Working Paper CeDER-06-02, New York University

The Dimensions of Reputation in Electronic Markets

Anindya Ghose, Panagiotis G. Ipeirotis, Arun Sundararajan
Leonard Stern School of Business, New York University, {aghose,panos,asundara}@stern.nyu.edu

We analyze how different dimensions of a seller's reputation affect pricing power in electronic markets. We do so by using text mining techniques to identify and structure dimensions of importance from feedback posted on reputation systems, by aggregating and scoring these dimensions based on the sentiment they contain, and using them to estimate a series of econometric models associating reputation with price premiums. We find that different dimensions do indeed affect pricing power differentially, and that a negative reputation hurts more than a positive one helps on some dimensions but not on others. We provide the first evidence that sellers of identical products in electronic markets differentiate themselves based on a distinguishing dimension of strength, and that buyers vary in the relative importance they place on different fulfilment characteristics. We highlight the importance of textual reputation feedback further by demonstrating it substantially improves the performance of a classifier we have trained to predict future sales. This paper is the first study that integrates econometric, text mining and predictive modeling techniques toward a more complete analysis of the information captured by reputation systems, and it presents new evidence of the importance of their effective and judicious design.

Key words: Reputation, Reputation Systems, Text Mining, Opinion Mining, Online Feedback, Electronic Markets, Internet, Ecommerce, Electronic Commerce, Econometrics, Panel Data

1. Introduction

We show how pricing power in electronic markets is affected by different dimensions of a seller's reputation, which are identified by mining the text feedback of buyers. We demonstrate that different dimensions affect pricing power differently, contrast positive and negative reputation along each dimension, and provide evidence that buyers value dimensions differentially, often purchasing based on the fulfilment dimension a seller is differentiated on. Our results are based on a series of econometric models and are further validated by demonstrating substantive improvements in a predictive model based on our theory and data.

The motivation for our study is simple. When buyers purchase products in an electronic market, they assess and pay not only for the product they wish to purchase, but for a set of fulfillment characteristics as well: packaging, timeliness of delivery, the extent to which the product description matches the actual product, and reliability of settlement, for example. In traditional (bricks and mortar) retailing where buyers and seller are often co-located, buyers have cues that help them determine retailers' fulfillment characteristics more easily. These characteristics cannot be reliably described or verified in an electronic market prior to a transaction. If the intermediary running the market does not guarantee these characteristics, such markets rely on reputation systems to signal the quality of the trade processes one takes for granted in face-to-face and collocated transactions. The importance of such systems is widely recognized in the academic literature (surveys are available in Resnick et al. (2000), Dellarocas (2003), and Resnick et al. (2006)).

Typically, reputation in electronic markets are encoded by a "*reputation profile*" that provides potential buyers with:

1. The number of transactions the seller has successfully completed,

2. A summary of scores (or ratings) from buyers who have completed transactions with the seller in the past, and

3. A chronological list of textual feedback provided by these buyers.

Studies of reputation systems thus far have typically used average numerical scores reported by buyers as their measure of reputation. However, trade processes are multidimensional. Sellers vary in their capabilities at delivery, packaging, customer service and so on. Buyers may value each of these fulfilment dimensions differently, and there may be heterogeneity across buyers about the relative importance of different dimensions

Our study is based on the notion that the qualitative information contained in *text-based feedback* can be used to unravel these different dimensions of reputation. Prior work has conjectured that feedback text might contain information of economic value (e.g., Ba and Pavlou 2002, Cabral and Hortaçsu 2005). Further, casual observation of mediated electronic markets does suggests that different sellers in these markets derive their reputation from *different* characteristics. Text-based descriptions of transaction quality contain buyer assessments of these characteristics which might augment and increase the richness of the information contained in numerical reputation scores.

We begin our analysis by developing our ideas of multidimensional reputation qualitatively, based largely on prior theories and prior studies of reputation systems. This is augmented by a simple model of pricing choices made by sellers who have different (and varying) abilities to fulfill different characteristics of transactions. Buyers make inferences about a seller's true characteristics based on their interpretation of its feedback profile, which comprises numerical and text-based information about the observed seller characteristics for each of their prior transactions. This analysis leads to a set of hypothesis, which briefly, conjecture that sellers with a higher frequency of positive assessments on each of these characteristics can successfully charge higher prices, that these characteristics may be of varying importance, that buyers value different characteristics differently, and that the "differentiating dimension" of a seller contributes disproportionately to its pricing power.

Next, using a text mining technique we have developed and implemented, we structure the text feedback in the reputation profile of each seller who has participated in at least one of over 9,500 transactions for the sale of packaged consumer software on Amazon.com. Our text mining technique identifies the simple *dimensions* associated with a seller's recorded reputation (examples of simple dimensions include "delivery" and "packaging") and then locates the *modifiers* associated with each dimension (examples of such modifiers include "*fast* delivery" and "*careless* packaging"), thus converting unstructured text feedback for a seller into a vector of dimension-modifier pairs. We aggregate the top-500 such pairs into a set of eight topics (or *fulfilment dimensions*), each corresponding to a well-defined real-world fulfilment characteristic or trade process, and associated with a specific sentiment (positive, neutral or negative).

We test the our hypotheses by estimating a series of models. Our baseline estimation associates the average numerical score associated with a seller's reputation and the level of experience (that is, the number of transactions in the seller's profile) with the premium in price the seller can command over other sellers who simultaneously have an identical product available at the time the transaction takes place. A higher average reputation and a higher level of experience sometimes increase pricing power, but this is not always the case. We conjecture that the baseline estimates that associate an increase in reputation with a decrease in pricing power do so because they do not account for unobserved heterogeneity across fulfilment dimensions.

Next, we calibrate a scoring function that assigns dimension-specific pricing premium scores to each of the modifiers we have mined. These scores isolate the information contained explicitly in the text feedback of a seller's profile, and are aggregated by topic into a normalized score for each fulfilment dimension. We test our main hypotheses using these aggregated dimensions and scores. We demonstrate that certain dimensions have more influence than others. For the most part, increases in the scores associated with positive reputation along each dimension lead to higher pricing premium, and vice versa. Further we show that the dimension that *differentiates* the seller —the fulfilment characteristics on which the seller's performance relative to its competitors is highest— independently explains variation in pricing power, even after accounting for average differences in scores across the dimensions. This is interesting because it is consistent with a theory that buyers are heterogeneous with respect to the dimensions they value, since revealed preference suggests that the successful buyer weighs what the seller is good at more heavily.

We provide additional evidence of the importance of the information contained in textual feedback by reporting on fairly substantial improvements in the power of a predictive model we have trained to assess which seller among a competing groups is likely to successfully transact. With just numerical reputation information, its accuracy is 0.74. When given access to the information in textual feedback, this performance improves by about 20%.

The rest of this paper is organized as follows. Section 2 places our research in two related literatures: empirical studies of online reputation systems, and opinion mining research computational linguistics. Section 3 develops our hypothesis.[1] Section 4 describes our data set and presents some baseline results. Section 5 describes our text mining approach, its application on identifying and scoring different dimensions of reputation, and Section 6 presents evidence that supports our main

[1] Appendix A contains our analytical framework and the theory that leads to our hypothesis.

hypotheses. Section 7 discusses the managerial implications of our findings by describing tools that can be built using our findings, which can improve the design of existing reputation systems. Finally, Section 8 concludes and outlines directions for future research.

2. Related work

Our paper adds to a growing literature on reputation systems. Most prior work has estimated hedonic regressions of absolute price that view reputation as a product characteristic (though their details vary in important ways). Their results often attribute a positive value to a good reputation: buyers pay more to sellers who have better histories. For example, Kalyanam and McIntyre (2001) study Palm Pilots and PDAs, and Melnik and Alm (2002) study gold coins; each of these studies finds that positive feedback increases prices while negative feedback decreases prices. Lucking-Reiley et al. (2000) finds that a 1% increase in negative feedback leads to a 0.11% decrease in the final bid price for gold coins. Dewan and Hsu (2004) find that the eBay reputation system has a significant but economically modest effect on final auction prices and likelihood of sale. However, the results of prior studies are not directionally consistent. For example, Eaton (2002) finds that negative feedback reduces the probability of sale of electronic guitars only for sellers with more than 20 feedback postings; Livingston (2002) finds that experienced sellers with positive feedback earn premiums over new sellers with no feedback, but cannot establish any effect of negative feedback. Cabral and Hortaçsu (2005) identify a significant effect of reputation on pricing power only after eBay changed its display in 2003. McDonald and Slawson (2002) find that more negative feedback actually increases the number of bidders in secondary market auctions. Ghose et al. (2006) are unable to find evidence relating seller reputation to used book prices at Amazon. Resnick et al. (2006) organized a series of controlled field experiments selling postcards on eBay to identify the effect of experience and reputation rating on sales. Their findings suggest that buyers are willing to pay approximately 8% more for lots sold by the more experienced seller identity rather than the new venders. They also find that negative feedback has little impact on sales.

We believe that a fraction of this conflicting evidence can be explained by a using a more robust measure of the value of reputation – a price premium, rather than simply absolute price, and by recognizing that reputation is multidimensional.[2] We explicitly incorporate both these features. In contrast with the prior literature, we use publicly available differences in *posted* prices for homogeneous products of constant quality as our measure of value, which mitigates some of the variation caused by product heterogeneity and by bidding psychology on the closing auction prices that form the basis for other studies.[3]

We also add to an emerging literature which combines economic methods with text mining. This includes Das and Chen (2006) who extract investor sentiment from bulletin boards on Yahoo! Finance and show it explains stock index movement well, Gu et al. (2007) who analyze the trade-offs between network size and information quality relating to how users value online communities, and Lewitt and Syverson (2005) who study how textual descriptions of housing characteristics affect final sale prices and time-to-sale by agents.

To the best of our knowledge, however, ours is the first paper that uses text mining techniques to analyze reputation feedback. This seems like a very natural context for these techniques, since there are few substitutes for the information contained in textual reputation feedback in an electronic market (this is in contrast with investor sentiment, housing or product reviews). Two earlier papers that have a similar motivation are Pavlou and Gefen (2005) and Pavlou and Dimoka (2006). The former relates aspects of text feedback to psychological contract violation and the second relates text feedback to buyer trust and pricing on eBay. However, they rely on manual (and expensive) content analysis techniques. Further, the latter paper pre-specifies two trust dimensions (credibility and benevolence). As a consequence, only a small proportion of text comments are categorized as providing evidence of a seller's outstanding credibility and benevolence. In contrast,

[2] This is consistent with prior marketing theory suggesting that retailers can be heterogenous on different attributes (Dawar and Parker 1994).

[3] Although it is implicit in our empirical approach, we do not explicitly address the fact that reputation also builds trust between traders (see, for instance, Resnick et al. (2000), Ba and Pavlou (2002), and Ba (2001)). This could be done by analyzing the network structure of buyer-seller relationships, an interesting direction for further work.

our techniques *automatically* identify the dimensions of reputation from unstructured text feedback, and *automatically* understand the positive or negative semantic orientation of each evaluation, together with the intensity of such evaluation. We are thus able to compare multiple dimensions of reputation, provide evidence consistent with buyers valuing different dimensions differentially, and predict future sales based on an analysis of seller reputation profiles, all of which move our paper's contribution substantially beyond existing work.

The techniques we use in this paper draw from research on opinion extraction, which has attracted substantial recent interest in the computational linguistics community. For instance, Hatzivassiloglou and McKeown (1997) use a supervised learning technique to identify the *semantic orientation* of adjectives. Despite the high accuracy of the proposed technique, the requirement for manual tagging of training data makes this technique prohibitively expensive in our setting. Turney (2002) notes that the semantic orientation of an adjective depends on the noun that it modifies (e.g. "unpredictable steering" for a car vs. "unpredictable plot" for a movie), and suggests using adjective-noun or adverb-verb pairs to extract semantic orientation, an approach we follow. Turney and Littman (2003) determine the semantic orientation of a word pair by computing the pairwise mutual information between the word pair and a set of unambiguously positive words (e.g., good, nice, excellent, positive, fortunate, correct, superior) and unambiguously negative words (e.g., bad, nasty, poor, negative, unfortunate, wrong, inferior). To compute the pairwise mutual information they issue queries to search engines for each evaluated word pair. Kamps and Marx (2002) use WordNet (Fellbaum 1998) to measure the distance of each word from "good" and "bad.". Hu and Liu (2004), whose study is the closest to our work, use Wordnet to compute the semantic orientation of product evaluations and try to summarize user reviews by extracting the positive and negative evaluations of the different product features. Lee (2004) uses text mining in the context of an ontology-based approach to analyzing product reviews. Our work is also related to recent papers about word-of-mouth that relate content in product reviews to demand (e.g., Godes and Mayzlin 2004, Senecal and Nantel 2004, Chevalier and Mayzlin 2006) by looking at attributes such as the length of the review or newsgroup. Our work uses a significantly more sophisticated text analysis methodology, allowing us to identify the nuances in the text that previous papers could not capture.

While our text mining of reputation profiles is inspired by these previous studies about opinion extraction and subjectivity analysis, it does differ in significant ways. First, we do not require any external resources for evaluating the (positive or negative) orientation of a word. Using search engines (e.g., Turney 2002, Turney and Littman 2003) is prohibitively expensive for large scale evaluation. The use of a lexical resource such as Wordnet is problematic when the adjectives can have different meanings that rely on their context (e.g., "fast shipping" vs. "fast packaging"). Second, by using price premiums to score our word pairs, we have the first truly objective measure of the positive or negative effect of the text in an evaluation. For example, on Amazon, the buyers tend to use superlatives to give a positive evaluation (e.g., "great packaging"), and therefore, simple evaluations (e.g., "good packaging") may actually have neutral or slightly negative connotation. This is in contrast with all previous opinion extraction techniques that would have unambiguously characterized "good packaging" as a positive evaluation.

3. Reputation and pricing power

This section develops our hypotheses, drawing on prior theory, past studies of reputation systems, and an outline of an economics model (which is presented in Appendix A).

Buyers and sellers are separated by time and distance. Furthermore, the quality of a seller on fulfilment characteristics a buyer may value is not known prior to transacting. Our underlying model is of risk-averse buyers who choose a seller to maximize their expected surplus from trade. All else being equal, these buyers are likely to choose sellers in a manner that lowers transaction risk. The premium a seller can charge will therefore be related to the degree of risk associated with a transaction. Sellers with lower numeric ratings and fewer completed transactions have less "information" about them available to prospective buyers. Better average scores and a longer history also assure buyers that sellers will not "sacrifice" their reputation and renege (Klein and Leffler 1981, Klein 2000) and in this context, reputation systems can act as assurance mechanisms (Kalyanam and McIntyre 2001). Risk averse buyers will therefore be inclined to buy at a price premium from a seller with higher average ratings and more prior transactions.

Motivated by the risk induced by fulfilment quality uncertainty and separation in time /distance, past research has conjectured that institutional feedback mechanisms facilitate trust on the Internet (Ba and Pavlou 2002, Pavlou and Gefen 2005), and how ecommerce sites design trust-building features by posting text (Kim and Benbasat 2003). Trust can be formed through familiarity (Gefen et al. 2003). Transaction frequency can make a seller more familiar to a buyer, or to others known to the buyer. Thus, prior work on reputation systems as methods for risk mitigation and trust building lead to the same baseline hypotheses:

H1 *Sellers with higher average numerical reputation scores will have higher price premiums associated with their successful transactions.*

H2 *Sellers with a more experience will have higher price premiums associated with their successful transactions.*

Next, we turn to the multi-dimensionality of reputation. It seems quite natural to us that order fulfilment has many dimensions. A couple of prior studies have explored this idea slightly. For example, Pavlou and Dimoka (2006) postulate credibility and benevolence as two dimensions of importance. If there are multiple fulfilment dimensions, it seems natural to conjecture that sellers may have different capabilities across them. It also seems natural that these dimensions could be valued, on average, differently, and will thus have varying effects on pricing power. Our basis for defining and scoring these fulfilment dimensions is textual feedback structured and scored as word pairs (bigrams) and categorized as positive or negative (more on this later). The next hypotheses follow naturally.

H3a *All else equal, sellers with a higher frequency of positive word pairs associated with a fulfilment characteristic will have higher price premiums associated with their successful transactions.*

H3b *All else equal, sellers with a higher frequency of negative word pairs associated with a fulfilment characteristic will have lower price premiums associated with their successful transactions.*

Based on what we know about decision making under risk from Kahneman and Tversky (1979), and given that our buyers are risk averse, it seems possible that negative words will have a relatively larger impact. We do not hypothesize this explicitly, but discuss it later.

Our final hypothesis is based on our model of buyers being heterogeneous in the relative importance they place on different fulfilment dimensions. If this is indeed the case, then the successful completion of a transaction reveals a buyer's preference for the dimensions that a seller is accomplished at. For example, suppose a seller scores high on delivery and low on customer service. If all buyers placed the same weights on delivery and service, there would be no real revelation of preferences in the fulfilled transactions. On the other hand, if some buyers weight delivery more highly than customer service, and others weight customer service more than delivery, it is very likely that a transaction fulfilled by the seller in question involves a buyer from the former set. The fact that the seller's *distinguishing dimension* is delivery will thus have additional explanatory power about its price premium (that is, over and above what is explained by the premiums attached to delivery on average. on a specific dimension.)

H4 *All else equal, a higher score on a seller's distinguishing characteristic leads to a higher price premium.*

4. Data and baseline results

Our data is of a cross-section of software resellers in several different categories gathered from publicly available information on software product listings at Amazon.com. Data from Amazon has many advantages for a study of our kind. Online feedback mechanisms rely on voluntary reporting of privately observed outcomes. This introduces potential reporting bias since traders may exhibit differing propensities to report different outcome types (Dellarocas and Wood 2006). On reputation systems like the one at Half.com, for instance, this leads to potential feedback manipulation by buyers and sellers, since buyers rate sellers, and sellers rate buyers, thus the truthfulness of reported reputation is affected by the fear of retaliation. Dellarocas and Wood (2006) provide evidence of both positive and negative reciprocation among eBay traders. On Amazon.com, on the other hand, sellers do not rate buyers. Consequently, reputations scores and feedback are more reliable.

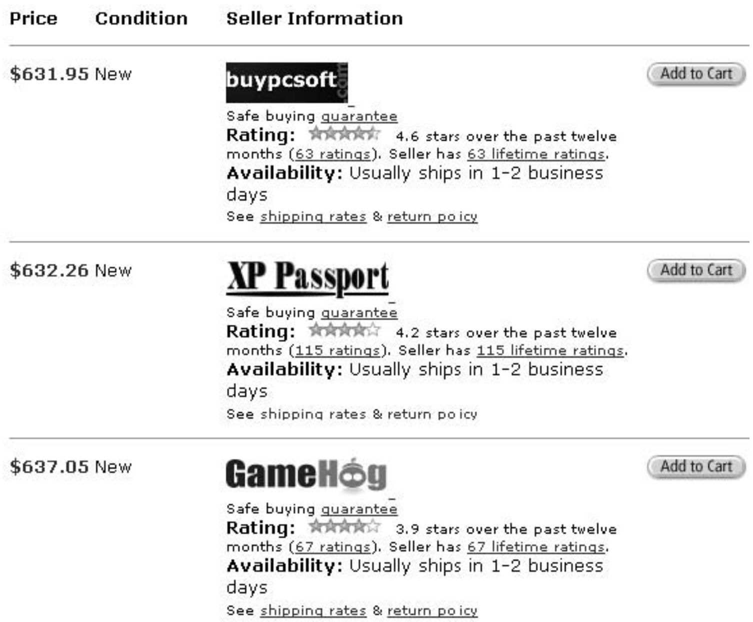
The data are gathered using automated Java scripts which access XML pages downloaded from the retailer. The panel includes 280 individual software titles, with an equal number of products from each of five major categories: Business/Productivity, Graphics, Development, Security/Utilities and Operating Systems. We gather data about all transactions for these software titles over 180 days between October 2004 and March 2005. Our set of sellers includes both individuals and Pro Merchants (who use Amazon's retail platform commercially). For each transaction, the variables associated with the seller and the product include the price at which the product was sold,[4] the ID of the seller, the seller's reputation at the time of the transaction (more on this later), the seller's reported condition of the product, and the duration for which the product was listed before it was sold. These conditions are coded in our dataset on a scale from 1 to 5, with 5 denoting the highest quality (New) and 1 denoting the lowest grade (Acceptable). Additionally, we also have variables associated with the *competing identical products* available at the time the product was sold.

We mine the XML feed every eight hours. We exploit the fact that Amazon.com associates a unique transaction ID with each listing. This transaction ID enables us to distinguish between multiple or successive listings of identical products sold by the same seller. When a transaction ID associated with a particular listing is removed, we infer this to mean that the listed product was successfully sold in the prior eight-hour window (Ghose et al. 2006).[5] The unique transaction

[4] While it is true that the posted prices are chosen by sellers, buyers have a choice of which seller to buy from. So, a rational buyer would buy at a price that maximizes the buyer's utility, and this is likely to be closer to their actual valuation for the product.

[5] Amazon indicates that their Pro Merchant seller listings remain on the site indefinitely until they are sold, while products listed by individual sellers are removed from the site after 60 days. We saw no unusual rise in inferred sales around the 60 day mark. Therefore, we include all inferred sales in our analysis regardless of the number of days before a sale occurs. We also ran our estimates by removing all imputed sales that occur exactly 60 days after listing, and this resulted in no appreciable change to our results. However, if the seller delists the product and then adds it again as a separate listing then,that would lead to some noise in the data. But given that seller has to pay a listing fee of $0.99 every time they list a product, it seems logical that they will simply change the prices without de-listing and re-listing the product.

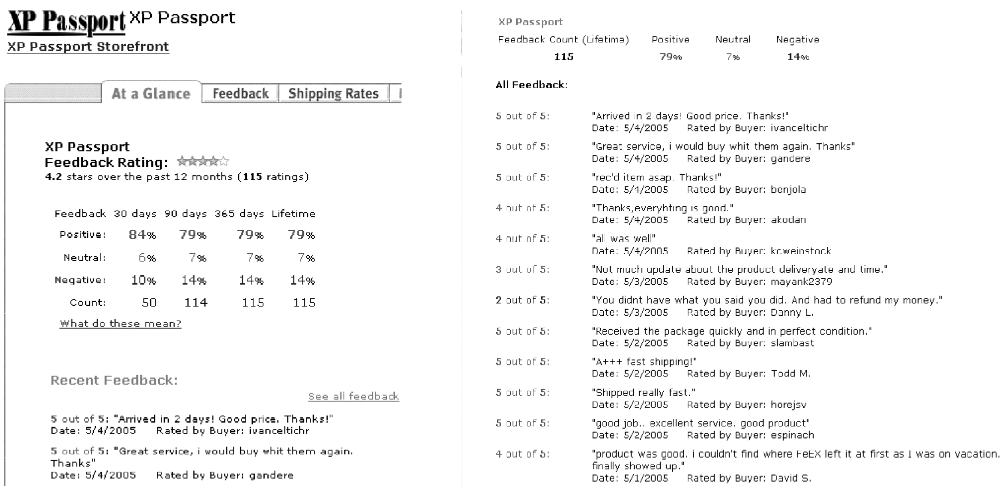
Figure 1 A set of sellers on the secondary market of Amazon.com selling an identical product for different prices



ID associated with each new listing lets us infer if a transaction has occurred even if the seller immediately lists another identical product for sale (while this listing would have the same seller and product, it would have a new transaction ID). We have data about 9484 unique transactions, along with reputation and pricing data for the successful seller and each of its competitors (defined as sellers who had listed the same product on Amazon.com at the time of any transaction). The "reputation" of a seller is constructed from its entire reputation history. Each of these sellers has a feedback profile, which, as described earlier, consists of numerical scores and text-based buyer feedback. The numerical ratings are provided on a scale of one to five stars. These ratings are averaged to provide an overall score to the seller. Amazon also reports similar averages over the last 30 days, 90 days and 365 days, for each of three categories: positive (4-5), neutral (3) and negative (1-2). Some typical positive, neutral, and negative comments are displayed in Figure 2.

Again, we collect all feedback (both numerical and text-based) associated with a seller over the entire lifetime of the seller, rather than simply over the 180-day period. This enables us to reconstruct each seller's exact feedback profile at the time each transaction. There were 1078 sellers, with an average level of experience of 4,932 postings. Of these 1078 sellers, 122 of them successfully

Figure 2 Fraction of the feedback profile for a seller, as displayed by Amazon



completed transactions in our 180-day period, while the remaining acted only as participants in the market, without actually selling a (monitored) product.[6]

We have chosen software as our product category because of the uniformity in its product quality across sellers (92% of our products were listed by the seller as “new”, and in general, software does not “degrade” with use). This is important because it implies that price variation we observe can be attributed primarily to variation in the seller’s performance on the fulfilment characteristics that buyers value, and the expected quality that buyers infer about the sellers’ potential performance based on the information contained in their feedback profiles. Second, many of the titles are pretty valuable. Thus, problems with fulfilment are likely to have an economic impact on the buyer, so it is reasonable to expect buyers to use seller feedback to infer reputation and pay a higher price for it.

4.1. Baseline results: average reputation and experience

Our preliminary estimates focus on numeric feedback scores and ignore all text-based feedback completely. Our dependent variable is *PricePremium*. We use two variants. The first is *RegPricePremium*, the difference between the price at which the transaction occurred and the price of each competing unsuccessful seller. This creates $N-1$ observations per transaction where N is the

[6] These sellers may have conducted transactions on products outside our 280-product panel.

total number of sellers. The second is *AveragePricePremium*, defined as the difference between the price at which the transaction occurred, and the average prices of competing unsuccessful sellers. This leads to one observation per transaction. If a seller with a higher average reputation or a higher level of experience can charge a higher price we should observe higher price premiums for this seller.

Our variables are summarized in Table 1, and their descriptive statistics in Table 2. The main variables are:

- ***ProductPrice***: the manufacturer's price at which software was listed on Amazon, used to control for the differences in the absolute values of the software products.
- ***Rating***: the average value of the seller's (*SellerRating*) and competitors' (*CompetitorRating*) *numerical* scores over their entire transaction history. *DiffRating* encodes the difference in ratings and is equal to *SellerRating-CompetitorRating*.
- ***Product Condition***: the average condition of the product as reported by the seller (*SellerCondition*) and competitor *CompetitorCondition*). As above, we also have *DiffCondition* that is equal to *SellerCondition-CompetitorCondition*.
- ***Life***: the total number of seller (*SellerLife*) and competitor transactions (*CompetitorLife*), measuring their level of experience on Amazon.com; the *DiffLife* is equal to *SellerLife-CompetitorLife*.[7]
- ***Competitors***: the number of competitors for each transaction. Intuitively, more competitors leads to more intense competition and lower price premiums.

We estimated models of the following form:

$$\ln(\mathit{RegPricePremium}) = \alpha + \beta_1 \ln(\mathit{ProductPrice}) + \beta_2(\mathit{RegDiffRating}) + \quad (1)$$

[7] Note that depending on the regression, each of *DiffRating, DiffLife and DiffCondition* are coded in two ways. For example, first, *DiffLife* is coded as the difference in the number of transactions between each seller and each competitor. This variable is used as an independent variable when the dependent variable is *Regular Price Premium*. In this case, for each transaction, there are N observations, where N is the number of competing sellers at time of sale. Second, *DiffLife* is coded as the difference between the number of lifetime feedback evaluations received by the seller and the AVERAGE of the lifetime feedback evaluations received by the competitors. This variable is used as an independent variable when the dependent variable is *Average Price Premium*. In this case, for each transaction, there is one observation.

Variable	Brief Description
RegPricePremium	Difference between the sale price and all of the competing prices.
AvgPricePremium	Difference between the sale price and the average of all competing prices.
ProductPrice	List price of the new product as listed on Amazon.
SalePrice	Price at which the used product was sold on Amazon.
RegDiffRating	Difference between the numerical reputation score of the seller and the reputation scores of all competitors at the time of sale.
RegDiffLife	Difference between the number of transactions completed by the seller and the number of transactions completed by all the competitors at the time of sale.
RegDiffCondition	Difference between the product condition reported by the seller and that reported by all the other competitors at the time of sale.
AvgDiffRating	Difference between the numerical reputation score of the seller and the average of the reputation scores of all competitors at the time of sale.
AvgDiffLife	Difference between the number of transactions completed by the seller and the average of the number of transactions completed by all the competitors at the time of sale.
AvgDiffCondition	Difference between the product condition reported by the seller and the average of the product condition of all the other competitors at the time of sale.
Competitors	Total number of unique sellers for a given product listing.

Table 1 Descriptions of the numeric variables that we use in our econometric models (see Equations 1 and 2).

$$\beta_3 \ln(RegDiffLife) + \beta_4(RegDiffCondition) + \beta_5 \ln(Competitors) + \mu + \epsilon.$$

$$\ln(AvgPricePremium) = \alpha + \beta_1 \ln(ProductPrice) + \beta_2(AvgDiffRating) + \qquad (2)$$

$$\beta_3 \ln(AvgDiffLife) + \beta_4(AvgDiffCondition) + \beta_5 \ln(Competitors) + \mu + \epsilon.$$

We use OLS regression with fixed effects controlling for unobserved heterogeneity across sellers and products, as well as regressions with product, seller, and competitor fixed effects.[8] Note that our data is at the transaction level–that is the unit of observation is at the level of a product i sold by a seller j when competing with other sellers, N, where N is the number of competing sellers at the time of the transaction. Here μ denotes the fixed effect and ϵ denotes the idiosyncratic error term. Both regressions described above yielded qualitatively similar results, and we report the results of both the sets, since they are both relevant for the text analysis that follows.[9]

[8] We verified that the fixed-effects transformation was in fact more suitable than the random-effects transformation using the Hausman test.

[9] To normalize the distribution and minimize the effect of any outliers, we take the log of the *ProductPrice*, *PricePremium*, *DiffLife* and *Competitors* variables.

The results of these estimations are presented in Tables 11 and 12. In both cases, these results support our hypothesis 2: the differences in the level of reputation score between sellers and competitors (*DiffLife*) has a positive and significant effect on pricing premiums, in Tables 11 and 12. Notice that the average price premium changes much more rapidly with changes in average reputation and experience than the price premium relative to one's nearest competitor. The coefficient of ln(*ProductPrice*) is less than 1 in each case, indicating that while the magnitude of the price premium increases with product price, it decreases in percentage terms. The other coefficients can be interpreted in a standard way: for instance, the coefficient of *DiffCondition* in Table 12 indicates that a 1 point increase in the difference between the seller's and competitor's average product condition increases the seller's price premium by about 11%.

Surprisingly, while the coefficients on *DiffRating* are consistently significant, their signs are mixed. Each of our regressions controls for seller and product fixed effects. When controlling for unobserved heterogeneity across competitors with whom the pairwise comparison is made as well (which is only possible when RegPricePremium is the dependent variable), a higher reputation is associated with a higher price premium (table 11, column 2). However, in the absence of competitor fixed effects, a higher reputation score is associated with a *lower* price premium (table 11 column 1 and table 12). One simple interpretation of this unusual finding is that differences across competitors represented by information contained in their text feedback is important. When there is a control for unobserved heterogeneity across competitors, this is partially accounted for (note that since we have not yet used text in the estimates, this feedback is still "unobserved"). If not, there may be information about fulfilment dimensions contained in a seller's feedback that is systematically causing a competitor to beat the seller despite a lower average rating.

The above argument is not the only one consistent with our findings (for example, there may be a few powerful sellers with high reputation scores who are just not good at software transactions). In any case, it adds to the body of mixed evidence relating average numerical reputation scores to prices, and further motivates a deeper analysis of the different dimensions constituting a seller's reputation. We turn to this task in the next section.[10]

5. Discovering and Analyzing the Dimensions of Reputation

In Section 5.1, we describe how we use linguistic analysis techniques to structure the textual feedback and to discover the dimensions of reputation (e.g., "*packaging*") mentioned in the text. In Section 5.2, we show how we can estimate the effect of the evaluations (e.g., "*cool* packaging") that buyers post, and see how such evaluations affect the pricing power of the sellers.

5.1. Retrieving the Dimensions of Reputation

This section describes a novel text analysis technique we have developed to structure the textual part of the feedback profiles. The goal of the technique is to discover the dimensions that contribute to the reputation of each vendor, identify the weight of that contribution, and quantitatively value the text-based feedback provided by buyers on each dimension (e.g., that "cool packaging" is way cooler than "good packaging").

Consider a seller was characterized by a vector of characteristics $X = (X_1, X_2, ..., X_n)$, representing their ability on each of n fulfilment characteristics (see Appendix A for further details). Our technique is based on the notion that each of these n characteristics (or *dimensions*) can be expressed by a noun or a verb phrase chosen from the set of all text feedback, and that a seller is evaluated on these n dimensions. For example, dimension 1 might be "shipping", dimension 2 might be "packaging" and so on. Further, in our model, each of these dimensions is assigned a numerical score. Of course, when posting textual feedback, buyers do not assign explicit numeric scores to each (or to any) dimension. Rather, they use *modifiers* (which are typically adjectives or adverbs) to evaluate the seller along each of these dimensions (we describe how we assign numeric scores to each modifier later in this section). Once we have identified the set of all dimensions that sellers are identified along, we can then parse each of the actual feedback postings from our data set, associate a modifier with each dimension, and represent this feedback set as an n-dimensional vector of

[10] Note, that the low R-squared values in these regressions as well as in the subsequent regressions in Section 6 are expected because this is the "within" (differenced) fixed effect estimator. If we had estimated the fixed effects instead of differencing them out, the measured R-squared would be much higher. However, this latter model is often computationally intractable in our data(due to the large number of fixed effects to estimate).

modifiers. To illustrate this, consider the following example: vector of modifiers. To illustrate this, consider the following example:

Example 1. Suppose dimension 1 is "*delivery*," dimension 2 is "*packaging*," and dimension 3 is "*service*." The feedback posting *"I was impressed by the speedy delivery! Great service!"* is then encoded as $\phi_1 = [speedy, NULL, great]$, while the posting *"The item arrived in awful packaging, and the delivery was slow"* is encoded as $\phi_2 = [slow, awful, NULL]$.

Let $\mathcal{M} = \{NULL, \mu_1, ..., \mu_M\}$ be the set of modifiers and consider a seller s_i with p postings in its reputation profile. We denote with $\mu^i_{jk} \in \mathcal{M}$ the modifier that appears in the j-th posting and is used to assess the k-th reputation dimension. We then structure the merchant's feedback as an $n \times p$ matrix $\mathbf{M}(s_i)$ whose rows are the p encoded vectors of modifiers associated with the seller. Briefly, our algorithm constructs $\mathbf{M}(s_i)$ as follows:

1. Retrieves the feedback postings associated with a seller.

2. Parses the postings to identify the dimensions across which the buyer evaluates a seller. For this task, we use a *part-of-speech (POS) tagger*[11], which parses each posting and identifies the part-of-speech[12] for each word. We keep the nouns, noun phrases, and verbal phrases as dimensions of the seller. We eliminate from consideration all dimensions that appear in the profiles of less than 50 (out of the 1,078) merchants, since we cannot extract statistically meaningful results for such sparse dimensions[13].

3. Retrieves adjectives and adverbs that refer to the nouns and verbs extracted in Step 2. To associate the adjectives and adverbs with the correct dimensions, we use a *syntactic parser*[14]. A syntactic parser analyzes the linguistic structure of each sentence and identifies the relations[15]

[11] We used the Stanford JavaNLP tagger.

[12] The parts-of-speech are: verb, noun, pronoun, adjective, adverb, preposition, conjunction, and interjection.

[13] The technique as described so far, considers words like "shipping" and " delivery" as separate dimensions, although they refer to the same "real-life" dimension of a seller. We discuss how we overcome this limitation in Section 5.3.

[14] We used the "Collins HeadFinder" capability of the Stanford JavaNLP package.

[15] The use of a syntactic parser allows us to identify adjective-noun and adverb-verb pairs, even if the two constituent words are not placed next to each other.

between the words. For our purposes, we keep the adjective-noun and adverb-verb pairs, which will serve as the basis for our further analysis.

We have implemented this algorithm on the feedback postings of each of our sellers. This transforms the set of all unstructured text feedback into a set of structured evaluations. Our analysis yields 151 unique dimensions, and a total of 142 modifiers (note that there is overlap between the modifier sets for each dimension). Tables 8 and 9 show a subset of the summary statistics for the most frequent modifier-dimension pairs.

The real question, of course, is how to "understand" the meaning of these dimension-modifier pairs and how to measure the effect of these comments on the pricing power of the merchants. We discuss these issues next.

5.2. Scoring the dimensions of reputation

In order to assign a "value" to the text reputation profile, we have developed and implemented a method for inferring the numerical scores that should be associated with each modifier, for each simple dimension (the elements of the matrix $\mathbf{M}(i)$ as discussed earlier). These elements belong to the (global) set of modifiers $\mathcal{M}$. We aim to compute the "score" $a(\mu, j, k)$ that a modifier $\mu \in \mathcal{M}$ assigns to dimension k, when it appears in the j-th posting.

Since buyers tend to read only the first few pages of text-based feedback, rather than all these pages, recent text postings should influence a buyer's assessment more heavily. We model this by assuming that K is the number of postings that appear on each page ($K = 25$ on Amazon.com), and that c is the probability of clicking on the "Next" link and moving to the next page of evaluations.[16] Intuitively, we weight down the contribution of old postings in the overall reputation score, by assuming that the user looks only at the first few pages of feedback postings and clicks "Next" i times with probability c^i. This assigns a posting-specific weight r_j for the j-th posting:

$$r_j = \frac{c^{\lfloor \frac{j}{K} \rfloor}}{\sum_{q=1}^{p} c^{\lfloor \frac{q}{K} \rfloor}}, \tag{3}$$

[16] We conducted experiments with the values $c = 0.0$, $c = 0.25$, $c = 0.5$, $c = 0.75$ and $c = 1.0$. The results across all values of c were similar. For conciseness, we report in the rest of the paper only results for $c = 0.5$.

where j is the rank of the posting, K is the number of postings per page, and p is the total number of postings for the given seller. Next, we set:

$$a(\mu, j, k) = r_j \cdot a(\mu, k), \tag{4}$$

where by $a(\mu, k)$ we denote the "global" score that the modifier μ assigns to dimension k.

Finally, we model buyers as placing different weights on each dimension of interest. Thus, each buyer is characterized by a type vectors w, the weights that the buyer uses to compute a weighted average of these modifier scores. The overall "reputation score" given by a buyer of type w to a seller i whose feedback set has been converted to the matrix of modifiers $\mathbf{M}(i) = \{\mu^i_{jk}\}$ is therefore:

$$\Pi(i) = r^T.A(\mathbf{M}(i)).w, \tag{5}$$

where $r = [r_1, r_2, ...r_p]$ is the vector of the posting-specific weights. (See Equation 3.) Or more explicitly:

$$\Pi(i) = [r_1, r_2, ...r_p] \begin{bmatrix} a(\mu^i_{11}, 1) & \dots & a(\mu^i_{1n}, n) \\ \vdots & \ddots & \vdots \\ a(\mu^i_{p1}, 1) & \dots & a(\mu^i_{pn}, n) \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}. \tag{6}$$

If we model the buyer's type distribution $F(w)$ as being independently distributed along each dimension, each modifier score $a(\mu, k)$ is also an independent random variable, and the random variable $\Pi(i)$ is a sum of random variables. Specifically, we have:

$$\Pi(i) = (w_1 \cdot a(\mu_1, 1)) \cdot R(\mu_1, 1) + \ldots + (w_n \cdot a(\mu_M, n)) \cdot R(\mu_M, n) \tag{7}$$

where $R(\mu_j, k)$ is equal to the sum of the r_i weights in all the postings in which the modifier μ_j modifies dimension k. From our data, we can easily compute the $R(\mu_j, k)$ values by simply counting the times that each dimension-modifier pair appears and summing appropriately the r_j values. Interestingly, we do not have to estimate the distributions of $w_k \cdot a(\mu_j, k)$; instead, we can just treat $w_k \cdot a(\mu_j, k)$ as part of the β_i coefficient that a regression assigns to each regressor. In our case, the regressors are the modifier-dimension pairs.

We first treat each modifier-dimensionpair as a separate regressor . (We present an alternative approach in Section 5.3.) Estimating equations similar to the ones described in Equation 1, after adding the $\Pi(\cdot)$ variables for the seller and the competitor in the regression. This process yields a weight associated with each modifier-dimension pair, which can be interpreted as an ordinal measure of the "value" (or increase/decrease in pricing power) of having that pair associated with a particular transaction. We discuss our findings in Section 6.2.

5.3. Clustering the Dimensions of Reputation

We observe that many simple dimensions refer to the same real-life fulfilment dimension. For example, the pairs "never sent," "never received", "never delivered", "not received", "not shipped", and "not delivered" have the same meaning and represent only one dimension (delivery). Ideally, we would like these (separate) dimensions to appear as a single independent variable.

We first attempted to apply one of many clustering techniques used to group together different verbs and nouns that correspond to the same dimension. We have experimented with Latent Dirichlet Allocation (LDA) (Blei et al. 2003), Non-Negative Matrix Factorization (NMF) (Lee and Seung 1999), and probabilistic Latent Semantic Analysis (pLSI) (Hofmann 1999). These represent the state-of-the-art machine learning techniques for clustering. Unfortunately, such techniques did not work well in setting: the underlying assumption for all these techniques is that nouns and verbs that refer to the same real-life dimension will appear frequently together. Buyer feedback exhibits exactly the opposite pattern, though: once buyers discuss one aspect of the transaction (e.g., shipping) they typically start discussing another dimension (e.g., packaging), not surprising given the limited amount of space available for the posting.

Given the difficulty of using automatic clustering techniques, we used a semi-automatic approach. We first derived the most important dimension-modifier pairs using our text mining techniques, ranked by frequency. We then manually assigned the top-500 most frequent pairs to categories. We first identified the eight different fulfilment dimensions by examining the list of modifier-dimension pairs. (This outcome is further supported by the results of an independent annotation study, described in Section 6.1.) Then, we classified each modifier-dimension pair as positive, neutral or negative. This leads to a fulfillment characteristics vector for each seller, which has $3n$ components, where n is the total number of dimensions.[17] We compute the scores for each (clustered) dimension in the same way that we did in Section 5.2 for the individual modifier-dimension pairs. We discuss our findings in Section 6.2.

6. Experimental Results

In this section, we discuss our findings, and describe the results of some robustness checks. We first report the results of our content analysis study that verify the high recall of our text mining technique (Section 6.1). Then, we present results that measure the importance of each dimension of reputation (Section 6.2). Finally, we describe the predictive power of text evaluations, and their use in a predictive model to determine future market outcomes. (Section 6.3). We conclude with a discussion of how the different dimensions of reputation affect the pricing power of online merchants (Section 6.4).

6.1. Recall of Extraction

Our first experimental evaluation step examines whether the opinion extraction technique of Section 5.1 indeed captures all the reputation characteristics expressed in the feedback (recall) and whether the dimensions that we capture are accurate (precision). For recall, we used two human annotators. The annotators read a random sample of 1,000 feedback postings, and identified the reputation dimensions mentioned in the text. Then, they examined the extracted modifier-dimension pairs for each posting and marked whether the modifier-dimension pairs captured the identified real reputation dimensions mentioned in the posting and which pairs were spurious, non-opinion phrases.

Both annotators identified nine reputation dimensions (see Table 16). Since the annotators did not agree in all annotations, we computed the average *human recall* $hRec_d = \frac{agreed_d}{all_d}$ for each dimension d, where $agreed_d$ is the number of postings for which both annotators identified the reputation

[17] $n = 8$, although we end up with 21 and not $3 \cdot 8 = 24$ dimensions, because our data does not contain neutral pairs for three dimensions.

dimension d, and all_d is the number of postings in which at least one annotator identified the dimension d. Based on the annotations, we computed the recall of our algorithm against each annotator. We report the average recall for each dimension, together with the human recall in Table 16. The recall of our technique is only slightly inferior to the performance of humans, indicating that the technique of Section 5.1 extracts the majority of the posted evaluations.[18]

Interestingly, precision is not an issue in our setting. In our framework, if an particular modifier-dimension pair is just noise, then it is unlikely to have a statistically significant correlation with price premium. Put differently, the noisy opinion phrases are statistically guaranteed to be filtered out by the regression.

6.2. The Effect of Text Evaluations

In this section, we discuss our findings about the importance and the effect of text evaluations on the price premiums that merchants can charge. We also show that typically there is one *differentiating* dimension of reputation that explains a disproportionate variation in price premium. This is consistent with our model of heterogeneous buyers, and evidence that the merchants can treat the most important dimension of their reputation as a differentiating characteristic.

Effect of individual phrases: Tables 13, 14 and 15 summarize the dimension-modifier pairs (positive and negative) that were statistically significant across all regressions. A weight of zero means that the modifier-dimension pair has no effect on the seller's pricing power, a weight above zero means that it has a positive impact, and a weight lower than zero means that it has a negative impact on the seller's (positive) price premium. Recall that these reflect changes in the seller's average pricing power across products *after* taking their average numerical score and level of experience into account, and highlight the importance of the value contained in text-based reputation.[19] Note also that the coefficients reveal characteristics of the reputation market that

[18] In the case of "Item Description," where the computer recall was higher than the human recall, our technique identified almost all the phrases of one annotator, but the other annotator had a more liberal interpretation of "Item Description" dimension and annotated significantly more postings with the dimension "Item Description" than the other annotator, thus decreasing the human recall.

[19] We checked the correlation matrix of the independent variables. Due to the huge number of independent variables, we are unable to report the correlation matrix. However, our analysis revealed that there is very little correlation between the modifier-dimension variables, with less than 1% of the pairs having a correlation higher than 0.2.

cannot be normally captured by existing opinion mining systems. For example, on Amazon, the buyers tend to use superlatives to give a positive evaluation (e.g., "great packaging"), and therefore, simple evaluations (e.g., "good packaging") actually have slightly negative effect. This is in contrast with all previous opinion extraction techniques that would have unambiguously characterized "good packaging" as a positive evaluation.

Effect of clustered dimensions: We create the *DiffDimension* variable where *DiffDimension* is the difference between the seller and the competitor's score on a given (clustered) dimension. Therefore, each *DiffDimension* (positive, neutral or negative) is a separate independent variable in our model. We estimate equations similar to those in Equation 1, adding these clustered dimension variables. Furthermore, before computing the difference, we normalized the reputation scores for each transaction, giving a score of 1 to the merchant with the highest score in this dimension and a score of 0 to the merchant with the lower score. This normalization allows us to compare the scores across different dimensions without the need to consult summary statistics for each variable. This process yields a weight associated with each dimension, which can be interpreted as an ordinal measure of the "value" (or increase/decrease in pricing power) of having that pair associated with a particular transaction.

The analysis with these 21 dimensions is presented in Table 19. We show that negative ratings on some topics matter more than positive or neutral ratings. As before, these estimates reflect changes in the seller's average pricing power across products *after* taking their average numerical score and level of experience into account, and highlight the importance of the value contained in text-based reputation.[20] We present further discussion on the importance of the different dimensions of reputation in Section 6.4.

Buyer heterogeneity and the distinguishing dimension of reputation: The last hypothesis (H4) from Section 3 states that "*All else equal, a higher score on a seller's distinguishing*

[20] Regressions involving the average price premium as the dependent variable yield similar results, and are omitted for brevity. Further, our results are qualitatively robust to different values of c. We checked the correlation matrix and found that 94% of the variable pairs had a correlation below 0.3. We also performed tests for multi-collinearity such as the VIF (Variance Inflation Factor) test. Our analysis reveals that multi-collinearity was not a significant concern in our dataset.

characteristic leads to a higher price premium." In order to test this hypothesis, we created a set of dummies, $Dummy_i$ that are equal to 1 when the respective dimension

has the highest score across the positive, neutral, and negative (clustered) dimensions. For example, if *positive delivery* is the best dimension for a seller, then the respective dummy for this seller will be equal to 1 for this variable, while all the other 20 variables will be 0. Thereafter, we constructed three new variables from the 21 dimensions and the respective dummy variables that represent which of the dimensions has the maximum score for a given seller. These three new variables are: $\sum PositiveDimension_i \cdot Dummy_i$, where $i \in (1, 8)$ $\sum NeutralDimension_j \cdot Dummy_j$, where $j \in (9, 13)$ and $\sum NegativeDimension_k \cdot Dummy_k$, where $k \in (14, 21)$. Recall that in any given transaction, only one of the 21 dummy variables can be equal to 1. Estimating an equation with the above additional independent variables helps us evaluate our last hypothesis that sellers can be identified by a single distinguishing characteristic, and that an increase in the score of that dimension increases pricing power.

The estimates from this analysis are presented in Table 20. In column (1), we present the estimates from running the regressions on the sample where c is equal to 0.0.[21] The main coefficients of interest in this Table are $\sum PositiveDimension_i \cdot Dummy_i$,$\sum NeutralDimension_j * Dummy_j$ and $\sum NegativeDimension_k \cdot Dummy_k$.

We find that the coefficient of $\sum PositiveDimension_i \cdot Dummy_i$ is positive and significant, while the coefficients of $\sum NeutralDimension_j * Dummy_j$ and $\sum NegativeDimension_k \cdot Dummy_k$ are not significant. This supports our hypotheses of the distinguishing dimension, and suggests that consistent with our conjecture, different buyers do place a different weight on different dimensions. As a robustness check, in column (2) we present the estimates from running the regressions on the sample where c is equal to 0.25. We find that all three coefficients are significant, and in the expected direction. However, the coefficient of $\sum PositiveDimension_k \cdot Dummy_k$ is larger in magnitude than $\sum NegativeDimension_k \cdot Dummy_k$, which provides further support for our hypothesis

[21] Recall that c is the probability with which a consumer clicks on the next page of the seller's feedback profile. We also ran regressions for other values of c, and found that our results are qualitatively very similar.

4. These results underline the importance of treating reputation as multidimensional rather than a single numerical score; buyers are heterogeneous in what they consider important, and sellers can differentiate themselves on either the dimension they see as mattering the most, or on what they consider themselves most accomplished at.

6.3. Predicting Market Outcomes Using Text Postings

A natural next step is to see if we could achieve similar results using just the reputation characteristics captured by the numeric variables that appear in each merchant's reputation profile. We do have evidence that text contains additional information since our R^2 values with the additional text-extracted and clustered dimensions are higher.

To strengthen this further, we train a predictive model to predict which of a set of competing sellers will make a sale. Such a prediction can be based on the posted prices and on the numeric and text reputation of each merchant. We used a decision tree classifier, specifically the C4.5 classifier Quinlan (1992). Our choice of using decision trees was motivated by the ability of decision trees to capture non-linear interactions of textual and numeric data. The goal of our classifier is to take as input a pair of merchants and then decide which of the two will make a sale.

For the classifier generation, the training set is the transactions that took place in the first four months and the test set is the transactions in the last two months of our data set. Table 17 summarizes the results for different sets of features used. The 55% accuracy when using only prices as features indicates that customers rarely choose a product based solely on price. Rather, as indicated by the 74% accuracy, they also consider the reputation of the merchants as expressed in the numeric scores in the reputation profile. The question is, of course, if customers consider the text of the postings. We observed that the prediction accuracy in the test set increases to 87%-89% when using the textual reputation variables. In fact, accuracy decreased only slightly (from 89% to 87%) when we removed the numeric variables from the input, as indicated by the results in Table 17. This is compelling and further evidence that the text information can capture the information in the numeric variables, but not vice versa.

6.4. Further Discussion of our Experimental Findings

Overall, we observed that there are a small number of "real-life" dimensions of reputation matter to the buyers. However, customers place different emphasis on different aspects of reputation and react differently to positive or negative comments across each dimension.

- ***Problem Response (Misc):*** Price premiums go down when a merchant is evaluated on miscellaneous problem responses. Interestingly, even when buyers provide positive evaluations on this dimension, price premiums decrease. Although this seems counterintuitive, it could indicate that buyers prefer a problem-free transaction, and any comments (even positive) about problems can affect a transaction adversely.

- ***Customer Service:*** This is an important dimension, associated with significant variation in price premiums. Notice that price premiums are affected significantly by simple modifier-dimensions such as "superb service", "quick service" or "happy service." Further, the coefficients in Table 19, we can also see that customers pay substantially more attention to *negative* comments about customer service. This is consistent with an electronic market environment where buyers and sellers are separated by time and distance. In contrast, buyers reward good customer service with comparatively smaller price premiums.

- ***Packaging:*** All transactions involve shipping, and it is not surprising that buyers pay attention to the packaging of the product. In our data we observed a statistical significant effect only on positive individual modifier-pairs on packaging (e.g., for comments like "great packaging," or "perfectly packaged"); we did not observe a statistically meaningful decrease when the comments on packaging were negative. A surprising result was the positive coefficient for the negative packaging dimension in Table 19: we attribute this result to the relative low importance of packaging for software products, and to possible noise introduced by the manual dimension clustering process. We should mention that the frequency of negative comments on packaging were comparatively rare in our data.

- ***Delivery:*** The reputation for consistent, fast and seamless delivery of a product considerably increases the overall pricing power of a seller. Not surprisingly, both, the extent of order fulfillment and the speed at which it is executed, matter in such transactions. In contrast, problems with the shipping, such as items being sent to the wrong address or even instances of outright fraud such as products that never arrived or were received, are a major source of frustration. Our results show that comments about reneging such as "never delivered" or " not shipped" significantly hurt the price premium achievable by a seller over its competitors in future transactions.

- ***Product-specific Comments:*** In general, we observed that product-specific comments tend to decrease price premiums. Given that the act of describing a product appropriately by the seller on an electronic market constitutes an implicit guarantee for the buyer, any deviation from the promised description is tantamount to product misrepresentation and constitutes a violation of trust. Positive feedback postings on this dimension like " just advertised" improve the ability of a seller to charge higher prices in the future. When this is not the case (for example, when a buyer receive a "wrong CD" or "wrong game" or "wrong book"), the respective negative postings decrease the pricing premiums. This highlights the importance of representation in electronic markets.

- ***Overall:*** A large fraction of the feedback postings commented on the overall quality of the transaction and on the overall sentiment towards the seller: not surprisingly, positive comments (e.g., "awesome transaction" or "totally satisfied") improve pricing power. Our results also show that comments like "very recommended", "A +++ seller" or "very impressive" affect price premiums, but the magnitude of the increase is relatively modest. More importantly, we observe that negative "overall" comments, although infrequent, decrease pricing power substantially. Comparing the magnitudes of positive and negative comments along this dimension, we observe that negative comments are almost *five times* more important than the positive ones.

using completely automated techniques for assessing the content of textual feedback, without the need for expensive, manual inspection of the comments left by the buyers.

We believe that our technique can be successfully deployed over existing reputation mechanisms, thereby significantly improving their impact. Since most reputation mechanisms rely on voluntary reporting of transaction outcomes, buyers do not leave feedback on every transaction.[22] An understanding of the different dimensions of reputation might actually provide stronger incentives for more buyers to leave feedback.[23] We discuss the managerial implications of our results further in what follows.

7. Contributions and Managerial Implications

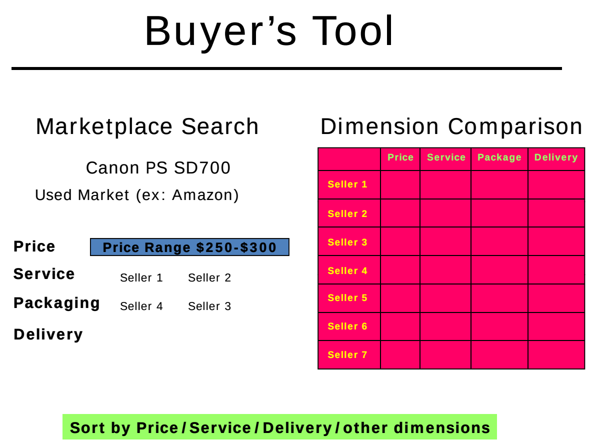
Our paper develops and implements a new framework for identifying the dimensions of a seller's reputation that online buyers actually value. Textual comments from buyers are important cues for improving the quality of electronic commerce. The study of the text feedback allows us to automatically identify the dimensions of reputation that are important. Our study is the first to examine the effect that positive and negative comments have across each dimension. Furthermore, we can now infer the actual economic value of each feedback posting (for instance, see Table 21, page 49). As documented in Section 6.4, page 26, positive and negative opinions have a different effect when evaluating a different dimension. By itself, this analysis sheds light on contradicting results from previous studies, where positive feedback seems more important than negative in some cases, and vice versa in others. We have shown that the relative importance of each is dimension-specific.

We present and experimentally confirm a new hypothesis: the dimension that differentiates a seller, or the fulfillment characteristics on which the seller's performance relative to its competitors is highest independently explains variation in pricing power, even after accounting for average differences in scores across the dimensions. In other words, we present evidence that customers look for a specific reputation characteristic when buying from a given seller, and they are not uniform in the importance they place on each of the dimensions of reputation. This is consistent with a natural economic theory that posits buyer heterogeneity. It is also of managerial significance because it provides a basis for understanding the extent to which a seller can benefit from improving on its distinguishing dimension, or from choosing a different positioning.

[22] Indeed, Zeckhauser and Resnick (2002) report that about 50% of eBay transactions in their data set were rated, whereas Dellarocas et al. (2004) report a higher, between 50-70% response rate.

[23] Amazon.com has recently adopted some elements of this strategy, asking users to evaluate merchants on specific dimensions (delivery, item representation, and customer service).

Figure 3 **A snapshot of the buyer tool.**



As the number of professional merchants who sell using retailing platforms like that of Amazon.com increases, our research insights become increasingly relevant. We rely on automated text-mining techniques which require minimal manual user effort to operationalize. The outcome of our techniques can also thus be more easily interpreted and acted upon. Our findings are also relevant to a platform like Amazon or eBay in informing their design of a more effective online reputation system. In particular, marketplace tools nased on the methods described in this paper are currently under preliminary development, and have attracted commercial interest, as illustrated below.

- **Buyers' tool:** More specifically, our techniques and results make it possible to build a tool which allows buyers to efficiently search the electronic market for sellers that satisfy particular criteria on each fulfilment dimension. For example, a buyer who wants to buy expensive electronic equipment might be looking for sellers with excellent scores in the *packaging* dimension, even if they are not that fast on shipping. This kind of system can substantially improve the efficiency of electronic markets and the accessibility of the reputation feedback. It will enhance buyer confidence that their opinion is important and is taken into account by other customers. This will make reputation systems more prominent parts of current electronic marketplaces.
- **Sellers' tool:** Our research also makes a tool that lets sellers assess the value of their own

Figure 4 **A snapshot of the seller tool.**

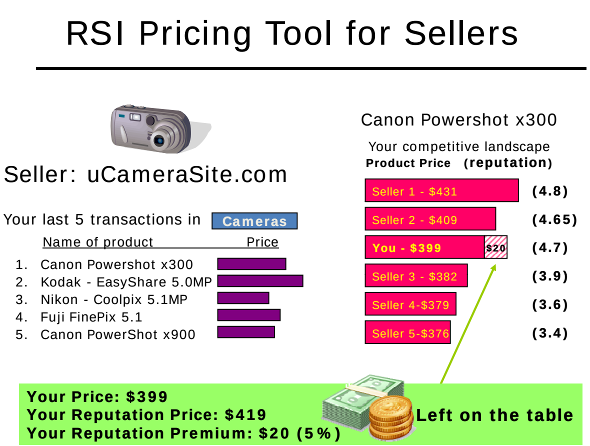
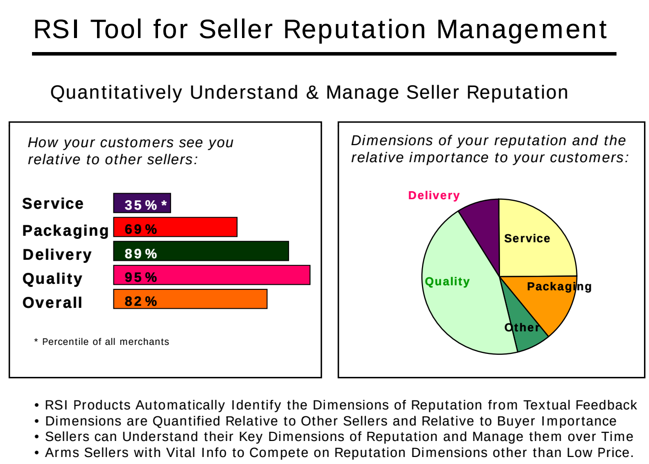


Figure 5 **A snapshot of the seller tool.**



reputation across each dimension viable, towards potentially comparing their reputation with that of their competitors. Such a tool can have two goals, one tactical and one strategic. The tactical goal is to allow sellers to price their products accurately, so that the posted prices incorporate the value of their reputation, compared to the one of the competitors. The tool, using the classifier described in Section 6.3, can increase confidence that a seller might make a sale at the right price point. A broader strategic purpose is to enable sellers to assess their relative competitive advantages and disadvantages ion the different dimensions of reputation. By realizing that customers look for a specific reputation characteristic when buying from a given seller, sellers can focus on improving their performance on their best dimensions, or decide to grow on another dimension of reputation that they deem is more profitable.

The relevance of these findings is not restricted to sellers on ecommerce platforms rather, they may generalize to adding to our understanding of what dimensions of transactions buyers value in electronic commerce in general. In this regard, our findings are of value to any business which seeks a deeper understanding of what differentiates online commerce from its traditional bricks and mortar counterpart.

8. Concluding remarks

We have presented a new approach for identifying and quantifying the dimensions of value from online reputation. We characterize how both numerical and qualitative measures of reputation affect a seller's pricing power in a mediated electronic market. We have validated the predictions of this theory by combining the results of the estimation of an econometric model with a novel text analysis technique, contrasted the relative importance of positive and negative reputation, provided new evidence of buyers valuing different fulfilment dimensions non-uniformly, and demonstrated the predictive power of text feedback by using it in a sales prediction model. To the best of our knowledge, this represents the first study of this kind, and the first set of results that establishes in multiple ways the value of information contained in the text-based feedback of an online reputation system.

Our analysis of the information in qualitative text feedback is likely to gain importance as the fraction of used good exchanges taking place on trading networks that are not mediated by a central market maker increases. As these networks evolve toward being the platform for more complex trade, rather than just for the free exchange of files, online reputation takes on an increasingly important role in these decentralized trading environments. A similar evolution occurred on Usenet groups in the 1990's, many of which were used as electronic secondary markets, and which, in the absence of a central mediating authority, used purely text-based feedback as their reputation mechanism.

Our study suggests several directions for further research. First, it may be beneficial to modify the method by which the "weights" associated with the dimensions of text-based feedback are estimated, toward being able to associate a dollar value with each modifier-dimension pair. A second extension could identify and account for patterns of manipulation of online reputation based on analysis of the associated text. Due to non-reciprocity based reputation system, the possibility of such systematic evidence of such manipulation on Amazon is low. However, it may well show up in a larger dataset, or from a market that uses a trading mechanism where there is reciprocity in the feedback system (for example, eBay allows both buyers and sellers to rate each other after the completion of a transaction).

We study software sales. It is also worth conducting a similar study on data from a different set of products to analyze if buyer behavior varies across different product categories. Following work done in the context of shopbots (for example, Montgomery et al. (2004)), another extension might develop economic measures of how of seller characteristics are evaluated from a consumer's perspective. The website interface of Amazon's secondary market has parallels with a shopbot system in which multiple goods from different sellers and different categories are displayed in response to a single query from a prospective buyer. The development of this feature followed our paper (although we do not have direct evidence that it was on account of our research).

An interesting extension to our approach would be to identify variation in trader trust based on an analysis of the content of the text feedback; however, this is not something we pursue in this paper. Finally, our study has not yet exploited the information contained in the structure of the buyer-seller network. Observing how feedback from buyer-seller pairs evolves over time may yield information that enables buyers to better "benchmark" the feedback. Specific buyers may prove to be reliable "indicators" of specific categories of trade (or, in the terminology of Kleinberg (1999), they may become "hubs"). This represents an exciting path for further research.

Appendix A: Reputation and pricing power: A theoretical framework

We model an electronic marketplace in which m competing sellers offer a single product. There are M buyers of this product, which, in addition to the product, they also value n different *fulfillment characteristics* for the transaction. Examples of these characteristics might be speed of delivery, quality of packaging, post-sale support and so on.

Each seller is indexed by a characteristics vector $X = (X_1, X_2, ..., X_n)$, where X_i represents the seller's ability to provide the i-th dimension of fulfillment. We assume that the products sold by each seller are of identical quality (which is consistent with our data set comprising consumer software, though adding an extra characteristic x_0 to represent product quality would not affect our results).

Buyers differ in the extent to which they place importance on each of these characteristics, and each buyer is therefore indexed by a type vector $w = (w_1, w_2, ..., w_n)$, where a higher value of w_i indicates that the buyer places a relatively higher value on fulfillment characteristic i. Each buyer's type is drawn from a common distribution with distribution function $F(w)$.

When a buyer purchases a product from a seller, there is a *realized value of fulfillment* $z = (z_1, z_2, ..., z_n) \in Z$ (the quality of fulfillment provided by the seller on that specific transaction). If the price charged by the seller is p, the value that the buyer gets from this transaction is:

$$u(w, z) - p, \tag{8}$$

where $u(w, z)$ is increasing in each component of its arguments. For example, $u(w, z)$ might be a weighted average of the realized fulfillment values.

After each transaction, the buyer posts a feedback set which contains the seller's ID, the buyer's ID, and information about the fulfillment on that transaction. There is consequently a *feedback set* $t(k) = \{s(k), b(k), \phi(k)\}$ associated with each transaction k, where the value of $s(k)$ identifies the seller, the value of $b(k)$ identifies the buyer, and $\phi(k)$ contains information about the quality of fulfillment. In most reputation systems, $\phi(k)$ contains a numerical score rating the overall quality of the transaction, along with unstructured text describing some of the dimensions of the transaction. We model $\phi(k)$ as a vector containing an average numerical score $\phi(k, 0)$ and a score $\phi(i, k)$ corresponding to each fulfilment characteristic i.

Prior to a transaction, a buyer may observe the entire set of feedback $\{t(k)\}$ for the electronic market. For any seller j, the buyer can thus identify the feedback profile

$$T(j) = \{t(k) : s(k) = j\}, \tag{9}$$

Denote the set of all such sets as T. Buyers map this profile $T(j)$ to an expected fulfilment characteristics vector using a common mapping $\Gamma : T \to Z$. We define a seller i as having a better reputation than seller j if

$$E[u(w, \Gamma(T(i))] \geq E[u(w, \Gamma(T(j))], \tag{10}$$

where the expectation is taken over the w according to the distribution of buyer types $F(w)$. Seller i could thus have a better reputation for one or more of the following reasons:

1. Seller i has a better average numerical reputation score than seller j.

2. The scores assigned to seller i on any one of the fulfilment characteristics are higher than those assigned to seller j.

3. The scores assigned to seller i on the set of fulfilment characteristics that most buyers care about are higher than those assigned to seller j.

4. $T(i)$ is a larger set than $T(j)$, and the estimate of average reputation generated by the mapping Γ reflects risk aversion among potential buyers.

Appendix B: Tables with Results and Robustness Checks

Variable	Observations	Mean	Standard Deviation	Min	Max
Avg PricePremium	9484	10.57	80.550	.006	648.12
Reg PricePremium	107922	8.87	104.272	-2190	1016.34
Product Price	107922	191.85	238.460	17.99	1699.99
Sale Price	107922	203.32	241.67	0.99	2200
Seller Rating	107922	4.43	0.35	1	5
Seller Life	107922	5040.68	34183.64	1	277309
Seller Condition	107922	4.89	0.40	1	5
Competitor Price	107922	194.44	222.32	1.25	1909.5
Competitor Rating	107922	4.45	0.16	1	5
Competitor Life	107922	22323.6	23242.2	1	275900
Competitor Condition	107922	4.88	0.135	2	5
Number of Competitors	107922	14.37	6.56	1	34

Table 2 **Summary statistics of numeric variables**

Variable	Observations	Mean	Standard Deviation	Min	Max
AvgPricePremium	2850	8.782418	64.17632	-1012.15	512.34
RegPricePremium	36707	5.60369	92.30909	-2091.02	1016.34
Product Price	36707	168.1202	196.0477	0	1079.99
Sale Price	36707	170.8201	166.7746	4.98	1115.98
Seller Condition	36707	4.879778	.4388896	1	5
Seller Rating	36707	4.447053	.3559633	1	5
Seller Life	36707	4649.975	32344.57	1	269502
Competitors Price	36707	165.2152	159.3986	5	2190
Competitor Rating	36707	4.448226	.5223408	1	5
Competitor Life	36707	17246.41	54566.9	1	277311
Competitor Condition	36707	4.898085	.3246085	1	5
Competitors	36707	15.416	6.421855	1	34

Table 3 **Summary statistics of numeric variables for Category 1 (Business and Productivity)**

Variable	Observations	Mean	Standard Deviation	Min	Max
AvgPricePremium	1683	13.36933	119.163	-655.7	464.59
RegPricePremium	17766	16.96234	165.0254	-946.18	779.89
Product Price	17766	364.0664	318.996	39.99	1499.99
Sale Price	17766	418.1018	339.6169	8.94	1944.18
Seller Condition	17766	4.919284	.3512876	1	5
Seller Rating	17766	4.354838	.3295088	1	5
Seller Life	17766	5620.205	35852.08	1	277309
Competitor Price	17766	401.1381	325.6468	8.94	1944.18
Competitor Rating	17766	4.373755	.5181269	1	5
Competitor Life	17766	23721.83	64063.19	1	277307
Competitor condition	17766	4.904086	.3367502	1	5
Competitors	17766	12.27789	5.503406	1	28

Table 4 **Summary statistics of numeric variables Category 2 (Graphics)**

Variable	Observations	Mean	Standard Deviation	Min	Max
AvgPricePremium	978	12.57742	30.13679	-78.49	127.5
RegPricePremium	10629	7.970027	35.2827	-231.85	211.03
Product Price	10629	103.6505	300.7201	29.99	1679
Sale Price	10629	113.4063	301.9741	.23	1698.98
Seller Condition	10629	4.895098	.366302	1	5
Seller Rating	10629	4.442477	.3486588	1	5
Seller Life	10629	5452.536	35935.64	1	274578
Competitor Price	10629	105.4356	294.2604	.15	1829.34
Competitor Rating	10629	4.479928	.5417333	1	5
Competitor Condition	10629	4.849939	.4041156	1	5
Competitor Life	10629	20722.8	62116.01	1	277312
Competitors	10629	14.4404	5.960315	1	23

Table 5 Summary statistics of numeric variables Category 3 (Utilities and Security)

Variable	Observations	Mean	Standard Deviation	Min	Max
AvgPricePremium	1892	13.88572	85.16937	-469.57	301.94
RegPricePremium	18560	9.781514	113.176	-531.89	531.89
Product Price	18560	221.7273	136.4061	91.99	599.99
Sale Price	18560	196.3616	135.3116	32.99	651.4
Seller Condition	18560	4.887128	.4187991	1	5
Seller Rating	18560	4.466811	.4226316	1	5
Seller Life	18560	4386.464	31691.68	1	269495
Competitor Price	18560	186.5796	121.5053	33.29	651.4
Competitor Rating	18560	4.490453	.4977473	1	5
Competitor Condition	18560	4.869105	.4063752	1	5
Competitor Life	18560	21983.1	62147.07	1	277306
Competitors	18560	16.34229	5.388391	1	27

Table 6 Summary statistics of numeric variables Category 4(Development)

Variable	Observations	Mean	Standard Deviation	Min	Max
AvgPricePremium	2101	11.15895	74.57983	-509.17	256.21
RegPricePremium	24260	9.242575	79.33625	-654.58	666.71
Product Price	24260	140.5953	170.3248	29.99	799.99
Sale Price	24260	152.6693	171.428	1.49	857.97
Seller Condition	24260	4.92939	.3317591	1	5
Seller Rating	24260	4.440655	.3258944	1.4	5
Seller Life	24260	4835.397	34017.11	1	277306
Competitor Price	24260	143.4254	161.4424	1.49	858.9
Competitor Rating	24260	4.464386	.4947219	1	5
Competitor Condition	24260	4.893116	.3582716	1	5
Competitor Life	24260	26536.48	69031.27	1	277312
Competitors	24260	14.83937	7.361648	1	31

Table 7 Summary statistics of numeric variables Category 5 (Operating Systems)

Variable	Observations	Mean	Std. Dev.	Min	Max
never sent	107922	0.0000658	0.0022875	- 0.0526316	0.0526316
very recommended	107922	- 0.0001903	0.0022041	- 0.0769231	0.0213049
never received	107922	- 0.0001495	0.0114964	- 0.25	0.25
never delivered	107922	- 0.0000926	0.0016085	- 0.0625	0.0666667
always excellent	107922	- 0.0000394	0.0008319	- 0.024	4.77e-09
never responded	107922	- 0.0000469	0.0041165	- 0.1666667	0.1666667
great buying	107922	- 0.0001057	0.0028721	- 0.0769231	0.0769231
super transaction	107922	- 0.0000656	0.0014872	- 0.0714286	0.0714286
great person	107922	- 0.0000858	0.0015508	- 0.0833333	0.0588235
never heard	107922	- 0.000176	0.0016118	- 0.03	0.04
not received	107922	0.0000864	0.0034954	- 0.0555556	0.0526316
excellent communications	107922	- 0.000075	0.001125	- 0.0233918	0.0233918
terrific condition	107922	-9.79e-07	0.0000511	- 0.005	0.0012508
not have	107922	- 0.0022077	0.0224192	- 0.6666667	0.5
not received	107922	0.000173	0.0053725	- 0.1	0.1
cancelled order	107922	- 0.0005719	0.0100573	- 0.3333333	0.0200049
definitely recommend	107922	- 0.0003166	0.0042976	- 0.1666667	0.0338983
not notified	107922	- 0.0000279	0.0007656	- 0.02	0.0026596
excellent seller	107922	0.0081056	0.0380053	-1	1
not delivered	107922	5.15e-06	0.0030023	- 0.1666667	0.1666667
not shipped	107922	- 0.0005664	0.0031205	- 0.0625	0.0625
poor condition	107922	- 0.0000625	0.0019898	- 0.0344828	0.0344828
fast seller	107922	0.0003116	0.0046761	- 0.0666667	0.0666667
not ordered	107922	- 0.0000805	0.0033164	- 0.1428571	0.1428571
perfectly packaged	107922	0.0004791	0.0040951	- 0.0350877	0.0264901
bad experience	107922	- 0.0002027	0.0019415	- 0.04	0.02
A+++ seller	107922	- 0.0001791	0.0028428	- 0.047619	0.04
wrong CD	107922	0.0000111	0.0005755	- 0.0205392	0.02
wrong address	107922	- 0.00027	0.0034527	- 0.0769231	0.0769231
wrong book	107922	- 0.0000628	0.0031658	- 0.3333333	0.0555556
wrong game	107922	- 0.0000296	0.0014374	- 0.0666667	0.0666667
awesome transaction	107922	- 0.0000558	0.0016608	- 0.0625	0.0200046
best seller	107922	- 0.0001497	0.0022142	- 0.0260756	0.0246914
wrong item	107922	0.0094492	0.0131771	- 0.1	0.1
awesome service	107922	0.0023027	0.0030124	- 0.0454545	0.037037
very slow	107922	- 0.0002231	0.002425	- 0.0201702	0.0100266
quickly advertised	107922	- 0.0002472	0.0028828	- 0.0454545	0.0909091
several weeks	107922	- 0.0001075	0.0030191	- 0.047619	0.047619
late shipped	107922	- 0.0005365	0.0025033	- 0.0236686	0.02
defective product	107922	- 0.0000367	0.0009372	- 0.125	0.0129032
bad shape	107922	-1.13e-06	0.0015334	- 0.0454545	0.0454545
not advertised	107922	- 0.0001909	0.0036066	- 0.0769231	0.0769231
happy service	107922	- 0.000217	0.0020874	- 0.0242424	0.0242424
very impressed	107922	- 0.0003181	0.0034756	- 0.1666667	0.1111111
later received	107922	- 0.0004482	0.0069506	- 0.2	0.0363636
not arrived	107922	- 0.0001429	0.0050211	- 0.1666667	0.0833333
top quality	107922	- 0.0003333	0.0032658	- 0.047619	0.0338983
friendly service	107922	- 0.002613	0.0218787	-1	1

Table 8 Summary statistics of text variables (modifier-dimension pairs)

Variable	Observations	Mean	Std. Dev.	Min	Max
slow shipping	107922	- 0.0002223	0.0039393	- 0.0526316	0.0526316
superb service	107922	- 0.0000246	0.0017151	- 0.0625	0.0236686
fair condition	107922	- 0.0000553	0.0029515	- 0.3333333	0.0246914
perfect item	107922	0.0053321	0.0064172	- 0.0714286	0.0714286
totally satisfied	107922	- 0.0000489	0.0017618	- 0.037037	0.037037
excellent service	107922	- 0.00004	0.0008413	- 0.02	0.0000195
great merchant	107922	- 0.0001023	0.0026192	- 0.0273973	0.0273973
very fast	107922	- 0.0002404	0.0021852	- 0.0285714	0.0200043
excellent condition	107922	- 0.0000458	0.0027119	- 0.1666667	0.1666667
ahead arrived	107922	- 0.0007818	0.0052645	- 0.1428571	0.1428571
received product	107922	- 0.0000403	0.0007823	- 0.02	0.005
great company	107922	- 0.0003496	0.0036476	- 0.0555556	0.0555556
promptly described	107922	- 0.0003732	0.0033739	- 0.0333333	0.0909091
easy service	107922	- 0.0000147	0.0022246	- 0.02	0.02
quick service	107922	- 0.0028728	0.0148672	- 0.5	0.5
great packaging	107922	- 0.0009077	0.007408	- 0.2	0.1428571
exactly arrived	107922	- 0.0008714	0.0070406	- 0.3333333	0.3333333

Table 9 Continued: Summary statistics of text variables (modifier-dimension pairs)

Variable	Observations	Mean	Std. Dev.	Min	Max
d0	107922	.2240692	.4169698	0	1
Dummy Negative Fulfillment	107922	.1229314	.3283599	0	1
Dummy Negative Misc	107922	.0266581	.161083	0	1
Dummy Negative Overall	107922	.0257871	.1585005	0	1
Dummy Negative Packaging	107922	.0364152	.1873218	0	1
Dummy Negative Pricing	107922	.0361743	.1867244	0	1
Dummy Negative Product Quality	107922	.5170771	.4997106	0	1
Dummy Negative Representation	107922	.0108875	.1037741	0	1
Dummy Neutral Fulfillment	107922	.1968088	.3975884	0	1
Dummy Neutral Misc	107922	.5882119	.4921594	0	1
Dummy Neutral Overall	107922	.1405923	.3476021	0	1
Dummy Neutral Representation	107922	.0375271	.1900504	0	1
Dummy Neutral Service	107922	.03686	1884188	0	1
Dummy Positive Fulfillment	107922	.4280592	.4947997	0	1
Dummy Positive Misc	107922	.0326532	.1777281	0	1
Dummy Positive Overall	107922	.0615167	.2402767	0	1
Dummy Positive Packaging	107922	.0638239	.2444401	0	1
Dummy Positive Pricing	107922	.0473953	.212484	0	1
d18	107922	0.1199941	0.3249561	0	1
Dummy Positive Representation	107922	.0831248	.2760721	0	1
Dummy Positive Service	107922	.1634328	.3697618	0	1
Negative Fulfillment	107922	-.0669293	.4097874	-1	1
Negative Misc	107922	.0727223	.4256386	-1	1
Negative Overall	107922	-.0702025	.3131249	-1	1
Negative Packaging	107922	-.0617229	.3157939	-1	1
Negative Pricing	107922	-.05948	2592682	1	1
Negative Product quality	107922	-.0270006	.356886	1	1
Negative Representation	107922	.346967	5318592	1	1
Negative Service	107922	-.077595	.3179903	-1	1
Neutral Fulfillment	107922	-.0695362	.3644121	-1	1
Neutral Misc	107922	.0668392	.4338325	-1	1
Neutral Overall	107922	-.0367441	.3115605	-1	1
Neutral Representation	107922	-.0702883	.2825511	-1	1
Neutral Service	107922	-.0753926	.3418242	-1	1
Positive Fulfillment	107922	.179798	4740336	1	1
Positive Misc	107922	-.0639868	.3451978	-1	1
Positive Overall	107922	.059836	4278125	1	1
Positive Packaging	107922	-.0284115	.3897397	-1	1
Positive Pricing	107922	-.0474962	.3812951	-1	1
Positive Product quality	107922	.0563512	.4251948	-1	1
Positive Representation	107922	-.0400379	.438151	1	1
Positive Service	107922	.082349	.183255	1	1

Table 10 Continued: Summary statistics of text variables (modifier-dimension pairs)

Variable	Estimates	Estimates
$Constant$	1.66 (1.24)	8.01^{***} (0.44)
$Log(ProductPrice)$	- 0.15 (0.19)	$.41^{***}$ (0.085)
$RegDiffRating$	- 0.14^{***}(0.007)	$.187^{***}$ (0.012)
$Log(RegDiffLife)$	0.011^{***}(0.0005)	0.051^{***} (0.003)
$RegDiffCondition$	0.27^{***}(0.01)	0.017^{***}(0.007)
$Log(Competitors)$	- 2.02^{***}(0.01)	- 1.42^{***}(0.026)
$R^2(\%)$	49.35	13.89

Table 11 The effect of average reputation and level of experience on pricing power with Regular Price Premium as the dependent variable. Robust standard errors are in parenthesis. The first column displays results with product-seller fixed effects. The second column displays results with product, seller and competitor fixed effects. *, ** and * denote significance at 10%, 5% and 1% respectively.**

Variable	Estimates
$Constant$	-1.27 (2.04)
$Log(ProductPrice)$	$.75^{*}$ (0.39)
$AvgDiffRating$	-0.37^{***} (0.06)
$Log(AvgDiffLife)$	0.048^{***} (0.008)
$AvgDiffCondition$	0.11^{***} (0.06)
$Log(Competitors)$	0.43^{***} (0.03)
$R^2(\%)$	13.2

Table 12 The effect of average reputation and level of experience on pricing power with Average Price Premium as the dependent variable. Robust standard errors are in parenthesis. This regression has OLS with product-seller fixed effects. *, ** and * denote significance at 10%, 5% and 1% respectively.**

Variable	Estimate
Log(ProductPrice)	$-0.84^{***}(0.18)$
RegDiffrating	$-0.11^{***}(0.012)$
RegDiffcondition	$0.021^{***}(0.006)$
Log(RegDifflife)	$0.038^{***}(0.0007)$
Log(Competitors)	$-1.87^{***}(0.011)$
[Modifier, Dimension] Pair	**Estimate**
[very recommended]	$12.47^{***}(2.38)$
[always excellent]	$141.98^{***}(34.52)$
[excellent communication]	$18.38^{**}(7.61)$
[terrific condition]	$857.55^{***}(108.15)$
[excellent condition]	$8.53^{***}(1.93)$
[excellent seller]	$0.7^{*}(0.37)$
[fast seller]	$9.76^{***}(2.12)$
[best seller]	$48.8^{***}(4.45)$
[perfectly packaged]	$37.4^{***}(4.26)$
[A+++ seller]	$49.15^{***}(3.87)$
[happy service]	$21.56^{***}(5.11)$
[friendly service]	$8.21^{***}(0.64)$
[easy service]	$46.17^{***}(5.87)$
[awesome service]	$26.05^{***}(4.24)$
[very impressive]	$21.3^{***}(2.91)$
[top quality]	$16.77^{***}(2.41)$
[perfect item]	$6.92^{**}(1.74)$
[excellent purchase]	$89.72^{***}(15.47)$
[great merchant]	$49.33^{***}(4.68)$
[great company]	$21.77^{**}(2.69)$
[promptly described]	$4.7^{**}(1.09)$
[not have]	$-3.48^{***}(0.17)$
[not notified]	$-103.69^{***}(23.75)$
[not ordered]	$-13.4^{***}(2.09)$
[not received]	$-9.44^{***}(3.82)$
[not delivered]	$-12.25^{***}(3.02)$
[wrong address]	$-17.54^{***}(3.13)$
[wrong item]	$-2.5^{***}(0.89)$
[never sent]	$-16.61^{***}(4.8)$
[never responded]	$-4.87^{**}(2.15)$
[still waiting]	$-6.38^{***}(1.14)$
[cancelled order]	$-5.01^{**}(2.4)$
[bad experience]	$-13.73^{***}(4.86)$
[not advertised]	$-9.93^{***}(1.41)$
[wrong CD]	$-32.58^{***}(12.36)$
[wrong Book]	$-17.39^{***}(1.87)$
[wrong Game]	$-41.54^{***}(5.31)$
[defective product]	$-13.19^{***}(4.4)$
$R^2(\%)$	62.7

Table 13 **Summary of the dimension-modifier pairs in text-based feedback that influence a seller's pricing power most strongly. Robust standard errors are in parenthesis. The dependent variable is the log of the Regular Price Premium. These estimates are based on regressions with product and seller fixed effects. *, ** and *** denote significance at 10%, 5% and 1% respectively.**

Variable	**Estimate**
Log(ProductPrice)	0.13(0.09)
RegDiffrating	$0.027^{*}(0.015)$
RegDiffcondition	$0.021^{***}(0.006)$
Log(RegDifflife)	$0.01^{***}(0.005)$
Log(Competitors)	$-1.38^{***}(0.027)$
[Modifier, Dimension] Pair	**Estimate**
[very recommended]	4.17(1.26)
[excellent seller]	1.6(0.43)
[A+++ seller]	$23.12^{***}(3.64)$
[superb service]	23.67(6.01)
[perfect item]	$3.83^{**}(1.77)$
[great merchant]	$8.31^{***}(3.3)$
[excellent purchase]	$32.83^{***}(9.17)$
[very fast]	$15.33^{***}(4.65)$
[great company]	$5.55^{**}(2.28)$
[easy service]	$4.26^{***}(1.25)$
[happy service]	$13.28^{***}(4.3)$
[great packaging]	$2.51^{***}(0.69)$
[excellent value]	$6.79^{***}(1.75)$
[excellent response]	$4.33^{***}(1.75)$
[slow shipping]	$-5.9^{***}(1.35)$
[very slow]	$-13.58^{***}(1.94)$
[never delivered]	$-20.37^{***}(4.17)$
[never heard]	$-65.16^{***}(7.70)$
[several weeks]	$-14.46^{***}(3.27)$
[yet have]	$-6.07^{***}(2.19)$
[not received]	$-9.36^{***}(1.69)$
[not arrived]	$-10.56^{***}(3.76)$
[not shipped]	$-5.15^{**}(2.29)$
[late shipped]	$-11.06^{**}(5.35)$
[bad experience]	$-16.43^{***}(3.3)$
[poor condition]	$-35.54^{***}(7.16)$
[not advertised]	$-18.40^{***}(3.82)$
[wrong address]	$-5.49^{*}(1.02)$
[wrong item]	$-1.31^{**}(0.61)$
[wrong CD]	$-22.75^{***}(9.08)$
$R^2(\%)$	19.66

Table 14 **Summary of the dimension-modifier pairs in text-based feedback that influence a seller's pricing power most strongly. Robust standard errors are in parenthesis. The dependent variable is the log of the Regular Price Premium. These estimates are based on regressions with product, seller and competitor fixed effects. *, ** and *** denote significance at 10%, 5% and 1% respectively.**

Variable	**Estimate**
Log(ProductPrice)	0.14(0.44)
AvgDiffrating	−0.11(0.095)
AvgDiffcondition	0.01(0.05)
Log(AvgDifflife)	0.099***(0.009)
Log(Competitors)	0.356***(0.045)
Modifier, Dimension] Pair	**Estimate**
[very recommended]	15.21***(3.14)
[always excellent]	60.8***(20.25)
[super transaction]	10.47**(5.17)
[great person]	10.81***(3.76)
[awesome transaction]	13.28***(3.28)
[excellent communication]	8.06**(3.77)
[excellent condition]	7.58***(2.3)
[terrific condition]	1237.42***(354.67)
[fast seller]	5.97**(2.87)
[A+++ seller]	14.12**(6.86)
[definitely recommend]	11.04***(3.78)
[happy service]	13.75**(6.87)
[easy service]	17.45**(3.94)
[quick service]	3.89***(1.33)
[top quality]	9.7***(3.21)
[ahead arrived]	8.11***(2.87)
[exactly arrived]	5.83***(1.46)
[great shape]	2.66**(1.2)
[excellent product]	3.1**(1.28)
[very satisfied]	1.5**(0.75)
[just advertised]	4.0***(1.34)
[great buying]	12.42***(3.61)
[good seller]	3.2**(1.05)
[superb service]	36.68***(13.01)
[perfect item]	13.914***(3.05)
[totally satisfied]	12.72**(6.1)
[not have]	−3.33***(1.09)
[poor condition]	−15.73***(4.74)
[fair condition]	−10.57***(3.8)
[not notified]	−42.15***(15.14)
[not shipped]	−8.84**(4.19)
[wrong item]	−6.36***(2.49)
[bad shape]	−10.53**(4.86)
[slow shipping]	−8.03***(2.98)
[wrong address]	−12.94***(3.52)
[never got]	−1.71*(0.98)
[never received]	−5.12**(2.13)
[never delivered]	−26.73***(8.89)
$R^2(\%)$	51.32

Table 15 Summary of the dimension-modifier pairs in text-based feedback that influence a seller's pricing power most strongly. Robust standard errors are in parenthesis. The dependent variable is the log of the Average Price Premium. These estimates are based on regressions with product and seller fixed effects. *, ** and * denote significance at 10%, 5% and 1% respectively.**

Dimension	Human Recall	Computer Recall
Product Condition (Product Quality)	0.76	0.76
Price (Pricing)	0.91	0.61
Package (Packaging)	0.96	0.66
Overall Experience (Overall)	0.65	0.55
Delivery Speed (Fulfillment)	0.96	0.92
Item Description (Representation)	0.22	0.43
Product Satisfaction (Product Quality)	0.68	0.58
Problem Response (Misc)	0.30	0.37
Customer Service (Service)	0.57	0.50
Average	**0.66**	**0.60**

Table 16 **The recall of our technique compared to the recall of the human annotators. In parenthesis the corresponding dimensions that we identified when clustering the individual modifier-dimensions pairs.**

Features	Accuracy on Test Set
Price	55%
Price + Numeric Reputation	74%
Price + Numeric Reputation + Text Reputation	89%
Price + Text Reputation	87%

Table 17 **Predicting the merchant who makes the sale.**

Variable	Estimate	Estimate	Estimate	Estimate
RegDiffRating	-0.983(.0137)***	-0.975(.014)***	-0.87(.0135)***	-0.34(.017)***
RegDiffCondition	0.411(.0179)***	0.413(.018)***	0.59(.0176)***	0.26(.0178)***
RegDiffLife	-0.0929(.0024)***	-0.092(.0024)***	0.064(.01)***	-0.066(.002)***
RegDiffLife Squared			-0.05(.0006)***	
Variance RegDiffLife				0.946(.0159)***
Log Product Price	-0.0114 (.006)*	-0.011(.006)*	-0.015 (.006)**	-0.0165(.006)***
Log (Competitors)	0.5547(.0503)***	0.555 (.05)***	0.398(.049)***	0.3985(.049)***
R^2(%)	4.6	5.25	6.9	16.65

Table 18 **Results from the numeric variables that influence a seller's pricing power most strongly. Robust standard errors are in parenthesis. The dependent variable is the log of the Regular Price Premium. These estimates are based on regressions with product-seller fixed effects. *, ** and *** denote significance at 10%, 5% and 1% respectively. The first column includes all sellers in our sample, the second columns includes only sellers with less than 100,000 transactions, the third column controls for the rate of increase in seller size and the fourth column includes the variance of the numeric rating.**

Variable	Estimate	Estimate	Estimate	Estimate
RegDiffRating	-0.91485 (0.014238)***	-0.91059 (0.014366)***	-0.84711(0.013946)***	-0.3577(.0173)***
RegDiffCondition	-0.11952 (0.002682)***	-0.11901 (0.002694)***	0.604662 (0.0174)***	0.29769(.01748)***
RegDiffLife	0.41727 (0.017597)***	0.42053 (0.017917)***	0.57962 (0.01026)***	-0.0862(.0027)***
RegDiffLife Squared			-0.04922 (0.0007)***	
Variance RegDiffLife				0.8742(.0159)***
Log Product Price	-0.00949 (0.00618)	-0.0092 (0.00621)	-0.01213 (0.006043)	-.01313(.006)**
Log (Competitors)	0.72182 (0.04935)***	0.71871 (0.049661)***	0.58068 (0.04827)***	.56738(.0487)***
Negative Fulfillment	0.9058 (0.02276)***	0.89201 (0.023039)***	0.78817 (0.0223)***	0.7945 (.0225)***
Negative Misc	-0.34512 (0.0214)***	-0.35065 (0.02158)***	-0.26099 (0.02094)***	-0.439 (0.0211)***
Negative Overall	-0.45022 (0.05138)***	-0.45008 (0.05177)***	-0.47257 (0.05021)***	-0.382 (0.05)***
Negative Packaging	0.1765(0.07991)**	0.18575 (0.080985)**	0.142676 (0.078092)*	0.2360(.0787)***
Negative Pricing	-0.37298 (0.097532)***	-0.36952 (0.09787)***	-0.15848 (0.09535)*	-0.265(.0961)***
Negative Product Quality	-0.90038 (0.119665)***	-0.90301 (0.120208)***	-0.52578 (0.11705)***	-1.179 (.118)***
Negative Representation	0.500851 (0.015115)***	0.490415 (0.01522)***	0.456036 (0.014783)***	0.404 (.0150)***
Negative Service	-0.33007 (0.04704)***	-0.3261 (0.047457)***	-0.3556 (0.045972)***	-0.194(0.046)***
Neutral Fulfillment	-0.27286 (0.02593)***	-0.26056 (0.02617)***	-0.1461 (0.0254)***	-0.40467(.025)***
Neutral Misc	-0.93181 (0.02334)***	-0.92069 (0.02356)***	-0.94764 (0.0228)***	-0.8886 (.023)***
Neutral Overall	-1.17169 (0.093789)***	-1.18781 (0.09503)***	-1.13184 (0.09164)***	-1.071(.0924)***
Neutral Representation	0.278814 (0.058851)***	0.278148 (0.05905)***	0.207855 (0.057516)***	0.222(.058)***
Neutral Service	-0.18202 (0.043417)***	-0.17207 (0.04449)***	-0.10965 (0.04243)***	-.1601(.0428)***
Positive Fulfillment	0.09167 (0.018601)***	0.086293 (0.018779)***	0.04088 (0.01819)**	.0026 (.0184)
Positive Misc	-0.68029 (0.040474)***	-0.68008 (0.04072)***	-0.58382 (0.03957)***	-0.719 (.039)***
Positive Overall	0.08494 (0.02353)***	0.0909 (0.023782)***	0.00807 (0.02302)	0.0708(.0232)***
Positive Packaging	0.618855 (0.03336)***	0.62857 (0.03359)***	0.67039 (0.0326)***	0.6435 (.032)***
Positive Pricing	0.03711 (0.03198)	0.0282 (0.032426)	0.07101 (0.03125)**	0.0232 (.0315)
Positive Product Quality	-0.1742 (0.017711)***	-0.16886 (0.01785)***	-0.25546 (0.01734)***	-0.1781(.0174)***
Positive Representation	0.08558 (0.017243)***	0.08849 (0.01739)***	-0.00124 (0.01689)	0.08819 (.017)***
Positive Service	0.16423 (0.019137***	0.16087 (0.01929)***	0.04672 (0.01877)*	0.1276 (.0188)***
R^2(%)	13.5	14.5	15.5	22.7

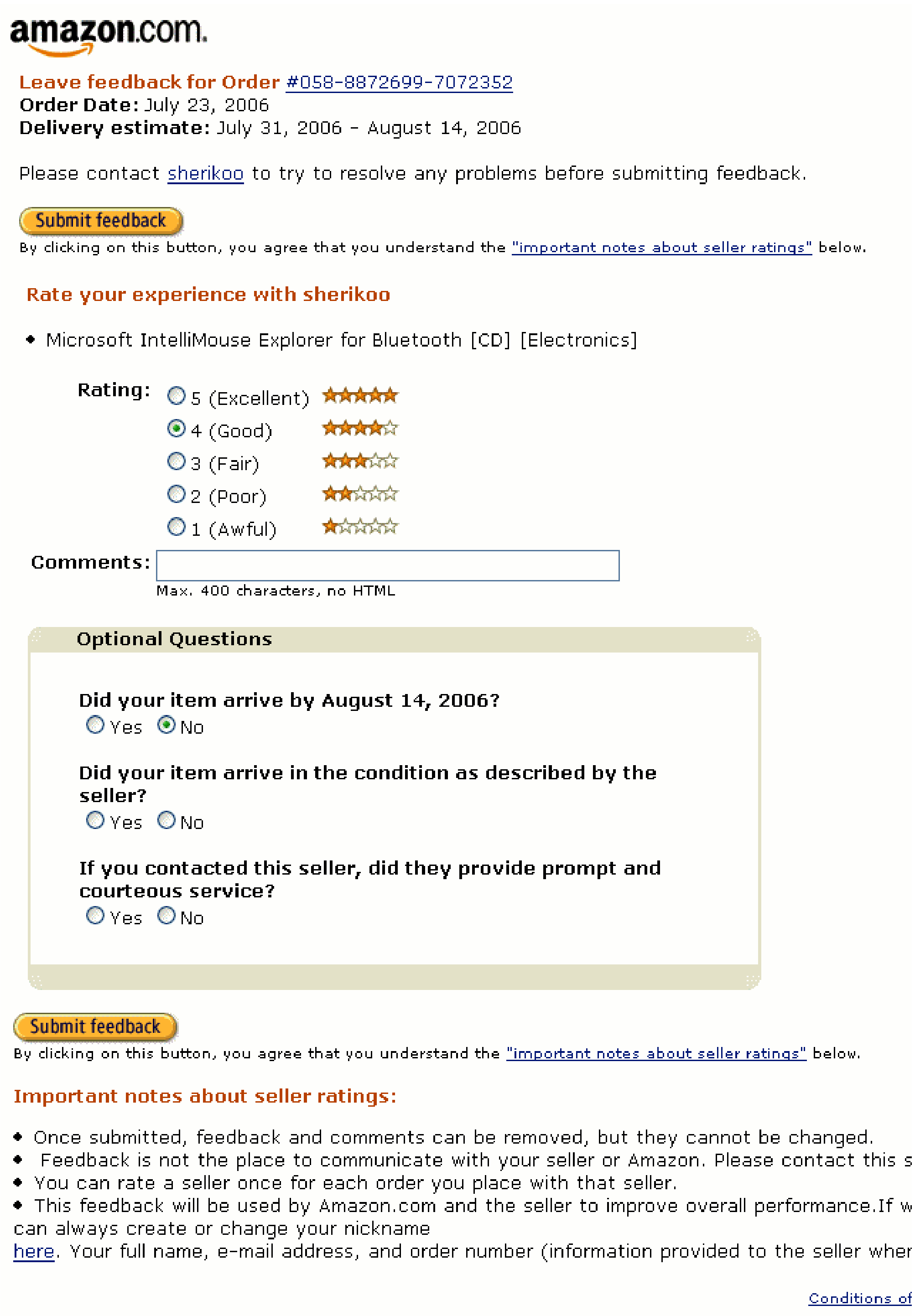
Table 19 **Results from the analysis of the dimensions in text-based feedback that influence a seller's pricing power most strongly. Robust standard errors are in parenthesis. The dependent variable is the log of the Regular Price Premium. These estimates are based on regressions with product-seller fixed effects. *, ** and *** denote significance at 10%, 5% and 1% respectively. The first column includes all sellers in our sample, the second columns includes only sellers with less than 100,000 transactions, the third column controls for the rate of increase in seller size and the fourth column includes the variance of the numeric rating.**

Variable	Estimate	Estimate
RegDiffRating	0.071(-5.760)***	-0.72(0.015)***
RegDiffCondition	-0.122(0.012)***	-0.18(0.002)***
RegDiffLife	0.161(0.051)***	0.42(0.017)***
Log Product Price	0.140(0.268)	-0.114(0.006)*
Log (Competitors)	0.241(0.093)***	0.45(0.05)***
Dummy Negative Fulfillment	-0.359(0.307)	-.066(0.161)
Dummy Negative Misc	-0.51(0.3)*	-0.2(0.16)
Dummy Negative Overall	-0.567(0.322)*	0.7(0.18)***
Dummy Negative Pricing	-0.58(0.31)*	0.27(0.17)
Dummy Negative Product Quality	-0.248(0.305)	0.27(0.18)
Dummy Negative Representation	-0.569(0.307)*	0.002(0.16)
Dummy Negative Service	-0.528(0.316)*	0.2(0.2)
Dummy Neutral Fulfillment	-1.067(0.989)	0.67(0.14)***
Dummy Neutral Misc	-1.119(0.997)	0.44(0.14)***
Dummy Neutral Overall	-1.192(0.994)	0.25(0.13)*
Dummy Neutral Representation	-1.143(0.985)	0.34(0.15)**
Dummy Positive Fulfillment	-0.24(0.127)*	0.16(0.1)
Dummy Positive Overall	-0.25(0.13)*	0.03(0.11)
Dummy Positive Packaging	-0.480(0.157)***	-0.051(0.11)
Dummy Positive Pricing	-0.383(0.091)***	0.002(0.01)
Dummy Positive Product Quality	-0.316(0.123)***	-0.026(0.1)
Dummy Positive Representation	-0.091(0.126)	0.13(0.11)
Dummy Positive Service	-0.381(0.133)***	-0.17(0.1)
Negative Fulfillment	-0.035(0.070)	0.98 (.027)***
Negative Misc	-0.055(0.063)	- 0.12 (.026)***
Negative Overall	0.094(0.080)	-0.529(.026)***
Negative Packaging	-0.412(0.081)***	-0.054(0.026)**
Negative Pricing	0.001(0.076)	-0.052(0.029)*
Negative Product Quality	-0.621(0.112)***	0.005(0.03)
Negative Representation	0.247(0.065)***	1.06(0.024)***
Negative Service	0.39(0.091)***	0.48(0.03)***
Neutral Fulfillment	-0.224(0.059)***	-0.99(0.025)***
Neutral Misc	-0.278(0.078)***	-0.798(0.04)***
Neutral Overall	-0.052(0.077)	-1.12(0.038)***
Neutral Representation	0.191(0.089)**	0.62(0.028)***
Neutral Service	0.258(0.091)***	0.82(0.026)***
Positive Fulfillment	-0.227(0.050)***	-0.45(0.027)***
Positive Misc	-0.208(0.086)***	-0.65(0.026)***
Positive Overall	-0.060(0.040)	0.087(0.024)***
Positive Packaging	0.313(0.073)***	0.62(0.025)***
Positive Pricing	-0.112(0.072)	-0.05(0.025)**
Positive Product Quality	-0.271(0.051)***	-0.092(0.022)***
Positive Representation	-0.342(0.051)***	0.018(0.02)
Positive Service	0.083(0.052)	0.29(0.025)***
Sum Positive	0.130 (0.059)**	0.323 (0.025)***
Sum Neutral	0.084 (0.079)	0.126(0.031)***
Sum Negative	-0.067(0.062)	-0.06 (0.03)**
$R^2(\%)$	19.2	18.3

Table 20 Regression showing that an increase in the score of a seller's distinguishing characteristic increases their pricing power. Robust standard errors are in parenthesis. The dependent variable is the log of the Regular Price Premium. These estimates are based on regressions with product-seller fixed effects. *, ** and * denote significance at 10%, 5% and 1% respectively.**

Figure 6 **A snapshot of the revised reputation system of Amazon.com, incorporating instructions that ask users to evaluate merchants across specific dimensions. (This version first appeared on Amazon.com in Spring 2006.)**

Figure 7 **A snapshot of the revised reputation system of Amazon.com, incorporating explicit form elements for the buyers to evaluate merchants across specific dimensions. (This version first appeared on Amazon.com in late Summer 2006.)**



amazon.com.

Leave feedback for Order #058-8872699-7072352
Order Date: July 23, 2006
Delivery estimate: July 31, 2006 - August 14, 2006

Please contact sherikoo to try to resolve any problems before submitting feedback.

Submit feedback
By clicking on this button, you agree that you understand the "important notes about seller ratings" below.

Rate your experience with sherikoo

- Microsoft IntelliMouse Explorer for Bluetooth [CD] [Electronics]

Rating: 5 (Excellent) 4 (Good) 3 (Fair) 2 (Poor) 1 (Awful)

Comments:
Max. 400 characters, no HTML

Optional Questions

Did your item arrive by August 14, 2006?
Yes No

Did your item arrive in the condition as described by the seller?
Yes No

If you contacted this seller, did they provide prompt and courteous service?
Yes No

Submit feedback
By clicking on this button, you agree that you understand the "important notes about seller ratings" below.

Important notes about seller ratings:

- Once submitted, feedback and comments can be removed, but they cannot be changed.
- Feedback is not the place to communicate with your seller or Amazon. Please contact this s
- You can rate a seller once for each order you place with that seller.
- This feedback will be used by Amazon.com and the seller to improve overall performance.If w can always create or change your nickname here. Your full name, e-mail address, and order number (information provided to the seller wher

Conditions of

Acknowledgments

The authors would like to thank Elena Filatova for the useful discussions and the pointers to related literature. We also thank Sanjeev Dewan, Alok Gupta, Bin Gu, and seminar participants at the Statistical Challenges in Ecommerce Research conference (SCECR 2006), Carnegie Mellon University, Columbia University, Microsoft Research, New York University, Polytechnic University, the University of California, Irvine and the University of Florida for their comments and feedback. We thank Rhong Zheng for assistance in data collection, and Ashley Tyrrel and Leslie Culpepper for help in data analysis. This work was partially supported by a Microsoft Live Labs Search Award, a Microsoft Virtual Earth Award, NYU Research Challenge Fund N-6011, and by NSF grants IIS-0643847 and IIS-0643846. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the Microsoft Corporation or of the National Science Foundation.

References

Ba, Sulin. 2001. Establishing online trust through a community responsibility system. *Decision Support Systems* **31**(3) 323–336. 6

Ba, Sulin, Paul A. Pavlou. 2002. Evidence of the effect of trust building technology in electronic markets: Price premium and buyer behavior. *MIS Quarterly* **26**(3) 243–268. 2, 6, 9

Blei, David M., Andrew Y. Ng, Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* **3** 993–1022. 21

Cabral, Luis, Ali Hortaçsu. 2005. The dynamics of seller reputation: Theory and evidence from eBay. Working Paper EC-04-05, Stern School of Business, New York University. 2, 5

Chevalier, Judith A., Dina Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* **43**(3) 345–354. 7

Das, Sanjiv Ranjan, Mike Chen. 2006. Yahoo! for Amazon: Sentiment extraction from small talk on the web. Working Paper, Santa Clara University. Available at http://scumis.scu.edu/~srdas/chat.pdf. 6

Dawar, Niraj, Philip Parker. 1994. Marketing universals: The use of brand name, price, physical appearance, and retailers' reputation as signals of product quality. *Journal of Marketing* **58**(2) 81–95. 6

Dellarocas, Chrysanthos. 2003. The digitization of word-of-mouth: Promise and challenges of online reputation mechanisms. *Management Science* **49**(10) 1407–1424. 2

Dellarocas, Chrysanthos, Ming N. M. I. Fan, Charles A. Wood. 2004. Self-interest, reciprocity, and participation in online reputation systems. Working paper. 29

Dellarocas, Chrysanthos N., Charles A. Wood. 2006. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. Working Paper, Robert H. Smith School Research Paper No. RHS 06-041. Available at http://ssrn.com/abstract=923823. 10

Dewan, Sanjeev, Vernon Hsu. 2004. Adverse selection in electronic markets: Evidence from online stamp auctions. *Journal of Industrial Economics* **52** 497–516. 5

Eaton, David. 2002. Valuing information: Evidence from guitar auctions on eBay. Working paper #0201, Department of Economics, Murray State University. 5

Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. MIT Press. 7

Gefen, David, Elena Karahanna, Detmar W. Straub. 2003. Trust and TAM in online shopping: An integrated model. *MIS Quarterly* **27**(1) 51–90. 9

Ghose, Anindya, Michael D. Smith, Rahul Telang. 2006. Internet exchanges for used books: An empirical analysis for product cannibalization and social welfare. *Information Systems Research* **17**(1) 3–19. 5, 11

Godes, David, Dina Mayzlin. 2004. Using online conversations to study word-of-mouth communication. *Marketing Science* **23**(4) 545–560. 7

Gu, Bin, Prabhudev Konana, Balaji Rajagopalan, Hsuan-Wei Michelle Chen. 2007. Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research* **18**(1) 68–85. 6

Hatzivassiloglou, Vasileios, Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL'97)*. 174–181. 7

Hofmann, Thomas. 1999. Probabilistic latent semantic indexing. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'99*. 50–57. 21

Hu, Minqing, Bing Liu. 2004. Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*. 168–177. 7

Kahneman, Daniel, Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263–292. 10

Kalyanam, Kirthi, Shelby McIntyre. 2001. Return on reputation in online auction market. Working Paper 02/03-10-WP, Leavey School of Business, Santa Clara University. 5, 9

Kamps, Jaap, Maarten Marx. 2002. Words with attitude. *Proceedings of the First International Conference on Global WordNet*. 332–341. 7

Kim, Dongmin, Izak Benbasat. 2003. Trust-related arguments in internet stores: A framework for evaluation. *Journal of Electronic Commerce Research* **4**(2) 49–64. 9

Klein, Benjamin, Keith Leffler. 1981. The role of market forces in assuring contractual performance. *The Journal of Political Economy* **89**(4) 615–641. 8

Klein, Daniel B. 2000. *Assurance and trust in a great society (FEE occasional paper)*. 1st ed. Foundation for Economic Education, inc. 9

Kleinberg, Jon Michael. 1999. Authoritative sources in a hyperlinked environment. *Journal of ACM* **46**(5) 604–632. 33

Lee, Daniel D., H. Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755) 788–791. 21

Lee, Thomas. 2004. Use-centric mining of customer reviews. *Workshop on Information Technology and Systems*. 7

Lewitt, Steven, Chad Syverson. 2005. Market distortions when agents are better informed: The value of information in real estate transactions. Working Paper, University of Chicago. 6

Livingston, Jeffrey Alan. 2002. How valuable is a good reputation? A sample selection model of internet auctions. Working Paper, University of Maryland. 5

Lucking-Reiley, David, Doug Bryan, Naghi Prasad, Daniel Reeves. 2000. Pennies from eBay: The determinants of price in online auctions. Working Paper, Vanderbilt. 5

McDonald, Cynthia Goodwin, Vester Carlos Slawson. 2002. Reputation in an internet auction market. *Economic Inquiry* **40**(3) 633–650. 5

Melnik, Mikhail Ion, James Alm. 2002. Does a seller's reputation matter? Evidence from eBay auctions. *Journal of Industrial Economics* **50**(3) 337–350. 5

Montgomery, Alan, Kartik Hosanagar, Ramayya Krishnan, Karen Clay. 2004. Designing a better shopbot. *Management Science* **50**(2) 189–206. 33

Pavlou, Paul A., Angelika Dimoka. 2006. The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research* **17**(4) 392–414. 6, 9

Pavlou, Paul A., David Gefen. 2005. Psychological contract violation in online marketplaces: Antecedents, consequences, and moderating role. *Information Systems Research* **16**(4). 6, 9

Quinlan, John Ross. 1992. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, Inc. 26

Resnick, Paul, Ko Kuwabara, Richard Zeckhauser, Eric Friedman. 2000. Reputation systems. *Communications of ACM* **43**(12) 45–48. 2, 6

Resnick, Paul, Richard Zeckhauser, John Swanson, Kate Lockwood. 2006. The value of reputation on eBay: A controlled experiment. *Experimental Economics* **9**(2) 79–101. 2, 5

Senecal, Sylvain, Jacques Nantel. 2004. The influence of online product recommendations on consumers' online choices. *Journal of Retailing* **80**(2) 159–169. 7

Turney, Peter D. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*. 417–424. 7, 8

Turney, Peter D., Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems* **21**(4) 315–346. 7, 8

Zeckhauser, Richard, Paul Resnick. 2002. Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. Michael Roy Baye, ed., *The Economics of the Internet and E-Commerce*. Elsevier Science, 127–157. 29