalent to assuming that the variable remains constant in the days between observations.

Table 1 shows summary statistics at the ISBN level. Prices for 5,273 ISBNs are included in the sample. There are 47,148 sellers observed at some point in the sample, ranging from one-time sellers to large independent bookstores such as Powell's Books. Of the 9,341 sellers that change prices at least once, 496 of them account for 95% of all observed price changes. I will focus on this group since I am interested in sellers' pricing strategies. Summary statistics for this group are given in Table 2.

2.1. Price Variability

Since transaction prices are not observed I focus on the variability of the lowest listed price for a given ISBN. The lowest price can be interpreted as the price faced by a fully price-sensitive buyer. Importantly, it is attainable. I also consider robustness to alternatives such as the lowest price listed by sellers of a certain size[7] and the average of the three lowest prices. I also consider multiple ways to measure price variability - (1) the range, defined as the percent difference between the minimum and maximum price over time and (2) the coefficient of variation (CV), defined as the standard deviation of prices over time divided by the mean. Each measure has its advantages. The range is more easily interpretable while the CV makes more efficient use of the data collected. In particular the range may be misleading if either the maximum or the minimum lowest price is short-lived while the CV has no such issue. My results are largely robust to the above choices.

The magnitude of variability I observe is surprisingly high. For the median book the range is 33.4%. That is, a buyer who purchased this book at the worst possible time would have paid 33.4% more than a buyer who purchased at the best possible time. This corresponds to a small dollar amount in the case of a paperback but might be more salient in, for example, the textbook market,

[7] In particular, the top quartile of sellers in terms of size. These sellers have received at least 35 ratings.

where prices are routinely greater than \$100. For the median book from a UChicago Syllabus, the range is 35.6%. Considering only prices listed by the top quartile of sellers in terms of size, the median book still has a range of 34.5% and the median textbook a range of 30.2%. See Figure 2 below for a histogram of price variability by type of book. Later, I give evidence that this variability is at least partially explained by the composition of sellers.

Amazon's prices also fluctuate over time, with its price changing at least once for 60% of the ISBNs observed. Of the ISBNs where prices change, the median range is 13.6% and, surprisingly, 53% have a higher price at the end of the observation period than at the beginning. This suggests that the observed price variability is not simply the result of prices uniformly decreasing over time. See Figure 3 below for an example plot of the price path for Wooldridge's textbook on Introductory Econometrics. This figure is representative of a common pattern. In most cases prices oscillate up and down but do not seem to trend in a particular direction.[8] That said, there are also cases, for example recent NYT Bestsellers, where prices steadily decrease.

Note that I have limited data pertaining to changes in market fundamentals that would influence price movements. In particular I lack any data on supply or even wholesale costs. Thus one presumably common strategy that I cannot identify is for a seller to maintain a constant markup. On the demand side I have data on the Amazon Best Sellers Rank, which is based on recent and historical sales. Though Amazon does not explain the exact methodology used to calculate Best Sellers Rank, its fluctuations may nonetheless act as a proxy for fluctuations in demand.

[8]Price tracking websites show similar patterns on a much larger scale. See for example camelcamelcamel.com, which collects prices from Amazon and other retail websites.

3. Determinants of Price Strategy

Here I discuss the determinants of seller price strategies, both theoretical and empirical, in order to provide a guiding framework for the rest of the paper. There are three broad types of factors discussed in the literature - competitive, market, and seller. (STUB I discuss these in turn) Although competitive factors could be viewed as a subset of market factors, I make the distinction because market factors might also influence entry and exit considerations whereas competitive will influence the regularity with which sellers reprice.

Dynamic pricing models suggest that sellers are sensitive to competitor actions in a number of ways. First, competitive repositioning may influence repricing decisions. Second, the recognition of rival responses to a seller's own actions might effectively deter frequent repricing (**?**).

Baye et al. (2004a) propose a well-studied model for markets where a price clearinghouse is active. In this setting buyers and sellers have access to a sorted list of prices. Here I discuss some of the model's implications in order to provide a guiding framework for the rest of the paper. The model generalizes earlier work (e.g. Varian (1980)) that rationalizes equilibrium price dispersion in offline markets. The distinguishing feature is that identical sellers target two types of customers and the incentive to price discriminate encourages randomized pricing behavior.

Suppose that $n > 1$ sellers compete in a market for an undifferentiated product. Suppose also that a clearinghouse exists through which they can advertise a price. There are two types of buyers and each buyer has unit demand and reservation price, r. There are $S > 0$ price-sensitive 'shoppers', who check the clearinghouse and buy at the lowest price as long as it does not exceed r. When multiple sellers tie with the lowest price the buyer chooses a seller randomly. There are $L \geq 0$ price-insensitive 'loyals' per seller, who are loyal to that particular seller and buy from them as long as the price does not exceed r.

In equilibrium sellers that target shoppers utilize randomized pricing strategies in the sense of drawing prices from a distribution at every time period. To see why they price randomly, suppose sellers were to utilize a predictable pricing strategy. Then any one seller could charge a price arbitrarily smaller than the lowest priced competitor and capture the entire market for 'shoppers' while suffering an arbitrarily small loss from its market for 'loyals'. The alternative is to abandon shoppers altogether and price uniformly at a high level. The randomized pricing strategy thus prevents systematic exploitation by rivals while permitting price discrimination between loyals and shoppers over time.[9]

Finally there are seller-specific factors. Differences in these factors manifest as differences in seller strategy, but there may be underlying reasons for favoring one pricing strategy over another, as discussed. In particular, small firms tend to reprice more often while large firms reprice less frequently and price at higher levels.

4. Characteristics of Seller Strategies

In this section I document interesting characteristics of seller strategies. As in Ellison and Snyder (2011), I find heterogeneity along multiple dimensions of pricing strategy, including the frequency of price change and the targeted price rank. This is unsurprising given that AM accomodates a wide range of seller types, from one-time amateurs to independent bookstores. Using tools from cluster analysis, I also identify relatively distinct groups of sellers. The first group targets low ranks and frequently changes prices, the second group targets high ranks and infrequently changes prices and the third group targets high ranks and frequently changes prices. A fourth group consists of sellers excluded from the cluster analysis because they change prices too infrequently. My purpose here is largely descriptive since the ability to freely post prices gives rise to such a wide variety of sellers.

[9]Baye et al. (2004a) show that it is optimal for the clearinghouse to give buyers free access but charge sellers.

4.1. Cluster Analysis

Given that pricing strategies may be complex, I seek to capture similarities across multiple dimensions by using tools from cluster analysis. I largely follow the method of Ellison & Snyder (2011), who consider sellers in a clearinghouse market for memory-modules.[10] The exercise is descriptive in that there is no true underlying classification to be compared against. I consider the following seller-level variables averaged over multiple listings: (1) target rank, defined to be the average rank of a given listing immediately following a price change (2) frequency of price change, defined to be the average number of days between price changes (3) number of ranks bumped, defined to be the average number of ranks a seller has been forcibly moved by competitor activity prior to a price change (4) size of price change, given in percentage change, and (5) synchronization of price changes, defined to be the average fraction of a seller's inventory that is changed on the same day. The first four variables are also considered in Ellison and Snyder (2011) but the fifth is not. I intentionally consider variables that capture aspects of pricing strategy directly controlled by sellers.

Clustering is performed using the R implementation of k-means clustering. I first transform each of the variables to the log scale. This is reasonable because the procedure calculates Euclidean distances between sellers and I do not want, for example, the difference between rank 39 and 30 to receive the same weight as the difference between rank 10 and 1. Variables are recentered and rescaled to have zero mean and unit variance.[11] Since I am interested in variables related to price changes, I require multiple observations for a given seller. Thus

[10] Ellison & Snyder (2011) cluster on seller-level averages of the following variables: (1) average rank immediately following a price change, (2) average placement immediately following a price change, (3) average number of ranks bumped since last price change, (4) average number of hours since last price change and (5) variance of number of hours since last price change, (6) total time present on the list, (7) amount of time spent in ranks 1-12 as a fraction of time present on the list.

[11] One potential issue is that the variables I choose to cluster on are not independent, and thus I may be over-weighting redundant information. The maximum pairwise correlation, between frequency of price changes and number of ranks bumped, is 0.80. The resulting clusters remain largely the same if number of ranks bumped is left out.

I exclude sellers with too few price changes from the clustering. In particular, I restrict my attention to the 99th percentile of sellers in terms of number of price changes. The remaining 496 sellers account for 95% of all observed price changes in the sample. Since the number of clusters, k, must be specified as well, I increment k until I begin to identify redundant clusters. Using $k = 3$ yields a qualitatively parsimonious classification. See Table 3 below containing averages for each cluster. See Figure 4 below for pairwise scatter plots of the variables considered, colored according to the resulting clusters. The table and figure both suggest that the method successfully identifies some separation along these dimensions.

Table 3 illustrates characteristics of the resulting clusters, which are named accordingly. The low rank, frequent price change cluster seems to be characterized by sellers who populate the lower ranks, in other words maintain lower prices, and make relatively small price changes relatively frequently. This group seems to drive the price variability on AM, as will be discussed later. The fact that a large proportion of FBA sellers fall in this cluster is encouraging. Such sellers have a monetary incentive to sell their products quickly and I would expect them to target low ranks.[12] One interpretation is that these sellers are market followers in the sense that they must frequently reprice in order to maintain low price ranks.

High rank, infrequent price change sellers populate the higher ranks and make relatively large price changes relatively infrequently. Moreover they only reprice a small percentage of their inventory on the same day. Interestingly, these sellers are, on average, relatively large in terms of number of ratings. Note that a seller can only receive a rating after a transaction is completed, so the number of ratings is a lower bound on the total number of completed transactions. This is somewhat puzzling if the lowest ranks accrue a disproportionate percentage of sales on AM as they do in the clearinghouse setting studied by

[12]Recall that the FBA listing applies to listings, so that the same seller may have both FBA and non-FBA listings. Here I identify an FBA seller as a seller where more than 10% of its listings are FBA. This excludes sellers who may have briefly tried the program.

Baye et al. (2009). How do these sellers sell any books when their prices are, on average, the 17th lowest? One possible explanation is that these sellers simply offer a large number of books, listing low prices for books they hold in inventory and high prices for books that they would acquire in the unlikely event of an order. Figure 5 shows a histogram of average price ranks for listings posted by sellers in this cluster, and indeed there appears to be some sort of bimodality. The average is taken over time for each listing. High rank, infrequent price change sellers post a relatively large number of listings in the first place, on average 1,000 amongst the observed ISBNs. Thus a large number of books are still listed at low ranks, or low prices.

High rank, frequent price change sellers also populate the higher ranks but change prices nearly as frequently as low rank, frequent price change sellers. They also reprice a large percentage of their inventory at once on the same day. Figure 5 suggests bimodality in the average price ranks for this cluster as well. Since these sellers post nearly the same number of listings as high rank, infrequent price change sellers, the explanation given above may apply to this cluster as well. An alternative explanation is that these sellers practice the "hit-and-run" pricing predicted by Baye et al. (2004b) in order to prevent systematic undercutting by competitors. That is, they regularly post a relatively high price but periodically cut to a relatively low price. In Baye et al. (2004b), this strategy is driven by an incentive to discriminate between loyal and price-sensitive buyers. However, it seems that sellers in this cluster are minimally branded and the largest sellers fall in the second cluster - high rank, infrequent price change. The fact that these sellers are able to synchronize changes across their relatively large inventories also suggests some level of sophistication. One possibility is that they price on multiple platforms and thus, seek to discriminate between buyers on different platforms as opposed to loyals.

One concern with the clustering methodology is that seller pricing strategies might not be uniform. That is, the same seller may price one of its listings very differently from another. Indeed, as discussed previously there are a number of sellers that have both FBA and non-FBA listings. Different market conditions for different books might influence sellers to price differently depending on the listing. The cost of monitoring market conditions may also influence sellers to actively reprice only their most profitable books. Clustering at a finer level would arguably take better advantage of the panel structure of my data, as well.

4.2. Other Seller Characteristics

Although the lowest price is quite variable, individual sellers do not continually change prices. Even sellers in the first cluster change prices every 10 days, on average. Davis and Hamilton (2004) study price inertia in the context of wholesale gasoline pricing and argue that sellers face significant non-administrative costs associated with changing prices. In particular they suggest that these costs reflect strategic considerations of both customer and competitor responses. These considerations seem less pertinent in the AM setting. Sellers are largely unbranded, precluding customer responses, and competitor responses seem inevitable given the large number of sellers active. Instead, sellers' price changes may be influenced by the costs of continually monitoring and responding to market conditions, as put forth by Ellison and Snyder (2011).

Indeed, sellers with more listings seem to change prices less often. I fit the following OLS regression model:

$$\text{log(Days Bet. Change)}_i = \beta_0 + \beta_1 \text{Synchronization}_i + \beta_2 \text{log(Num. of Listings)}_i$$

where i indicates seller. Results are reported in Table 4. The coefficient on number of listings is positive and statistically significant. In particular, a 10% increase in a seller's number of listings roughly corresponds to a 1.5% increase in the average number of days between price changes. Sellers with more listings to monitor change prices less often. Moreover, the coefficient on synchronization is negative and statistically significant. A unit increase in synchronization, that is a 1% increase in the average fraction of inventory changed on the same day, corresponds to a roughly 2.6% decrease in the average number of days between price changes. One plausible explanation is that some sellers use automated pricing algorithms, which allows them to simultaneously change a large fraction of their inventory at once.

Automated pricing software does exist, though standard programs seem to be limited in their sophistication.[13] A search for the term "amazon repricing" reveals a number of programs that advertise features for strategic pricing. For example, a program called RepriceIt includes the ability to compare against competing prices, to take into account competitor characteristics such as rating, and to specify the number of competing prices to compare against. This particular program costs at least $9.95 a month, with prices increasing in the size of the inventory to be managed. However, the types of pricing algorithms used by, for example, Amazon likely require a high level of technical sophistication and significant capital investment.[14] Even with the transparency of prices listed on AM, there are nontrivial costs that scale with the number of products to be monitored.

Finally, I find evidence that some large sellers may charge different prices across different platforms. I manually examine the largest observed third-party sellers sorted by number of ratings and identify a number of associated websites. Websites were verified by matching seller descriptions and locations on Amazon to those on the website. Prices for 100 ISBNs are manually compared. I do not attempt to standardize the ISBNs checked on each website, since different sellers have different inventories and the purpose of the exercise is simply

[13]See http://www.michaeleisen.org/blog/?p=358 for an amusing case of automated pricing gone awry.

[14]A look at Amazon's job postings is instructive. Not only does it implement automated pricing algorithms, but it also runs large-scale pricing experiments, which may perhaps account for some of the unexplained variation in observed prices. A posting for the title, 'Senior Software Engineer, Pricing' includes the following description: "The Pricing systems are responsible for determining and publishing prices automatically, with little to no manual intervention, for the millions of items that Amazon sells worldwide ranging from Books to Consumer Electronics to Shoes." Another posting for the title 'Pricing Analytics - Software Development Engineer' includes: "The Pricing Analytics team builds the infrastructure to run large-scale pricing experiments and simulations on millions of products ranging from books to electronics to shoes."

to check for the existence of discrepancies. Instead they are chosen by randomly selecting 100 ISBNs for which a given seller has listed prices on AM. Surprisingly, some discrepancies do exist. See Table 5 for a list of such sellers and the average percentage difference between prices on their independent website and on AM. A simple explanation is that this is done to account for differences in listing or commission fees. Nonetheless, it is surprising that a seller would persistently charge different prices across different platforms. Of course, I do not rule out the possibility that this price discrepancy is simply an empirical curiosity.

5. Sources of Price Variability

Here I study the sources of price variability at the seller-book level. Although I previously clustered seller strategies using seller level averages, it seems natural to expect that sellers take market conditions into account when setting prices.

In particular I focus on the extent to which Amazon's price changes drive the variability of the lowest price. Although Amazon tends to list relatively low prices, the lowest price on AM is generally listed by a third-party seller. The primary result is shown in Figure 6 below. To avoid overplotting, the variable on the x-axis is produced by rounding values of Amazon's CV to the nearest digit and binning. The variable on the y-axis is then produced by calculating the median lowest price CV for ISBNs in each bin, with the interquartile range indicated by bars. The plot suggests that the lowest price is more variable when Amazon's price is more variable, but tapers off when Amazon's CV is greater than about 0.4. Moreover, Amazon's distinctive leadership position within AM suggests that its price changing decisions influence third-party sellers and not the other way around.

To formally consider this relationship, I fit the following OLS regression model:

$$\text{Variability}_i = \alpha_{S(i)} + \beta \text{ Controls}_i + \gamma \text{ No Amazon Change} + \delta \text{ Amazon's CV}_i$$

where i indexes the ISBN. $S(i)$ indicates the source of the ith ISBN and constitutes including dummy variables for each source - Bookrags, NYT Bestseller or UChicago Syllabi. I further distinguish between different NYT Bestseller Lists. Controls indicates ISBN-level characteristics that I control for and includes list price, star rating, average Best Sellers Rank, standard deviation of Best Sellers Rank and number of sellers, broken down by type according to the cluster analysis. I view Best Sellers Rank as a proxy for demand, and thus include its standard deviation to control for changes in demand. The number of sellers is broken down by type. For example, if there are 20 sellers listing prices on average, the breakdown by type might be (10,5,5,5). In this case there are 10 sellers from the low rank, frequent price change cluster.

The results from estimating the regression model are given in Table 6. Models 1 and 2 consider the lowest price on Marketplace but for different measures of price variability. Model 3 considers the coefficient of variation for the average of the 3 lowest prices. The response is multiplied by 100 for readibility so that, for example, $\%$ Difference$_i = 25\%$. Note that the coefficients for models 1 and 2 are not directly comparable, though the models are consistent in the sense that relative magnitudes of the coefficients remain the same.

The explanatory variables of interest is the coefficient of variation for Amazon's price over time. The coefficient on Amazon's CV is positive and statistically significant in all three models, though it is smaller in model 3 than in model 2. This suggests that when Amazon's price is more variable, third-party sellers largely follow suit so that the resulting lowest price is also more variable. Note that the plot suggests some nonlinearity in that there is no effect from increasing Amazon's CV above a certain level, around 0.4 or 0.5. Thus it may not be appropriate to give a direct interpretation of the coefficient on Amazon's CV. One alternative possibility is to fit a piecewise linear model, though care must be taken in choosing the breakpoint.

An interesting thought experiment is to consider what relative magnitude a cleanly estimated version of this coefficient might have in markets for other products, or perhaps in the marketplaces recently implemented by other retailers. I imagine that it might influence the marketplace operator's, whether Amazon's or some other retailer's, decision on how vigorously to compete as opposed to collecting commissions from third-party sales. In particular the degree to which variability in the lowest price is driven by variability in the operator's price may gives an indication of the operator's ability to influence the lowest price, which it can advertise to buyers.

Coefficients on the statistically significant ISBN-level control variables have the expected signs. As expected, prices are also more variable when demand, as proxied by the Best Sellers Rank, is more variable. Unsurprisingly, prices are also more variable as the number of low rank, frequent change sellers increases. This provides another form of validation for the clustering, as these are the sellers that would tend to list the lowest price. The coefficients for both clusters of high rank sellers change signs depending on the model. The average Best Sellers Rank, average star rating, and list price do not have significant effects on variability.

Note that in the above regression model, Amazon's price variability is determined exogenously. I am largely interested in the effects on third-party sellers within the relatively self-contained environment of AM, in which case the exogeneity assumption likely provides a good approximation. However a richer model would other important factors into account. Amazon uses sophisticated pricing algorithms and although I have controlled for demand and certain book characteristics, it is likely that there are influential characteristics that I have missed. In particular Amazon uses a detailed, hierarchical categorization scheme for books, which I ignore. More importantly, Amazon itself competes with large, branded retailers such as Barnes and Noble or Walmart so that I do not account for higher level strategic interactions.

6. Discussion

The features of Amazon Marketplace naturally give rise to seller heterogeneity, given the wide range of seller types that AM accomodates. Using tools from cluster analysis I identify and characterize three groups of seller strategies differentiated primarily by how often prices are changed and the price levels targeted. Note that the strategies I observe were not previously unknown and indeed are quite mundane in and of themselves. See, for example, Noble and Gruca (1999) for a review of pricing strategies from the theoretical literature, which includes those that I observe. My paper's main contribution is to study the context and prevalence of strategies within a single, relatively self-contained setting.

In addition I find that prices, both the lowest third-party price and Amazon's price, vary nontrivially over time. There is a 30% difference between the highest and lowest price listed for the median textbook. Of the books for which Amazon changes prices, the median range is 13.6%. Prices are not uniformly decreasing, either, with more than half of the these books having a higher price at the end of the observation period than at the beginning. This suggests that there may be gains to timing purchases of expensive books. In particular patient buyers can save money by consulting price tracking or even price prediction websites.[15] Finally, I find that Amazon's own pricing practices account for some of the variability of the lowest third-party prices. This reflects its clear leadership position within AM.

There are a number of interesting issues that I have not addressed in this paper. In particular, how do sellers form pricing strategies? To what extent are these strategies fixed across listings and over time? For example, Einav et al. (2011) find that independent sellers on eBay conduct targeted, small-scale experiments. This illustrates the richness of individual seller behavior, of which we still know relatively little about.

[15] A website called Decide.com provides price predictions for expensive electronics and appliances, as well.

References

Baye, Michael, John Morgan, and Patrick Scholten, "Price Dispersion in the Small and in the large: Evidence from an Internet Price Comparison Site," *Journal of Industrial Economics*, 2004, *52* (4), 463–496.

__ , __ , and __ , "Temporal Price Dispersion: Evidence from an Online Consumer Electronics Market," *Journal of Interactive Marketing*, 2004, *18* (4), 101–115.

__ , Rupert Gatti, Paul Kattuman, and John Morgan, "Clicks, Discontinuities, and Firm Demand Online," *Journal of Economics and Management Strategy*, 2009, *18* (4), 935–975.

Borenstein, Severin and Nancy Rose, "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 1994, *102* (4), 653–683.

Brown, Jennifer, Tanjim Hossain, and John Morgan, "Shrouded Attributes and Information Suppression: Evidence from the Field," *Quarterly Journal of Economics*, 2010, *125* (2), 859–876.

Brynjolfsson, Erik and Michael Smith, "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 2000, *46* (4), 563–585.

Clay, Karen, Ramayya Krishnan, and Eric Wolff, "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," *Journal of Industrial Economics*, 2001, *49* (4), 521–539.

Davis, Michael and James Hamilton, "Why are Prices Sticky? The Dynamics of Wholesale Gasoline Prices," *Journal of Money, Credit and Banking*, 2004, *36* (1), 17–37.

Einav, Liran, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan, "Learning From Seller Experiments in Online Markets," *NBER Working Paper No. 17385*, 2011.

Ellison, Glenn and Sara Ellison, "Search, Obfuscation and Price Elasticities on the Internet," *Econometrica*, 2009, *77* (2), 427–452.

Ellison, Sara and Christopher Snyder, "An Empirical Study of Pricing Strategies in an Online Market with High-Frequency Price Information," *Working Paper*, 2011.

Haynes, Michelle and Steve Thompson, "Price, Price Dispersion and Number of Sellers at a Low Entry Cost Shopbot," *International Journal of Industrial Organization*, 2008, *26* (2), 459–472.

Hosken, Daniel and David Reiffen, "Patterns of Retail Price Variation," *RAND Journal of Economics*, 2004, *35* (1), 128–146.

Lewis, Matthew, "Price Dispersion and Competition with Differentiated Sellers," *Journal of Industrial Economics*, 2008, *56* (3), 654–678.

Nakamura, Emi, "Pass-Through in Retail and Wholesale," *American Economic Review: Papers & Proceedings*, 2008, *98* (2), 430–437.

Noble, Peter and Thomas Gruca, "Industrial Pricing: Theory and Managerial Practice," *Marketing Science*, 1999, *18* (3), 435–454.

Stigler, George, "The Economics of Information," *Journal of Political Economy*, 1961, *69* (3), 213–225.

Varian, Hal, "A Model of Sales," *American Economic Review*, 1980, *70* (4), 651–659.

Figure 1: Screenshot of Amazon Marketplace

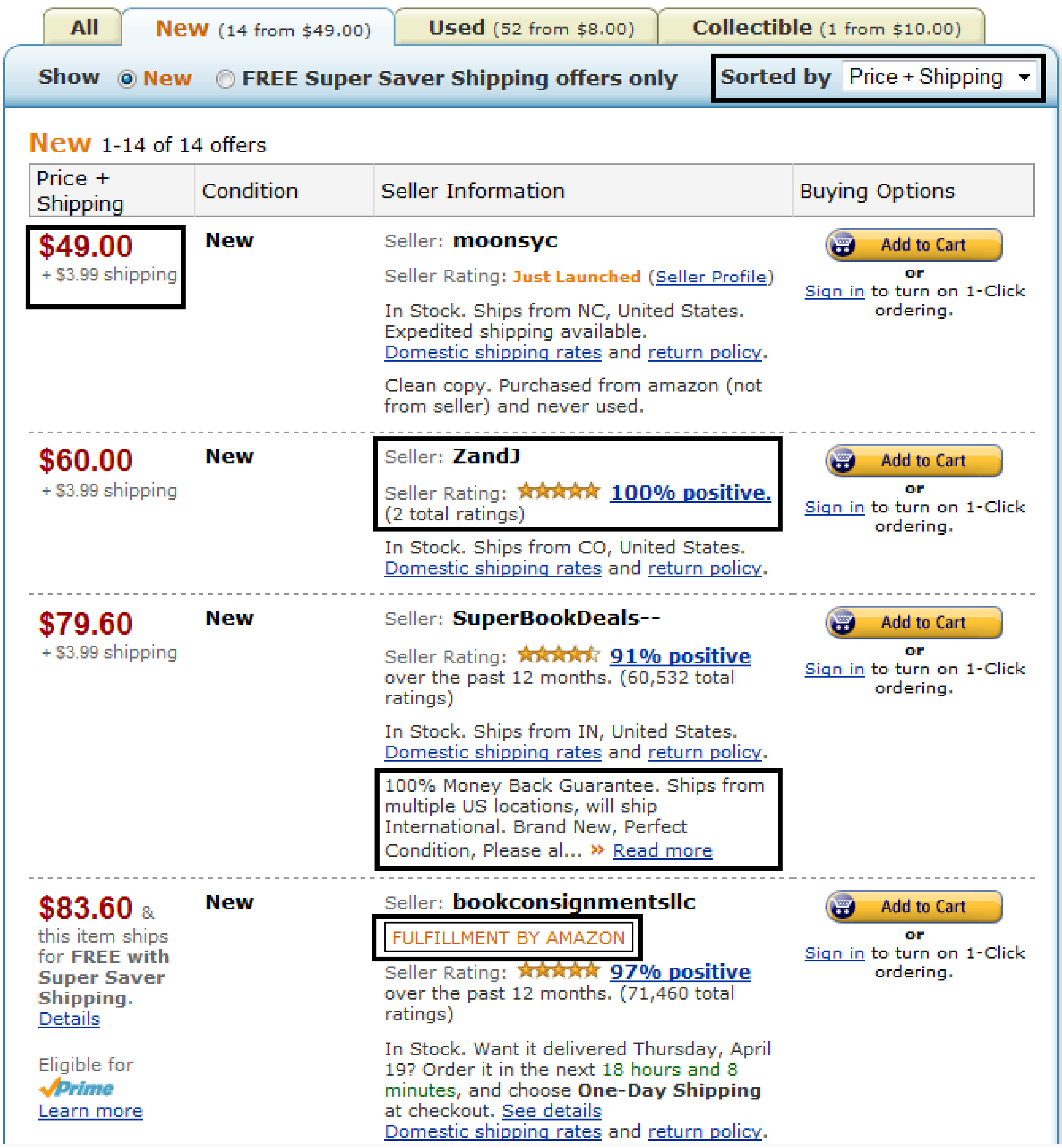


Figure 2: Percent Differences by Type

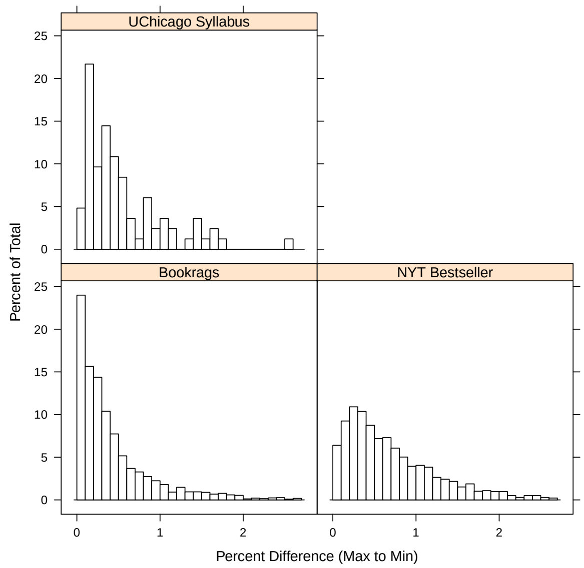
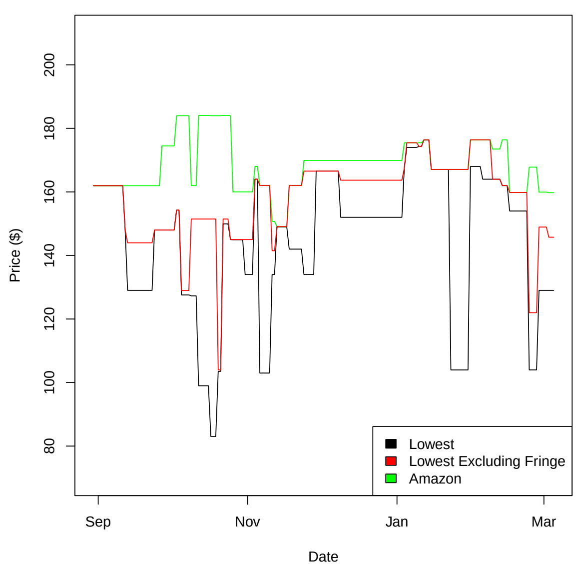
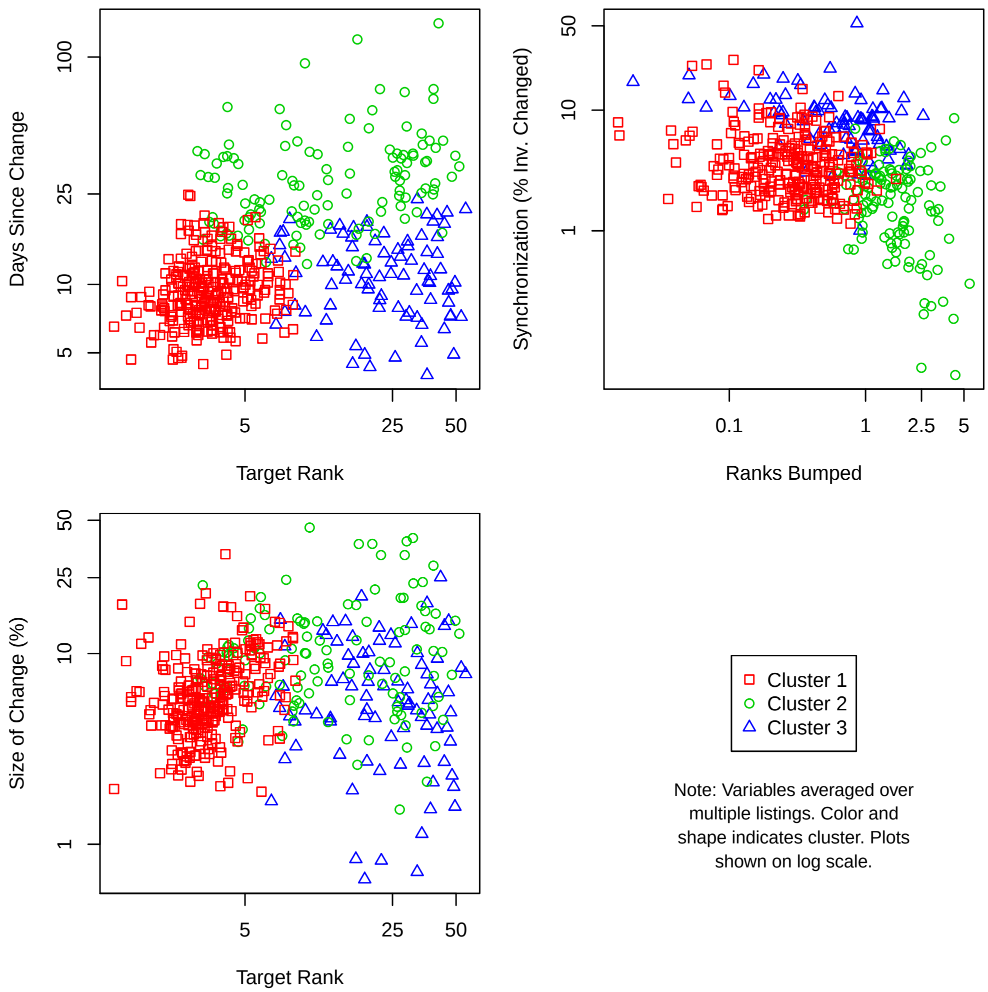


Figure 3: Price Path for Introductory Econometrics by Wooldridge



Note: Prices are observed semi−daily. Missing days are filled in assuming prices do not change.

Figure 4: Pricing Strategy with Clusters



Note: Variables averaged over multiple listings. Color and shape indicates cluster. Plots shown on log scale.

Figure 5: Histograms of Price Ranks by Cluster

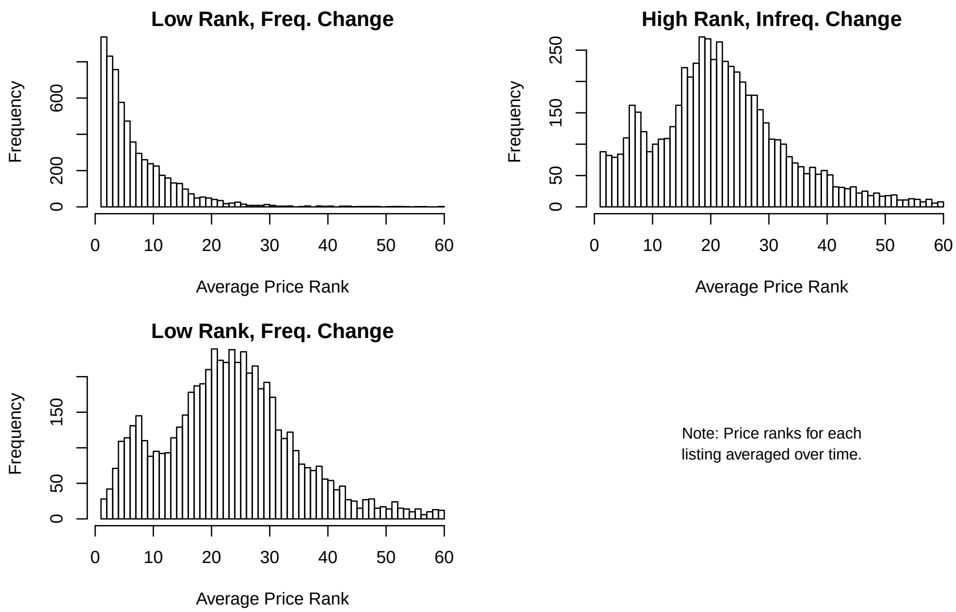
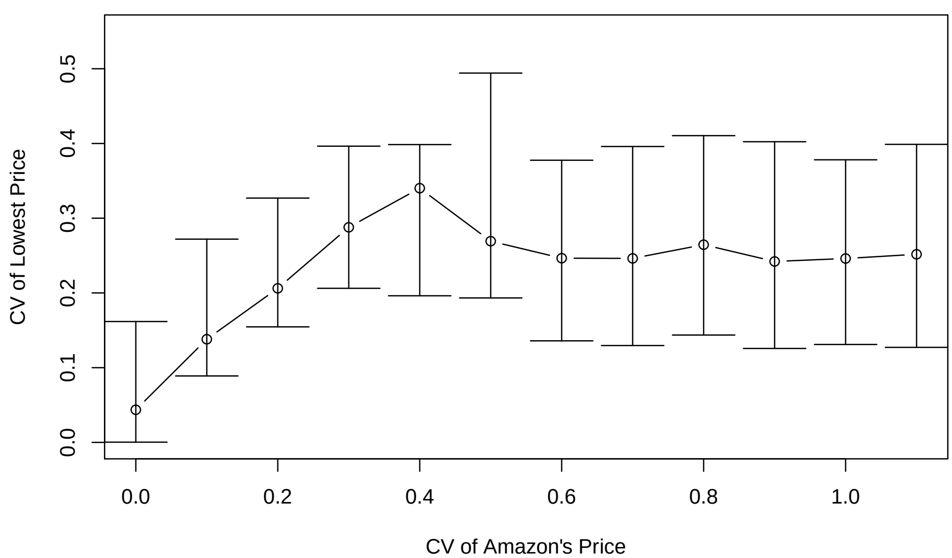


Figure 6: Variability of Lowest Price vs. Amazon's



Note: Collapsed by rounding CV of Amazon's price to 1 digit. Bars indicate interquartile range.

Table 1: Summary Statistics for ISBNs (Averaged Over Time)

	Mean	Std. Dev.	Median	Min	Max
List Price	$24.2	$19.4	$23.95	$5.99	$258.95
Best Sellers Rank	141,776	176,899	75,354	7	997,740
Number of Sellers	40.0	22.0	37.3	2.0	299.4
Number of Sellers (Top Quartile)	30.0	10.4	30.7	1.5	106.8
Range (Lowest Price)	53.2%	62.9%	34.2%	0.0%	1,140.0%
CV (Lowest Price)	24.6%	21.5%	19.6%	0.0%	166.7%
CV (Top Quartile)	18.1%	16.1%	14.4%	0.0%	168.9%
CV (Avg. of 3 Lowest)	17.5%	22.5%	13.5%	0.0%	62.1%

* N = 5273

Top Quartile refers to the top quartile of largest sellers in terms of size.

Table 2: Summary Statistics for Sellers

	Mean	Std. Dev.	Median	Min	Max
Number of Listings	533	1,144	125	6	5273
Number of Ratings	29,743	105,124	3,257	7	1,517,104
Rating (% Positive)	96.0%	3.4%	97.0%	60.0%	100.0%
Price Changes Per Listing	3.7	3.5	2.6	0.1	28.0
Days Observed Per Listing	65.1	44.0	51.3	2.0	223.7

* N = 496

Table 3: Averages by Cluster

	(1) Low Rank, Freq. Change	(2) High Rank, Infreq. Change	(3) High Rank, Freq. Change
Target Rank	3.7	17.0	25.8
	(1.5)	(13.1)	(13.1)
Days Bet. Change	9.8	31.6	11.7
	(3.4)	(19.1)	(4.6)
Ranks Bumped	0.4	1.8	0.8
	(0.2)	(0.9)	(0.5)
Size of Change	6.6	11.5	7.0
	(3.7)	(8.4)	(4.7)
% Inv. Changed	4.2	2.2	9.2
	(3.3)	(1.5)	(6.3)
Number of Ratings	17,750	66,000	18,250
	(46,500)	(188,000)	(62,000)
Number of Listings	185	1,056	929
	(417)	(1,534)	(1,621)
Number of FBA	63	17	8
Number of Sellers	285	123	88

Table 4: OLS Model for Frequency of Price Change

Variable	Coefficient
Intercept	1.83***
	(0.11)
Synchronization	-2.61***
	(0.58)
Log Number of Listings	0.16***
	(0.02)
Observations	496
R^2	.23

Standard errors in parentheses.

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

Table 5: Top 20 Largest Sellers By Number of Ratings

Seller	**Site**	**% Difference†**	**Cluster**
goHastings	gohastings.com	5.6 (15.6)	2
internationalbooks	NA	NA	2
any_book	NA	NA	2
the_book_depository_	bookdepository.com	16.9 (9.2)	3
bookcloseouts_us	bookcloseouts.com	-8.3 (23.3) ‡	2
ExpressMedia	NA	NA	1
alibris	alibris.com	NA	2
best_bargain_books3	bestbargainbook.com	0.0 (0.0)	1
powells_books	powells.com	0.0 (0.0)	2
—SuperBookDeals	NA	NA	2
massbookstore	NA	NA	1
bargainbookstores-	bargainbookstores.com	NA	2
bellwetherbooks	NA	NA	2
–textbooksrus–	textbooksrus.com	4.4 (11.7)	1
MovieMars-Books	moviemars.com	14.9 (8.6)	3
SuperBookDeals-	NA	NA	2
pbshopus	NA	NA	2
Booknackrh	thebestlittlebookstore.com	NA	3
collegebooksdirect	collegebooksdirect.com	NA	2
your_online_bookstore	youronlinebookstore.com	NA	1

* Average over 100 ISBNs displayed

* Standard errors in parentheses

† $(P_{amzn} - P_{indep})/P_{amzn}$

‡ Pre-shipping cost

Table 6: OLS Models for Price Variability: Lowest Price

	(1)	(2)	(3)
	% Diff.	CV	CV†
Amazon's CV	14.41***	5.77***	2.33***
	(1.97)	(0.65)	(0.65)
No Amazon Change	-7.43**	-5.73***	-0.11
	(2.56)	(0.84)	(0.84)
Best Sellers Rank			
Log Avg.	-0.61	-0.25	-4.49***
	(1.24)	(0.38)	(0.68)
Log Std. Dev.	4.75***	2.32***	5.06***
	(0.99)	(0.32)	(0.63)
Avg. Star Rating	-1.03	-0.29	-0.01
	(1.16)	(0.38)	(0.38)
Log List Price	1.78	1.06	-7.13
	(4.26)	(1.40)	(4.17)
Number of Sellers			
Low Rank, Freq. Change	6.16***	2.19***	1.78***
	(0.64)	(0.21)	(0.21)
High Rank, Infreq. Change	1.20***	0.09	-0.43***
	(0.29)	(0.09)	(0.09)
Low Rank, Freq. Change	0.99*	0.66***	-0.08
	(0.50)	(0.17)	(0.17)
Not Clustered	0.12*	0.04*	0.01
	(0.05)	(0.02)	(0.02)
Observations	5273	5273	5255
R^2	.18	.23	.08

Standard errors in parentheses, coefficients on dummy variables for source omitted.

† Coefficient of variation for the average of the 3 lowest prices.

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