Shipping News:
An Analysis of Internet Retailers' Shipping Strategies*

Emin M. Dinlersoz† — University of Houston

Han Li‡ — University of Houston

March 2004

Abstract

Managing shipment of goods to consumers is one of the central aspects of retail competition on the Internet. In this article, we use data from the Internet book retailing industry to analyze firms' choices of shipping options. We discuss and test the implications of alternative theories on the relation between a firm's product price, the quality of its shipping, and its shipping fee. We find that firms with lower product prices offer higher shipping qualities and lower shipping fees than firms with higher product prices. Compared with perfect information models with product differentiation, models based on imperfect consumer information explain better the observed patterns.

*We thank Aimee Chin, Adriana Kugler, Pedro Pereira, Roy Ruffin, Rebecca Thornton, and many others for comments, and especially Roger Sherman, for several useful conversations and suggestions.

†Corresponding author. Department of Economics, 204 McElhinney Hall, Houston, TX 77204-5019. E-mail: edinlers@mail.uh.edu

‡Department of Economics, 204 McElhinney Hall, Houston, TX 77204-5019. E-mail: han.li@mail.uh.edu

1 Introduction

Internet retailing is virtually impossible without shipping, for most Internet retailers have no physical presence where consumers can pick up orders. This reliance on shipping makes the delivery of goods to consumers a potential strategic component of competition. What distinguishes shipping from other firm characteristics such as reputation, brand name, and reliability, is that a firm's shipping service is often explicitly priced in the form of a separate shipping fee, rather than being implicit in the total price. There is significant variation in the quality of shipping options and associated fees among Internet retailers, in addition to the variation in basic price of the item sold, which we refer to as the base price. Consumers are sensitive to both the base price and the shipping fee when deciding which Internet retailer to buy from. Shipping and handling charges, like sales taxes, are important factors in whether customers view online prices as lower than prices in traditional stores. In this paper, we analyze the nature of firms' shipping strategies using data from Internet book retailing and assess the relevance of theories that offer alternative forms for the relationship between firms' choices of their base prices and their shipping options.

In guiding our empirical analysis, we focus on the implications of two broad strands of theory. The first is the theory of vertical product differentiation under perfect consumer information about firms' prices and other attributes. The shipping policies of Internet retailers can create or enhance differentiation for an otherwise homogenous product. Firms can choose different levels of quality, as measured by average delivery time, and mitigate competition to some extent by serving consumer segments that have different sensitivities for delay. Basic models of vertical product differentiation, e.g. Shaked and Sutton (1982), Choi and Shin (1992), Motta (1993), Lehmann-Grube (1997), when applied to the context of shipping, have three main predictions: *i*) controlling for other firm characteristics, firms with higher quality offer higher total prices, *ii*) controlling for shipping fee, firms with higher quality charge higher base prices, and *iii*) controlling for quality, base price and shipping fee are negatively correlated. In a perfect consumer information setting, these predictions simply state that consumers pay for a quality premium and, keeping quality constant, firms with higher base price must compensate their customers by offering a lower shipping fee, or vice versa, so that each consumer is indifferent between firms. Since consumers only care about total price, the partition of total price into a base price and a shipping fee is arbitrary. As a result, firms have several choices: they can fully absorb the cost of shipping, share part of the cost with consumers, shift all the cost onto consumers, or make profit on shipping by charging a fee above the cost of shipping. Empirical as well as anecdotal evidence suggests that some firms do indeed engage in a variety of practices.[1] The main question is

[1] For instance, Blackwell's, one of the retailers in our sample, always offers free shipping. Tedeschi (2001) reports that CDNow made profit on shipping in the early days of e-commerce by charging customers $3 for the first item and $1 for each additional item, a price that was then far above the average cost of shipping a CD. The marketing manager reported that shipping fees brought roughly the same profit margin (15-20%) as each compact disk sale. Campanelli (2002) notes

what makes each firm adopt one of these various practices if all are equivalent in terms of profit.

One answer comes from models of imperfect consumer information and search. One main reason for the non-identification of base price and shipping fee in the standard vertical product differentiation models is that consumers have perfect information of base prices and shipping fees, and they simultaneously consider base price and shipping fee before purchase decision. If consumers have imperfect information about base prices and shipping fees, and if they learn about shipping options only after spending additional effort and time at a firm's website, separate identification of base price and shipping fee is possible. Such a framework is plausible for Internet retail markets. Some sophisticated web navigators who have low search cost search many websites before they purchase. Some others are less sophisticated high search cost consumers, who randomly select a seller and do not spend much time in trying to find a bargain. Consumers typically do not learn about a seller's shipping options until they complete their purchases because sellers do not usually disclose their shipping options up front, often not until customers finalize their orders. This may be an effort to lock in consumers who have relatively high search costs. We demonstrate, in an extended version of Salop and Stiglitz (1977), that when consumers have different information gathering costs, base price and shipping fee can be positively correlated after controlling for other differences across firms, a prediction that is the opposite to that predicted by vertical product differentiation models. The presence of informed consumers with low search costs constrains the firms targeting these consumers to charge low base prices and shipping fees, whereas those firms targeting uninformed consumers who have high search costs can charge high base prices and shipping fees. This creates dispersion in both base prices and shipping fees across firms, and results in a positive correlation between base price and shipping fee. We show that this conclusion can hold even when shipping quality choice by firms is considered.

We test the predictions of the theories using data from Internet book retailing. We first estimate the shipping fee-quality schedule for retailers in our sample and find that there are systematic differences in these schedules across retailers. Discount book retailers and well-known retailers such as Amazon.com and Barnes and Noble tend to offer lower fees than average, holding quality constant. Less-well known, smaller, and more specialized retailers charge much higher fees than average. We also find that, controlling for a variety of observables, low base-price sellers offer higher shipping quality, i.e. lower average delivery time. At the same time, sellers with lower base prices tend to charge lower shipping fees, controlling for shipping quality and other seller characteristics. Our controls suggest that the results are unlikely to be driven by scale economies that may result in lower prices and lower shipping costs for larger retailers, or by specialization of retailers on certain book categories

that an on-line Christmas tree retailer charged a below cost shipping fee with the hope of attracting consumers. Some firms are aware of the fact that savvy consumers can find out whether a firm is charging a consistently higher than expected shipping fee for a given quality, and may abandon that firm. These concerns are important for firms trying to build trust and loyalty in a dynamic setting. Campanelli (2002) reports that some retailers play safe and adopt the practice of cost shifting, i.e. charging their customers exactly their cost of shipping.

that may allow them to charge higher base prices and shipping fees for their strongly differentiated products. The observed patterns favor imperfect information-based models rather than the vertical product differentiation models with perfect information. Our findings provide additional evidence that informational asymmetries across consumers may play a role in how firms price the services they attach to their products. The price dispersion in the Internet book retailing is closely related to the dispersion in quality and pricing of shipping services, and all of these dispersions may be driven by the fact that consumer information gathering is a prominent consideration in these markets, as emphasized by recent research (e.g. Baye and Morgan (2001), Baye, Morgan and Scholten (2002), Ellison and Ellison (2001)).

Several recent papers have investigated the role of shipping indirectly when analyzing price dispersion in Internet markets. In an analysis of the role of brand awareness in consumer decision making in Internet markets, Smith and Brynjolfsson (2001) find that consumers are very sensitive to shipping fees and delivery times. Their findings point to the importance of shipping strategies in influencing demand, but they do not focus on the supply side, i.e. on how shipping options are chosen by retailers. In a reduced-form empirical investigation of the drivers of price dispersion, Pan, Ratchford and Shankar (2003) find that the shipping options adopted by retailers contribute significantly to the variation in firms' prices. While this finding hints at the importance of shipping policies in generating variations in price, the direction of the relationship and the channel through which the relationship emerges, are left unexplored. Our approach differs in that we single out the role of shipping and test models that lead to explicit, and different, predictions regarding the relationship between price, quality, and fee. Our findings also add to the literature on the role of consumer information in generating dispersion in prices and product quality in retail markets, traditional as well as online. In an experimental setup involving online wine sales to agents, Lynch and Ariely (2000) conclude that consumer search costs affect firms' choices of quality as well as prices for this differentiated good. Here, we use actual data from a homogenous product industry and find support for the theory that search costs and imperfect information can affect the pricing of firm attributes and services, contributing to the price dispersion of the total price.

The road map for the rest of the paper is as follows. In Section 2, we present alternative theories and contrast their testable implications, especially on the relationship between base prices and fees. Section 3 contains our empirical analysis based on the models presented. We first analyze the structure of the fee-quality schedule for retailers. We then test the predictions of the models regarding the base price-shipping fee relationship, followed by a discussion of the results and alternative explanations. Section 4 concludes.

2 Theoretical motivation

In this section we discuss models that have different implications on the relationship between base price, shipping fee, and shipping quality across firms. We begin in a perfect consumer information setting where firms choose their shipping options and the corresponding fees to differentiate themselves from their rivals in the shipping quality dimension. We then consider a model of imperfect consumer information that leads to a different set of implications. Finally, we briefly discuss issues that we do not explicitly model, such as free shipping and dynamic considerations.

2.1 Vertical product differentiation under perfect information

Consider a general vertical product differentiation model (e.g. Shaked and Sutton (1982), Tirole (1988), Choi and Shin (1992), Motta (1993), Lehmann-Grube (1997)). Two retail firms offer a single homogenous product for sale and they engage in a two-stage competition. In the first stage, they choose their shipping qualities and shipping fees, followed by choice of prices in the second stage. The product thus becomes differentiated once firms choose shipping quality. The two-stage nature of competition is reasonable. A firm's shipping options typically apply uniformly to all products of the firm within a given product category. For instance, in the Internet book retailing we consider, a firm's shipping options generally apply uniformly to all the books it sells. Furthermore, shipping fees tend to vary much less frequently over time compared to prices, and can be taken as given when a firm a firm is updating its prices.

In the first stage, each firm chooses a shipping quality, q, and an associated fee, f. The quality q specifies a range of delivery times. The average delivery time is strictly decreasing as quality increases. Shipping a unit of a product at quality q costs $c(q)$, where $c(\cdot)$ is increasing and strictly convex, representing a technology that is common to all firms. This last assumption is for simplicity and it rules out the possibility that a firm can engage in a special contract with a shipping company based on scale or other firm characteristics.[2] In this case, not all shipping technologies would be available to all firms, a possibility we allow in our empirical analysis.

In the second stage of the competition, each firm chooses a price, p, given the first stage choices of shipping options. There is a unit measure of consumers. Consumers can purchase at most one unit of the product.[3] Each consumer obtains a gross surplus of s from consuming the good itself, but they have varying tastes for shipping quality. A consumer's marginal utility from quality is given by α, distributed uniformly in the interval $[0, 1]$. A type α consumer obtains a utility of

$$u(p, f, q; \alpha) = s + \alpha q - p - f$$

[2]It is straightforward to allow for different shipping technologies across firms by assuming that $c_i(q) \equiv c(q, X_i)$, where X_i is a set of firm characteristics that affect the cost of shipping.

[3]We discuss later the case of multiple unit purchases.

from consuming the product at price p and having it shipped with quality q at fee f. If the consumer does not purchase, the utility is zero. To ensure market coverage, s is sufficiently large that even the consumer with $\alpha = 0$ obtains positive utility from a purchase.[4]

Firms are heterogenous with respect to their marginal cost, c. Without loss of generality, firm 1 has a lower marginal cost, i.e. $c_1 < c_2$. We look for a subgame perfect Nash equilibrium for the choice of shipping options and prices. Given the shipping options chosen in the first stage, the consumer type, α^*, who is indifferent between purchasing from the two firms, is given by $\alpha^* = \dfrac{\Delta p + \Delta f}{\Delta q}$, where $\Delta x = x_1 - x_2$ for $x = p, f, q$. All consumers with types in $[\alpha^*, 1]$ buy from firm 1, and the rest from firm 2. The profit maximization problem for firm 1 in the second stage is then given by

$$\max_{p_1}(p_1 - c_1 + f_1 - c(q_1))\left(1 - \frac{\Delta p + \Delta f}{\Delta q}\right),$$

which generates the first order condition for p_1

$$\left(1 - \frac{\Delta p + \Delta f}{\Delta q}\right) - \frac{1}{\Delta q}(p_1 - c_1 + f_1 - c(q_1)) = 0. \tag{1}$$

Similarly, for firm 2 we have

$$\left(\frac{\Delta p + \Delta f}{\Delta q}\right) - \frac{1}{\Delta q}(p_2 - c_2 + f_2 - c(q_2)) = 0. \tag{2}$$

From (1) and (2), we obtain

$$\begin{aligned} p_1 &= \frac{2}{3}\Delta q + \frac{2}{3}c_1 + \frac{2}{3}c(q_1) + \frac{1}{3}c_2 + \frac{1}{3}c(q_2) - f_1, \qquad (3)\\ p_2 &= \frac{1}{3}\Delta q + \frac{1}{3}c_1 + \frac{1}{3}c(q_1) + \frac{2}{3}c_2 + \frac{2}{3}c(q_2) - f_2,\\ \Delta p &= \frac{1}{3}\left(\Delta q + \Delta c + \Delta c(q)\right) - \Delta f, \end{aligned}$$

where $\Delta c(q) = c(q_1) - c(q_2)$. From (3), the total price, P_i, a consumer pays if he buys from firm i is

$$P_i = p_i + f_i = (1 - \frac{i}{3})\Delta q + \frac{2}{3}(c_i + c(q_i)) + \frac{1}{3}(c_j + c(q_j)), \tag{4}$$

for $j \neq i$. It is clear from (4) that how P_i is partitioned into p_i and f_i does not matter, because a consumer is indifferent between any change in either of these variables as long as that change is offset by an equal but opposite change in the other.

Now consider the first stage, where firms choose their qualities and associated fees. Firm 1 solves

$$\max_{(q_1, f_1)}(p_1 - c_1 + f_1 - c(q_1))\left(1 - \frac{\Delta p + \Delta f}{\Delta q}\right),$$

[4]This simplifies analysis, but otherwise has little impact on our qualitative results.

which, after substituting for p_1 and Δp using (3), is equivalent to

$$\max_{q_1} \left(\frac{[2\Delta q - \Delta c - \Delta c(q)]^2}{\Delta q} \right). \tag{5}$$

Note that shipping fee f_1 does not appear in the objective function, making it impossible to identify f_1, or equivalently, p_1. A similar conclusion applies to the profit maximization problem of firm 2. As a result, a firm has several options in choosing shipping fee: it can entirely absorb the shipping cost, share it with the consumer, make the consumer pay for it in full, or make profit on shipping. We call these choices cost absorption or free shipping ($f_i = 0$), cost sharing ($f_i \in (0, c(q_i))$, cost shifting ($f_i = c(q_i)$), or profit from shipping ($f_i > c(q_i)$). In general, we can write

$$f_i = \theta_i c(q_i), \tag{6}$$

where $\theta_i \geq 0$ is a firm specific constant. The fee choices available to firms then correspond to cases $\theta_i = 0$, $\theta_i \in (0, 1)$, $\theta_i = 1$, and $\theta_i > 1$, respectively. The model does not identify θ_i. To complete the solution of the model, we need to determine the choice of quality levels by each firm. This problem is analyzed by Motta (1993) for firms with equal marginal costs and the results apply here with straightforward modification, which we omit to save space. In general, either firm can become the high quality firm, depending on the parameters.

The model has three main testable implications. First, controlling for seller cost structure, the high quality firm commands a higher total price. This follows because $P_1 - P_2 = \Delta P > 0$, if the marginal costs are equal, i.e. $\Delta c = 0$. Second, controlling for shipping fee and sellers' cost structure, the high quality firm commands a higher base price. This follows because $\Delta p > 0$, if the marginal costs and the shipping fees are equal, i.e. $\Delta c = 0$ and $\Delta f = 0$. Higher quality leads to a higher price as a result of two factors: the extra cost of higher quality, $\Delta c(q)$, and the quality premium itself, Δq.[5] Finally, controlling for quality and seller cost structure, the firm with a higher shipping fee must charge a lower base price. This follows because Δp is negatively related to Δf when $\Delta q = 0$. From the second implication, it is essential to control for shipping fee in analyzing the price-quality relationship. The model implies that the shipping fee is taken into consideration by firms when base price choices are considered, even though the model does not posit any reason for choosing a particular fee given a quality level. Unless firms adopt a common markup θ over the cost of shipping quality, not including it as an explanatory variable results in a bias.[6]

While the model emphasizes the use of shipping options for product differentiation, the scope for doing so can be limited in practice. If there are only a finite number of shipping options available to

[5] Unfortunately, these two effects cannot be disentangled empirically, unless one has a good measure of the underlying cost of quality.

[6] Several recent investigations explaining base price variation across firms consider either a measure of shipping quality (e.g. Xing, Ratchford, and Shankar (2002)) or the shipping fee (e.g. Baylis and Perloff (2002)), but not both simultaneously.

firms and if each firm has equal access to all options, with an adequate number of firms in the market, some firms will be restricted to choosing the same options. The standard Bertrand result then implies that for firms choosing the same option, the quality premium, Δq, should vanish. However, since the firms bunching at higher quality have higher costs than firms with lower quality, i.e. $\Delta c(q)$ is positive, they should still charge a higher total price, and a higher base price controlling for shipping fee. Baye and Morgan (2003) find that the price premium can indeed vanish in markets where firms can differentiate themselves only along a single dimension that can easily be imitated by rivals.[7] In the case of shipping options, this may not apply exactly. As discussed earlier, not all options may be available to all firms, because some firms may engage in contracts with shipping companies based on their scale or other distinguishing characteristics. This can prevent or limit imitation. The data used here reveal that there is indeed substantial differentiation in shipping quality.

The problem of identification between base price and shipping fee arises for two main reasons. First, the shipping fee and the base price enter the utility function additively and symmetrically, implying that consumers focus only on total price and not on its components. However, consumers may perceive and evaluate base price and shipping fee differently. Smith and Brynjolfsson (2001) find that consumers are roughly twice as sensitive to shipping fee as to base price when making purchase decisions on the Internet.[8] One reason for such discrepancy is pyschological: consumers may perceive high shipping fees as unfair, even though the total price is the same (see, e.g., Thaler (1985)). This may be one explanation for why some retailers offer permanent free shipping, where the cost of shipping is presumably entirely reflected in the base price. There is also some recent experimental evidence that partitioned prices may alter demand. Morwitz, Greenleaf, and Johnson (1998) find that demand increases when price is presented to consumers in a partitioned form rather than as a total.[9] There is little guidance, however, on why one firm may choose a particular partition over others.

The second reason for non-identification is that consumers have perfect information of the base prices and shipping fees in the market, and they consider them simultaneously before making a purchase. If consumers differ in their information about prices and fees, and if they discover these two sequentially, then identification may be possible. Sequential discovery is plausible in on-line markets where consumers visit a firm's website based on base price and learn about the shipping fee only after completing certain steps of shopping. We examine this possibility next.

[7] They consider whether a firm is listed at CNET's certified merchant program. Being listed signals consumers reliability of the retailer, and differentiates it from others.

[8] The authors note the possibility that this may be an artifact of the econometric specification they use.

[9] The authors also present some experimental evidence that firms with higher reputation and brand name effects benefit more in terms of demand from partitioned prices compared to less well-known rivals.

2.2 Imperfect information with sequential price discovery

Recent research has emphasized the role of information and consumer search in generating price dispersion in Internet retail markets (e.g. Baye and Morgan (2001), Baye, Morgan and Scholten (2002), Ellison and Ellison (2001)). In many standard models of consumer search, such as Salop and Stiglitz (1977), Reinganum (1979), Burdett and Judd (1983), as long as base price and shipping fee are discovered simultaneously, the identification problem between base price and shipping fee remains, because a consumer cares only about the total price in deciding whether to search further. One way to overcome this problem is to introduce sequential and costly discovery of base prices and shipping fees. We formalize this approach by extending the model of Salop and Stiglitz (1977) to incorporate quality choice and sequential price discovery.

As in the previous model, there is a unit measure of consumers with unit demand and each consumer obtains a gross surplus of s from consuming a homogenous good. Consumers also value shipping service. There are only two shipping options available: a low and a high quality option, with respective quality levels $q^L < q^H$. Higher quality brings higher gross surplus to the consumer, $u^L < u^H$.

There are n firms in the market, and for each, the total cost not including shipping is $c(x) = F + cx^2$ from selling $x \geq 0$ units, where F is the fixed cost and $c > 0$.[10] Marginal cost of shipping is constant and equal to c^L and c^H, respectively, for low and high quality shipping, where $c^L < c^H$. Assume also that $u^H - c^H > u^L - c^L$, i.e. high quality brings a higher welfare per unit of good shipped.[11]

A fraction, ϕ, of the consumers are informed and know all the base prices in the market. These consumers are savvy web navigators and enthusiastic bargain hunters who already have access to shopbots, search engines and other sources that provide them with a list of base prices. The remaining fraction, $1 - \phi$, are uninformed and do not know individual firms' base prices. These consumers have a high cost, $\sigma > 0$, of gathering price information, due perhaps to a high opportunity cost of time, low skills in web navigation, or simply ignorance. Since this cost can be the cost of time and effort invested in obtaining price information from a variety of sources, an uninformed consumer can choose to become informed by incurring the cost σ.

No consumer, informed or uninformed, initially possesses perfect information on shipping qualities and shipping fees offered by firms.[12] A consumer discovers these variables only after spending some time at the firm's website. Even for informed consumers who know all the base prices, consumer information about the attributes of firms is incomplete. Several studies find that navigating websites and learning about firm attributes takes time so consumers cannot easily rank all firms in the market,

[10] For search to be meaningful, we consider $n > 2$ firms. With only two firms, consumer search can be trivial.

[11] We discuss alternatives to this assumption later.

[12] We discuss later the case when both the base price and shipping fee are listed simultaneously in a shopbot search. Some search engines provide this service, some don't.

even though they may have relatively easier access to price information.[13] Firms also do not typically disclose their shipping policies on the first page of their websites, but rather inform the consumer much later, usually just prior to purchase. By engaging in such practices, firms try to lock in consumers to some extent before they face the final decision. For uninformed consumers, accessing the desired information is burdensome and costs $\omega > 0$ for each firm visited. For informed ones, the cost of discovering a firm's shipping policy after visiting a firm is zero, because they do not value time as much and/or because they are skillful web-surfers and can access to such information more swiftly.

An informed consumer always visits the seller with the minimum base price, $p_{(1)}$, first. The consumer purchases from this firm if

$$u_{(1)} - p_{(1)} - f_{(1)} \geq E[u] - p_{(2)} - E[f], \tag{7}$$

where the left hand side is the expected utility offered by the firm with the lowest base price, and the right hand side is the expected utility offered by the firm with the second lowest base price, $p_{(2)}$. Otherwise, the consumer proceeds to visit costlessly the firm with the second lowest base price, and this process continues until the consumer settles with a firm.

An uninformed consumer first has to decide whether to randomly sample a seller i, or to become informed and visit the firm with the minimum base price, $p_{(1)}$. The consumer prefers visiting a firm randomly if

$$E[u] - E[p] - E[f] - \omega \geq E[u] - p_{(1)} - E[f] - (\sigma + \omega). \tag{8}$$

The left hand side of (8) is the expected utility from visiting a randomly selected firm. The right hand side is the expected utility from becoming informed and visiting the firm with the lowest base price. Condition (8) is equivalent to

$$E[p] \leq p_{(1)} + \sigma. \tag{9}$$

If condition (9) holds, the consumer visits a firm randomly and learns its price. The consumer proceeds to discover the firm's shipping quality and shipping fee if

$$E[u] - p_i - E[f] - \omega \geq E[u] - p_{(1)} - E[f] - (\sigma + \omega), \tag{10}$$

otherwise, the consumer becomes informed. Condition (10) is equivalent to

$$p_i \leq p_{(1)} + \sigma. \tag{11}$$

Finally, after discovering the shipping fee and quality, the uninformed consumer purchases if

$$u_i - p_i - f_i \geq E[u] - p_{(1)} - E[f] - (\sigma + \omega), \tag{12}$$

[13] See, e.g., Baylis and Perloff (2002), Dellaert and Kahn (1999), Mandel and Johnson (1998), Menon and Kahn (1998).

otherwise, the consumer becomes informed.

We look for a two-price equilibrium where: *i*) a fraction, μ, of firms charge a base price $\underline{p}$, and offer high quality, q^H, at a fee, f^H, and the remaining fraction $(1-\mu)$ charge a price $\overline{p} > \underline{p}$, and offer low quality, q^L, at a fee, $f^L > f^H$, *ii*) given prices, fees, and quality levels, informed consumers always purchase from low price firms and uninformed consumers purchase randomly, *iii*) firms maximize their profits, and *iv*) the number of firms serving the market, n, is optimally determined under free entry, i.e. each firm makes zero profit.[14]

In Appendix A we show that such a two-price equilibrium exists and we characterize it. In equilibrium, the low price, $\underline{p}$ is equal to zero, and the fee f^H for high quality shipping is equal to p^*+c^H, where p^* is the competitive base price equal to the minimum average cost. The competition for informed consumers, and the fact that informed consumers first visit the firm with the lowest base price, drive the base prices of low price firms to zero, and their fee to the sum of the competitive price and competitive fee. On the other hand, for firms with high base price we have $\overline{p} = \underline{p}+\sigma$, and $f^L = f^H - \Delta u + \frac{\omega}{\mu}$, where $\Delta u = u^H - u^L$ is the gap between the gross surpluses from high and low qualities. The cost of information gathering, σ, allows some firms to charge prices that are just high enough to keep uninformed consumers indifferent between becoming informed and buying randomly. Firms with price $\overline{p}$ sell only to uninformed consumers, and firms with price $\underline{p}$ sell to informed consumers and lucky uninformed ones who happen to visit them.

In general, the fee charged by high price firms, f^L, can be higher or lower than the fee charged by low price firms, f^H. The part of the difference $f^L - f^H$ accounted for by quality difference across low and high price firms is $-\Delta u < 0$, and the remaining part, $\frac{\omega}{\mu} > 0$, reflects the cost associated with discovery efforts. These effects work in opposite directions. High price firms compensate their customers for the lower quality they offer, but also charge a premium to take advantage of the fact that uninformed consumers incur discovery costs. If $\frac{\omega}{\mu} > \Delta u$, then we have $f^L > f^H$. Thus, f^L is higher than f^H when discovery cost, ω, is relatively high, there is a small fraction, μ, of low price firms, and the surplus gap, Δu, is not very high. It is important to observe that, controlling for quality difference between the two types of firms, we always have $f^L > f^H$, because once the part of f^L accounted by the quality difference is accounted for, the remaining part, $\frac{\omega}{\mu}$, is always positive.

The total price offered by firms with high quality, $\underline{P} = p^* + c^H$, is lower than that of the firms with low quality, $\overline{P} = p^* + c^H + \sigma + \frac{\omega}{\mu} - \Delta u$, if $\sigma + \frac{\omega}{\mu} > \Delta u$, i.e. the information gathering and discovery costs are high enough to offset the compensation for quality gap. Otherwise, higher quality actually commands a lower total price. Therefore, the relationship between total price and quality can go either way in equilibrium.

[14]This is not necessarily the only equilibrium. As detailed in Salop and Stiglitz (1977), depending on the parameters, a single price equilibrium can obtain, or an equilibrium fails to exist. There may also be other two-price equilibria with different quality choice configurations. Our goal is to illustrate one equilibrium that lines up with the empirical evidence to follow.

Appendix A discusses what happens if σ or ω, or both, are large. In these cases, the positive relationship between base price and shipping fee hold even more strongly, controlling for quality. It is also easy to verify that, if $\sigma = 0$, then the two-price equilibrium breaks down. If $\omega = 0$, the two-price equilibrium prevails, but the positive relationship between base price and fee disappears, as expected.

To sum up, in this equilibrium base price and shipping fee are positively associated, controlling for shipping quality. This is in contrast to the perfect information model presented earlier, which predicts a negative relationship. Furthermore, in the imperfect information model base price and shipping quality are negatively related, a prediction opposite to that of the perfect information model. Finally, the total price can be positively or negatively associated with shipping quality. If information gathering and discovery costs are relatively high, then a negative association is expected. This again contrasts with the implication of the perfect information model that total price should always increase with shipping quality.

The assumption that consumers do not discover base price and shipping fee simultaneously deserves further comment. If becoming informed is equivalent to learning simultaneously all the base prices, the shipping qualities, and the shipping fees, say, by using a search engine, then it is not possible to identify the base price and shipping fee. In this case, $\underline{p}$ and f^H can each take any value in the range $[0, p^* + c^H]$, subject to the constraint $\underline{P} = \underline{p} + f^H = p^* + c^H$. Therefore, the base price and the fee are then negatively associated for firms with low base price. However, as long as information gathering and discovery costs, σ and ω, are large enough, we show in Appendix A that it is possible to identify the base price, $\overline{p}$, and the shipping fee, f^L, separately. We also show that, controlling for quality difference, $\overline{p}$ is higher than any base price $\underline{p} \in [0, p^* + c^H]$ charged by firms targeting informed consumers, and f^L is higher than any shipping fee $f^H \in [0, p^* + c^H]$ charged by firms targeting informed consumers, as long as information gathering and discovery costs, σ and ω, are large enough. Thus, our main results can still hold in the case where informed consumers know all the shipping qualities and the fees, in addition to the base prices.

Finally, we note that depending on the parameter values, other equilibria may exist in our framework, an issue we do not expand on for brevity. We emphasize that the model presented is not the only way to obtain the empirical implications mentioned. Similar models can be set up based on imperfect information and search. The quality dimension can also be modelled in alternative ways. For example, consumers with high search costs can also be the ones that value higher shipping quality, because they have high opportunity cost of time and need faster shipping. This may result in firms with high base price offering higher quality, as opposed to lower quality in the current model. The implications on the base price-shipping fee relationship is not affected under alternative scenarios for quality, as long as the quality difference between firms is controlled for.

2.3 Multi-unit purchases, free shipping, and dynamic competition

To simplify analysis, we have so far assumed that consumers can purchase at most one unit. In reality, firms have much more complicated shipping strategies when multi-unit purchases are considered. For instance, some firms use a two-part tariff for shipping. Such affine pricing is familiar. Others offer additional discounts for the third unit and beyond in a highly non-linear scheme. Some firms also practice free shipping beyond a threshold purchase amount, defined in dollars or in units. This threshold varies across firms, and free shipping sometimes only applies to selected items. In short, firms differ considerably in pricing of additional units beyond the first. Competition becomes complicated when firms can choose non-linear fees for shipping, and a full theoretical treatment of shipping competition with several non-linear shipping fees is beyond the scope of this paper. However, some theoretical arguments can still be made about free shipping.

A common strategy used by retailers is the practice of free shipping beyond a certain dollar amount of purchase. In general, a firm can benefit from offering free shipping because it induces more sales. However, firms that offer lower base prices are expected to benefit more from doing so compared to the firms that cannot, because a firm with a lower price in general sells more units, implying that the firm experiences scale economies in shipping and, hence, faces lower average shipping costs. The firm can then afford more easily to offer free shipping beyond a certain size of purchase compared to high price firms, and can attract even more consumers and generate more sales per consumer by doing so. Thus, we expect a low price and a low free shipping fee threshold to reinforce each other. There is also some evidence that firms use free shipping much like a penetration pricing strategy upon entry. For instance, in the early days of the diffusion of e-commerce, most entrants offered free shipping at the expense of major short run losses, that led to their demise.[15] By offering free shipping, a firm can attract consumers, even though it might make a loss in the short run. From a dynamic perspective this loss can be justified if firms can create loyalty and repeat purchases for the long run.

Overall, we expect a negative association between a firm's likelihood of offering a free shipping threshold and its price, and a positive association between a firm's free shipping threshold and its price. These predictions are also consistent with the imperfect information model in the previous section, where lower shipping fees are associated with lower base prices, controlling for quality.

3 Empirical analysis

We now proceed to examine the models' implications using data from Internet book retailing. We first analyze the shipping fee-quality schedules offered by retailers in this industry so as to have a better understanding of what determines shipping fees, and to find out to what extent the schedules differ across firms. Then we test the predictions of the models regarding the relationship between base

[15] See Tedeschi (2001).

price, shipping quality and shipping fee.

We focus on Internet book retailing because it is especially suitable for the kind of analysis we undertake. First, it is one of the most studied industries in the empirical e-commerce literature. It is highly concentrated (Brynjolfsson and Smith (2000), Smith and Brynjolfsson (2000), Latcovich and Smith (2001), Chevalier and Goolsbee (2003)), so strategic effects are likely to be important. It is a relatively more mature Internet industry, so our results are unlikely to be driven by transitory effects that may have been influential in the early days of e-commerce. There is also important price dispersion in the industry to be explained (Brynjolfsson and Smith (2000), Smith and Brynjolfsson (2001), Clay, Krishnan, and Wolf (2001)). Furthermore, Smith and Brynjolfsson (2001) document that search plays a role in this industry. Second, books are homogenous goods, which makes services attached to the product, such as shipping quality, all the more central in creating product differentiation. Third, books are mostly not very expensive items so shipping fee usually constitutes an important fraction of a book's price. Therefore, we expect consumer search to play a role in the determination of prices and fees in this market. Finally, and most importantly, shipping options and schedules offered by firms are relatively simple and well-defined in this industry. In general, there is a wide range of practices across industries that can complicate analysis. Shipping fees can be based on weight of the product sold, on the total value of the shipment, and sometimes on customer's geographic location as well. Our analysis is simplified because in book retailing the products come in well-defined units, shipping fees are mostly based on the number of units shipped, and are independent of consumers' geographic location for a majority of retailers. A brief description of the dataset used in this paper is given in Appendix B, together with the details of data collection. Table 1 describes the variables used in the analysis.

3.1 Analysis of shipping options

Table 2 lists shipping fees and delivery times for the standard shipping options of selected sellers and comparable shipping options offered by main specialized shipping companies. If sellers were offered these quoted rates by shipping companies, the shipping technology frontier for the industry could be easily determined. However, publicly posted rates of shipping companies have in general little to do with the actual rates offered by the book sellers. For instance, Amazon.com and Barnes and Noble report a fixed fee of $3.00 and a fee of $0.99 for each book, implying a shipping fee of $3.99 for a single book. This fee structure and associated delivery ranges are different from any of those offered by shipping companies. As discussed earlier, this pattern is potentially a result of special contracts between sellers and shipping companies. While such contracts are not observable, further evidence on their existence comes from the website of Laissez Faire Books:[16]

[16] Please visit the retailer's website, *http://www.lfb.com/shipping.asp*, to see this quote in the first page, and also to read more on its shipping strategy.

> "...In this business, it is hard to avoid supporting the [U.S.] Post Office, but when we can use a private carrier instead, we try to do so. We have recently negotiated rates with United Parcel Service that allow us to offer our customers much faster and more reliable delivery for only $2 per order..."

We expect such contracts to depend on firm characteristics, such as scale and scope, the structure and location of retailers' warehouse and distribution systems, as well as the quality of the shipping method(s) chosen. For instance, large firms may be able to negotiate better rates because of the scale economies they confer on shipping companies.

Table 3 provides summary statistics on the quality and fees across retailers for four different shipping options in our sample of retailers taken in January 2004.[17] We measure quality by the average delivery time in days, calculated from the reported minimum and maximum expected delivery times by the retailers. Option I is the slowest and option IV is the fastest in terms of average delivery time. Keep in mind that the quality of an option offered by one firm may correspond to the quality of a different option offered by another firm. Two things are notable in Table 3. First, not all firms offer all options, and they differ in the number of options they offer to their customers. Second, there is significant variation in quality and fee across firms for a given option, although the variation in quality diminishes as we move from Option I to Option IV. This is probably because not too many choices are available from shipping companies for highly expedited shipping. In fact, in all of the 87 options offered by 36 sellers in our sample, there are 20 distinct quality levels, implying that quality levels for one or more options coincide for some firms. A simple analysis of variance reveals that 45% of the variation in quality comes from *within* firm differences in quality of shipping options. The remaining 55% is due to the variation of quality *across* firms. Thus, there is significant variation in quality across firms. Similarly, shipping fee varies considerably across firms. About 35% of the variation fee is *within* firm variation, whereas 65% is attributable to variation *across* firms.

To understand the nature of the shipping fees, we now analyze the determinants of the shipping fee. If shipping quality is the main determinant of the shipping fee, there should be little systematic relation between firm characteristics and shipping fee. Differences across firms in shipping fee can then be mostly explained by different choices of quality. But if observable firm characteristics affect the contracts secured from specialized shipping companies, we expect significant variation in fee-quality schedule across firms to be explained in part by them. Towards a resolution, assume that the cost of a given quality of shipping is given by

$$c(q_i, \mathbf{x}_i) = g(\mathbf{x}_i)c(q_i) = g(\mathbf{x}_i)q_i^{\beta},$$

[17] As noted in Appendix B, we only considered retailers that offer well-defined shipping options based on the quantity of books purchased. Retailers that do not have reported shipping options, or options that depend on weight and/or distance are not included. These excluded retailers are very small retailers and we estimate that they constitute a very small fraction of Internet book sales.

where q_i is firm i's shipping quality, $\mathbf{x}_i$ is a vector of firm characteristics, and $\beta > 1$. This specification implies a convex cost of quality, as in the vertical product differentiation model, and a multiplicatively separable effect of other firm characteristics.

A typical firm offers several shipping options. In our data, options range from one to four, differentiated by quality. In accordance with the choice of shipping fee by a firm under the vertical product differentiation model summarized in equation (6), assume that the shipping fee charged by firm i for a particular shipping option j is given by $f_{ij} = \theta_{ij} c(q_{ij}, \mathbf{x}_i)$, where $\theta_{ij} \geq 0$. Assume also that θ_{ij} only depends on a set of variables, $\mathbf{y}_i$, that influence a firm's choice of fees, i.e. $\theta_{ij} = \theta_i = \theta(\mathbf{y}_i)$. Thus, a firm's 'markup' over shipping cost applies uniformly to each shipping option j. This assumption is restrictive, but simplifies estimation for a preliminary analysis.[18] Suppose that the observed shipping fees are given by

$$f_{ij}^O = \theta(\mathbf{y}_i) c(q_{ij}, \mathbf{x}_i) \varepsilon_{ij} = \theta(\mathbf{y}_i) g(\mathbf{x}_i) q_i^\beta \varepsilon_{ij},$$

where $\varepsilon_{ij} > 0$ is a log-normal error term that accounts for unobservables. Taking the logarithm of both sides, and assuming that $\theta(\mathbf{y}_i)$ and $g(\mathbf{x}_i)$ are multiplicatively separable in the individual variables included in $\mathbf{x}_i$ and $\mathbf{y}_i$, we can write

$$\ln f_{ij}^O = \alpha + \beta \ln q_{ij} + \boldsymbol{\delta} \ln \mathbf{y}_i + \boldsymbol{\gamma} \ln \mathbf{x}_i + \epsilon_{ij}, \tag{13}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are vectors of coefficients, and $\epsilon_{ij} = \ln \varepsilon_{ij}$ has normal distribution with zero mean.

The interpretation of the model in (13) is important. The estimate of β essentially tells us how the fee responds to a unit change in shipping quality both *within* and *across* firms. The estimates of the parameter vectors $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ describe how much and in what direction the postulated convex fee-quality schedule varies across firms with observable firm characteristics. As an alternative, we also use firm dummies to quantify the inter-firm variation in fee attributable to firm characteristics.

In practice, it is hard to specify separately what variables should be in $\mathbf{x}_i$ and $\mathbf{y}_i$. In theory, information on what shipping companies a firm contracts with and the nature of these contracts should be included in $\mathbf{y}_i$, while information on firm's characteristics that matter only for the markup for shipping should be in $\mathbf{x}_i$. However, the same variables that may affect a firm's shipping cost structure can also affect its choice of the markup θ_i.[19] Then, the coefficients of the common variables in $\mathbf{x}_i$ and $\mathbf{y}_i$ cannot be separately identified. Without an obvious way to classify variables into $\mathbf{x}_i$ and $\mathbf{y}_i$, and given the unavailability of data on identities of shipping companies a firm contracts with and the nature of these contracts, we will use a common set of variables for $\mathbf{x}_i$ and $\mathbf{y}_i$.[20]

[18] In fact, we do not have many observations for faster shipping options, so the separate estimation for θ_{ij} for each j would be imprecise.

[19] For example, a larger firm may be offered lower shipping rates because of scale economies it provides to the shipping company it contracts with. At the same time, a larger firm might choose to charge a lower markup for shipping.

[20] We leave the more ambitious attempt to estimate separately the markup function and the part of the cost function that depends on firm characteristics to future research. This can be accomplished with detailed data on the nature of

In implementing the regression (13), an issue is how to handle cases where the shipping fee is zero. Although this happens for only for very few gross outliers in our data, one cannot apply the logarithmic transformation when $f_{ij}^O = 0$. An observation of zero fee is not a consequence of an obvious censoring or truncation of a latent variable, either, which would make a Tobit approach appropriate. Since the fraction of observations with zero fee is very small in our data, we chose to maintain the regression framework in (13) with a small positive number added to the dependent variable when it is zero.[21] The ordering of the dependent variable remains unchanged in this case. As an alternative, we also use a more restrictive specification in levels

$$f_{ij}^O = \alpha + \beta_1 q_{ij} + \beta_2 q_{ij}^2 + \boldsymbol{\delta}\mathbf{y}_i + \boldsymbol{\gamma}\mathbf{x}_i + \epsilon_{ij}, \tag{14}$$

where the square of the quality is added to account for non-linearities in the relationship between the cost of shipping and quality. Note that this specification in levels does not allow firm characteristics to affect the slope of the fee-quality schedule for a firm, while the logarithm specification in (13) does. Interaction terms can be added to (14), but we chose to keep the number of regressors small because we do not have a very large dataset. Finally, we also estimate (13) and (14) after dropping the zero-fee observations.

Estimates from several alternative specifications of (13) and (14) using 37 firms and a total of 87 shipping options are presented in Table 4. The left panel contains the logarithm and level specifications including all observations. The right panel contains results excluding two outlier observations with zero shipping fee. Note that the coefficient of SHIPTIME, the average delivery time in days, is negative and highly significant in all log specifications, implying that lower average delivery time implies higher fees. In the level regressions, both the coefficient of SHIPTIME and the coefficient of its square are significant, implying a negative and convex relationship between fee and SHIPTIME. Fee is decreasing as average shipping time increases, but at a decreasing rate.

The coefficients for firm characteristics are not highly significant, suggesting that, given a quality level, the fee does not change in a highly systematic way with most of the firm characteristics included in the regressions. The only exceptions are the variables PHYSICAL and PUBLIC, which have relatively high significance in logarithm specification II when observations with zero fee are dropped. In particular, PUBLIC has a significant negative coefficient. PUBLIC is a dummy variable that takes on a value of 1 for relatively large sellers that have made public offerings: Amazon.com, Barnes and Noble, Walmart, Booksamillion and Overstock.com. Amazon.com accounted for 77% of the U.S. online sales for books, music and video in 1999, and Barnes and Noble accounted for 12% (Latcovich and Smith

the contracts firms secure with shipping companies. For instance, knowing whether a firm works with UPS, FEDEX, or USPS can help further in controlling for the shipping costs. Unfortunately, such information is not available from firms' websites and they are unwilling to release it. A lot of firms appear to work simultaneously with more than one shipping company, depending on the location of their warehouses.

[21] We added 0.0001 to zero before making the logarithm transformation.

(2001)). Walmart is a well-known, traditional, low-cost retailer. Booksamillion is also a relatively large bookseller, although not nearly as large as Amazon.com or Barnes and Noble. Overstock.com is an online discount book retailer. Thus, we have reason to believe that the variable PUBLIC controls effectively for relatively larger size and likely lower cost structure of these retailers. We also experimented with a "big 2" dummy which takes on a value of 1 for Amazon.com and Barnes and Noble to further control for the dominant position of these two retailers. This dummy had little effect when considered next to the PUBLIC variable. At most 8% of the total variation is explained by firm specific variables, as evidenced by the R^2's from logarithm specifications I and II in the left panel, implying that the included firm specific variables contribute little to the variation in shipping fee.

Firm characteristics used in specification II may be too crude to account for all firm effects. To address this, we also ran regressions with firm dummies, as shown in specifications labelled III. In 3 of the 4 cases, going from specification I to III implies an increase in R^2 of at most 18%. That is, at most 18% of the variation in fee is explained by firm fixed effects, as opposed to between 56% to 73% explained by quality. The only exception to this general pattern occurs in the case of the logarithm specification including all observations. For this specification, R^2 increases 55 percentage points when firm fixed effects are added. However, this result is driven largely by the two zero-fee observations in the data. Recall that in the logarithm specification a small positive number was added to these observations to include them in the regression. This leads to a large, negative value for the dependent variable for these two observations. The variation induced by these two observations are captured when firm specific effects are introduced. Therefore, with firm effects the R^2 increases dramatically.

Overall, these results suggest that firm characteristics matter to some extent in determining fee given a quality level. Figure 1 shows the fitted fee-quality schedules for well-known retailers and the average schedule across all retailers. For these estimates, we used the logarithm specification III excluding the zero-fee observations. From (13), these estimates are given by $\widehat{f}_{ij} = \widehat{\theta}_i q_{ij}^{-\widehat{\beta}}$. The estimated schedules for these three retailers are not far from each other. Barnes and Noble appears to offer a slightly more expensive schedule compared to Amazon.com, and Walmart has the cheapest estimated schedule, without taking into account the standard deviations. All of these three sellers have schedules that lie below the average schedule across firms. Table 5 reports the estimated firm specific fixed constant, $\widehat{\theta}_i$, for all retailers except for the two outliers. Note that estimates for all big retailers (in bold) lie below the median and the mean, and discount book retailers, such as Overstock.com, Gene's Books, Discountpcbooks, and Softdiscountbooks tend to offer the lowest fees. Less known, small retailers such as Motorbooks and The Book Cell tend to charge much higher fees.

Figure 2 shows the relationship between fee and quality for Option I, the standard and most commonly used option by firms and possibly by consumers. There is a high degree of variation in quality and fee across firms for this option, with 14 distinct quality levels offered by firms. A negative relationship between shipping fee and shipping quality is observed up to an average shipping time of

10 days. After that, the fee seems to increase somewhat abruptly, but then continues to decline with average shipping time. Well-known firms, such as Amazon.com, Barnes&Noble, and Walmart offer higher quality and lower fees, whereas less well-known firms such as Motor Books or Bargain Books offer very low quality and much higher fees than we would expect for these quality levels. These latter sellers are responsible for the abrupt jump in fees for lower quality levels. It is interesting to note that there is bunching of sellers at certain fee-quality combinations. For instance, Ecampus and Booksamillion have fee-quality combinations that almost match Amazon.com's. In fact, they differ only by a few cents. The two largest sellers, Amazon.com and Barnes and Noble, also offer very similar quality levels and fees. Note also that Blackwell's and Digitalguru offer free shipping, and they are the only two retailers that seem to be the gross outliers in the general fee-quality relationship. For the full range of observations in Figure 2, both the logarithm and level specifications in (13) and (14) still yield a mildly significant negative relationship between fee and quality, with and without controlling for firm characteristics.[22]

The pattern for other shipping options is similar, as shown in Figure 3, although they exhibit much less variation across firms, because these are offered by a smaller number of sellers compared to the standard shipping option in Figure 2. The pattern for Option II also reveals that as quality declines, fee does not decrease by much. This pattern, as in Figure 2, suggests that lower quality firms offer fees that are above what is expected. Option III exhibits a much clear decline in fee with quality, and Option IV has very little variation in quality. Since the standard shipping option in Figure 2 is one of the most commonly used by both firms and consumers, and exhibits much more variation across firms compared to other options, it will be our main focus in the price analysis to follow.

In summary, our main conclusions from the analysis of the fee-quality relationship are that *i*) the number and the quality of shipping options and the corresponding shipping fees vary considerably across firms, *ii*) quality is a major determinant of the shipping fee, but firm characteristics can matter as well, *iii*) well-known retailers and discount retailers tend to have lower fees for a given quality, and these firms also offer relatively high quality, and *iv*) at least for the standard shipping option (Option I) and the next faster one (Option II), the shipping fees offered by very low quality sellers are not much lower compared to the fees offered by relatively high quality sellers. Observations (*iii*) and (*iv*) hint to the possibility that low quality retailers may be offering fees above what their quality levels would justify, or alternatively, high quality retailers offer fees below what their quality levels would require. The vertical product differentiation model implies that shipping fee should increase as shipping quality increases, controlling for firm characteristics. However, the imperfect information model implies that the shipping fee can be higher for low quality firms, depending on the parameters. Although not conclusive, the patterns observed so far give more support to the imperfect information

[22]Excluding the two outliers, we ran an OLS regression similar to (13) but only for the the first shipping option, i.e. $j = 1$. For the logarithm specification without controls, we obtain a coefficient estimate of -0.17 with a standard error of 0.10, significant at 10%. With controls, the estimated coefficient is -0.23 with a standard error of 0.09, significant at 2%.

model. Of course, at this point these patterns can also be consistent with a scale economies based explanation, where large firms offer lower fees for higher quality because they have much lower costs, or with a specialization based explanation, where specialized retailers have differentiated products so that they can charge both higher base prices and higher fees, and offer lower quality. In the following section, we investigate more closely the relevance of the two main theories we discussed by examining the relationship between base price, shipping quality and shipping fee. In doing so, we control for a variety of firm characteristics so that we can address concerns about other potential influences, such as scale economies or specialization.

3.2 The relationship between prices and shipping options

Figure 4 looks at the simple bivariate relationship between shipping fee for the standard option (Option I) and the average base price of a firm. For this figure, we only focus on *common* books across sellers, because we want to assess the relationship between a firm's price and its shipping fee for product markets where many firms participate. Note again the outliers with zero fee. Their influence on the relationship between fee and base price is substantial. Furthermore, these zero-fee outliers charge relatively high average base prices. At first sight, the behavior of these outliers appears to be consistent with the vertical product differentiation model, which predicts a negative relationship between base price and fee. However, once these outliers are dropped, the figure suggests that, regardless of the kind of books (technical or bestseller), time period, and the number and composition of sellers competing in the market, there is a strong positive association between base price and shipping fee, as seen from the estimated slope coefficients. This appears to be more in favor of the imperfect information model. Of course, the main question here is whether this relationship prevails when we control for other observables. This is what we undertake next.

The relationship between base price and shipping fee can be analyzed in a linear framework, as suggested by both models. Using (3), we consider the following projection for the base price, p_{ik},of book k at seller i

$$p_{ik} = \alpha + \beta q_i + \lambda f_i^O + \boldsymbol{\gamma} \mathbf{x}_i + \nu_k + \varepsilon_{ik}. \tag{15}$$

where v_k is a book specific fixed effect, and ε_{ik} is a normally distributed, mean zero error term that varies across firms and products. As before, $\mathbf{x}_i$ is a set of firm-specific variables uncorrelated with ε_{ik} and q_i is firm i's shipping quality. Earlier we found that a firm's shipping fee is explained in part by the firm's quality choice and to some extent by firm characteristics. The coefficient of fee, λ, should be viewed as the remaining effect of fee on price after such firm characteristics and firm's quality choice are controlled for. The vertical differentiation model implies a negative sign for this remaining effect, whereas the imperfect information model implies a positive relationship.[23]

[23] A potential problem with (15) is the possible correlation between the fee f_i^O and the error term ε_{ik}. The identification

Similarly, both models predict a linear form for the total price, P_{ik}. We consider the following projection inspired by (4)

$$P_{ik} = \alpha + \beta q_i + \boldsymbol{\gamma} \mathbf{x}_i + v_k + \varepsilon_{ik}. \tag{16}$$

In calculating the total price, we take into account the free shipping threshold. If a book qualifies for free shipping, its total price is equal to its base price; otherwise, the total price is the sum of the base price and the shipping fee. In the base price regression, we chose not to make this adjustment to the fee for free shipping, because we wanted to focus on the general relationship between prices and fee levels across firms. Nevertheless, our results below change little if we adopt this approach also in the base price regression.

Table 6 provides estimates from several specifications that are variants of (15). For robustness, we use two separate datasets assembled in October 2003 and January 2004. We have 21 sellers in both data sets, but seller identities do not coincide perfectly because some of the sellers went out of business between the sample dates and others entered.[24] We include both technical and bestseller books in our sample controlling for book fixed effects, because the theories discussed are expected to apply to different types of books, albeit possibly with different degrees. We also combine the datasets measured in two separate points in time and report the resulting coefficients after including a time dummy, as we expect the cross sectional behavior to be persistent over time. In all regressions, we eliminated two outlier sellers with zero-fee observations (Blackwell's and Digitalguru), as these two extreme cases influence the results unduly. In addition to OLS, we also use a median regression (LAD) and an OLS regression in logarithms (LOG) for further robustness against remaining potential outliers.

In all specifications, the coefficients of quality, SHIPTIME, and fee, SHIPFEE, are consistently positive and highly significant, while magnitudes vary somewhat across specifications. Nevertheless, for a given specification, the magnitudes of the coefficients are not far apart for the two different times of measurement. Based on the pooled OLS specification III, a one day improvement in shipping fee is associated with up to a \$1.19 decrease in price (4.2% of average price in the sample). Based on the same specification, a one dollar increase in shipping fee is associated with a price increase of \$3.19 (11% of average price in the sample). In general, these results suggest strongly that a lower base price is associated with higher quality and, at the same time, a lower fee.

In all specifications, free shipping has a negative and highly significant coefficient. Recall that

of λ hinges on the assumption that the part of f_i^O accounted by unobservables is uncorrelated with ε_{ik}. This is equivalent to assuming that unobservables that affect firm's price are uncorrelated with unobservables that affect firm's fee. If an unobserved variable causes both the fee and base price to be higher for a firm, then the estimates would be biased upwards. However, lack of a straightforward instrument for shipping fee prevents us from assessing the magnitude of such a potential bias.

[24]The number of sellers is less than the number in the analysis of fee in the previous section because the comprehensive price search engine we used to obtain book prices did not contain sufficient number of books from these sellers for statistical precision.

FREESHIP takes on a value of 1 for retailers with permanent free shipping for all purchases, or for those which offer free shipping for purchases below \$25, and a value of 0 if a retailer does not offer free shipping at all, or offers it for a purchase threshold that is more than \$25.[25] According to the estimates from pooled OLS specification III, a firm that adopts a free shipping threshold at or below \$25 has a price that is lower by \$1.38 (4.8% of average price in the sample). This negative association appears to be in line with our argument that lower price and a lower free shipping threshold go together and reinforce each other.

Other variables included also have varying magnitudes and degrees of significance. Some variation is expected because we do not have a very large number of sellers, and the identities of sellers are different across the two time periods. In general, larger firms, firms that have a product scope beyond books, and firms with a physical presence offer higher prices. These results are consistent with the previous findings that larger and better-known firms in Internet book retailing command a higher price (e.g. Smith and Brynjolfsson (2001)), and that firms with a physical presence tend to charge higher prices (e.g. Pan, Venkatesh and Ratchford (2002)). Thus, results from our sample are consistent the relationships between prices and firm characteristics found in earlier studies. Older firms, and firms that sell many categories of books, offer lower price. Books that have lower availability at a given seller have a higher base price, which is consistent with standard demand-supply theory. Some service dimensions also command a price premium. Firms with a higher number of shipping options and, to a lesser extent, those that allow return without a restocking fee, command a higher price. Note also that the coefficients of SHIPTIME and SHIPFEE in general have much higher absolute value compared to almost all the other variables included, indicating their economic significance. Regressions in logarithms also confirm this. Note from pooled LOG specifications that a 10% decrease in SHIPTIME induces about 3% increase in base price, and a 10 percent increase in SHIPFEE leads to about 4.5% increase in base price.

Overall, the base price regressions suggest a strong negative relationship between a firm's quality choice and its price, and a strong positive relationship between its fee and its price, controlling for quality. This is inconsistent with the standard vertical product differentiation model with perfect information, which implies a positive relationship between base price and quality, and a negative relationship between base price and fee, controlling for quality. The results lend more support to models with imperfect information, an example of which was given in Section 2.2. The imperfect information model with sequential price discovery explains the joint variation in base price, shipping fee and shipping quality much better than the vertical product differentiation model. Moreover, the negative association between base price and free shipping threshold also supports the imperfect information model, and is in line with our conjecture that lower price and a lower free shipping threshold should reinforce each other. As discussed earlier, this pattern may be consistent with a

[25] A few firms offer free shipping above \$50, and some others do so above \$99.

number of different ways these models can be cast.

Finally, consider the behavior of total price. Table 7 reports estimates from several specifications of (16), in a manner similar to Table 6. Note that the total price reflects free shipping if the book is listed as qualified for free shipping by the seller. In most specifications with controls, SHIPTIME has a negative coefficient, and the coefficient usually is highly significant. However, in some cases with controls, the coefficient is positive, although significantly positive in only one case. The October 2003 data have much more consistency in terms of the sign and significance of this coefficient. In general, these results suggest a mild positive association between total price and quality. Higher quality firms tend to offer a higher price, but the results are not as strong as in the case of base price regressions. From the base price regressions, we know that firms with a low base price tend to offer higher shipping quality. Since fee tends to increase in general with quality, it is quite possible that total price is higher for low base price firms. Thus, the pattern for the total price depends on the relative rates of increases in the fee and the base price with quality. Since the data in two points of time differ in the number of firms, it is likely that these rates are somewhat different as well, contributing to the discrepancy in sign and magnitude of SHIPTIME in the total price regressions. Overall, we do not find a very strong relationship between total price and quality. Thus, there is no overwhelming evidence in favor of the vertical product differentiation model, which predicts a positive relationship between total price and quality. In contrast, the imperfect information model implies that total price and quality can be negatively correlated, but does not rule out a positive relationship, so it is difficult to reject the imperfect information model.

While we emphasized consumer information based theories in explaining the observed empirical regularities, other plausible theories can also be offered. Dynamic theories of reputation and loyalty suggest that firms may choose to offer higher shipping quality at a lower fee with the hopes of creating repeat purchases and consumer loyalty in the long run. Firms that wish to signal they are going to be around in the long run may engage in such strategies, whereas firms that adopt myopic, hit-and-run type of strategies may choose to offer higher prices and lower quality. We leave exploration of these richer dynamic models to future research.

4 Conclusion

We examined Internet retailers' strategies for shipping, an essential component of competition among Internet retailers. In arriving at our results, we exploited the fact that shipping service is almost always separately priced, unlike other services attached to the good or other desirable firm characteristics, whose prices are implicit in the base price of the good. Evidence presented suggests that firms differ considerably in the quality of shipping they offer and shipping fees they charge. Firms that charge lower base prices tend to offer lower shipping fees and higher shipping quality, as measured by average delivery time. These results are robust over time and prevail after controlling for a variety of observables. Most notably, scale economies and retailer specialization do not appear to be driving our results, as we control for these and other retail characteristics. The observed patterns cannot be readily reconciled with vertical product differentiation models under perfect consumer information. Imperfect information models appear to explain them better. Our findings suggest that informational asymmetries among consumers may play a role not only in generating the well-known price dispersion in Internet markets, but also in how firms price their attributes and the services attached to their products. The shipping fee does not appear to mitigate the dispersion in base price, as one would expect in a perfect consumer information setting.

The analysis was simplified by focusing on book retailing, where shipping options have a relatively simple structure. Extending the analysis to other products and industries is desirable for robustness and further insight. We expect the strength of the association between base prices and shipping fees to vary by product type. Search costs do matter for information acquisition by consumers. Books are relatively cheap products, so consumers may not have a very high incentive to search, as search costs may constitute a large fraction of a book's value. This may increase a seller's ability to charge higher prices to non-searching, uninformed consumers. For expensive goods, consumers have more incentive to search and gather information, as search costs may be relatively small compared to the savings that may be realized when a lower price seller is found for an expensive item. Incentives to search more can reduce the price dispersion across sellers, and may limit sellers' ability to potentially exploit uninformed consumers by offering lower quality and charging higher fees for services. The analysis would also be enhanced by the availability of data on the size of consumer purchases. Having average quantity or sales information by firm would be useful in assessing the separate role of shipping policies in affecting demand. Such data would also be helpful in further analysis of the nature of free shipping offers.

References

[1] Baye, Michael and Morgan, John (2001) "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets", *American Economic Review*, **91**: 454-74.

[2] Baye, Michael, Morgan, John and Scholten, Patrick (2002) "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site", forthcoming, *Journal of Industrial Economics.*

[3] Baye, Michael and Morgan, John (2003) "Red Queen Pricing Effects in E-retail Markets", Manuscript, Indiana University and University of California at Berkeley.

[4] Baylis, Kathy and Perloff, Jeffrey M. (2002) "Price Dispersion on the Internet: Good Firms and Bad Firms", *Review of Industrial Organization,* **21:** 305-324.

[5] Brynjolfsson, Erik and Smith, Michael (2000) "Frictionless Commerce? A Comparison of Internet and Conventional Retailers", *Management Science,* **46:** 563-585.

[6] Burdett, Kenneth and Judd, Kenneth (1983) "Equilibrium Price Dispersion", *Econometrica,* **51**: 955-969.

[7] Campanelli, Melissa (2002) "Who Pays to Get It There?", *Entrepreneur,* February.

[8] Smith, Michael and Brynjolfsson, Erik (2001) "Consumer Decision-Making at an Internet Shopbot: Brand Still Matters", *Journal of Industrial Economics,* **49:** 541-558.

[9] Chevalier, Judith and Goolsbee, Austan (2003) "Measuring Prices and Price Competition Online: Amazon and Barnes and Noble", forthcoming, *Quantitative Marketing and Economics.*

[10] Choi, C. J. and Shin, H. S. (1992) "A Comment on a Model of Vertical Product Differentiation", *Journal of Industrial Economics*, **40:** 229-232.

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

[12] Dellaert, Benedict, and Kahn, Barbara (1999) "How Tolerable is Delay? Consumers' Evaluations of Internet Web Sites after Waiting", Working paper, Center for Economic Research, Tilburg University.

[13] Ellison, Glenn and Ellison, Sarah Fisher (2001) "Search, Obfuscation, and Price Elasticities on the Internet", mimeo, MIT.

[14] Latcovich, Simon and Smith, Howard (2001) "Pricing, Sunk Costs, and Market Structure Online: Evidence from Book Retailing", *Oxford Review of Economic Policy,* **17:** 217-234.

[15] Lehmann-Grube, Ulrich (1997) "Strategic Choice of Quality When Quality is Costly: the Persistence of the High-quality Advantage", *Rand Journal of Economics*, **28:** 372-384.

[16] Lynch, John G., and Ariely, Dan (2000) "Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution", *Marketing Science,* **19:** 83-103.

[17] Mandel, Naomi, and Johnson, Eric (1998) "Constructing Preferences Online: Can Web Pages Change What You Want?", Working Paper, University of Pennsylvania.

[18] Menon, Satya, and Kahn, Barbara (1998) "Cross-Category Effects of Stimulation on the Shopping Experience: An Application to Internet Shopping", Working Paper, Wharton School, University of Pennsylvania.

[19] Morwitz, V., Greenleaf, E. A., and Johnson, E. J. (1998) "Divide and Prosper: Consumers' Reactions to Partitioned Prices", *Journal of Marketing Research,* **35:** 453-463.

[20] Motta, Massimo (1993) "Endogenous Quality Choice: Price vs. Quantity Competition", *Journal of Industrial Economics*, **41:** 113-131.

[21] Pan, Xing, Venkatesh, Shankar and Ratchford, Brian (2002) "Price Competition Between Pure Play vs. Bricks-and-Clicks e-Tailers: Analytical Model and Empirical Analysis", Working paper, University of Maryland.

[22] Pan, Xing, Ratchford, Brian T., Shankar, Venkatesh (2002) "Can Price Dispersion in Online Markets Be Explained by Differences in E-tailer Service Quality?", *Journal of the Academy of Marketing Science,* **30:** 433-445.

[23] Pan, Xing, Ratchford, Brian T., Shankar, Venkatesh (2003) "Why aren't the Prices of the Same Item the Same at Me.com and You.com?: Drivers of Price Dispersion Among E-tailers", Working Paper, Sloan School of Management, MIT.

[24] Reinganum, Jennifer (1979) "A Simple Model of Equilibrium Price Dispersion", *Journal of Political Economy,* **87:** 851-858.

[25] Salop, Steve and Stiglitz, Joseph (1977) "Bargains and Ripoffs: A Model of Monopolistically Competitive Prices", *Review of Economic Studies,* **44**: 493-510.

[26] Shaked, Avner and Sutton, John (1982) "Relaxing Price Competition through Product Differentiation", *Review of Economic Studies*, **49:** 3-13.

[27] Smith, Michael and Brynjolfsson, Erik (2001) "Consumer Decision-Making at an Internet Shopbot: Brand Still Matters", *Journal of Industrial Economics,* **49:** 541-558.

[28] Tedeschi, Bob (2001) "E-commerce Report; Shipping Fees: Some Scrimp, Some Profit", *The New York Times,* June 11.

[29] Thaler, Richard (1985) "Mental Accounting and Consumer Choice", *Marketing Science,* **4:** 199-214.

[30] Tirole, Jean (1988) *The Theory of Industrial Organization.* The MIT Press.

A Equilibrium in the imperfect information model

In this appendix we show that the proposed equilibrium for the imperfect information, sequential price discovery model in Section 2.2 indeed exists and characterize its properties. We follow the basic arguments in Salop and Stiglitz (1977).

In equilibrium, the base price and fee charged by a high price firm must be such that an uninformed consumer does not prefer to become informed.[26] From (9), (10) and (12), this implies

$$(1-\mu)(\overline{p}-\underline{p}) \leq \sigma, \tag{17}$$

$$\overline{p}-\underline{p} \leq \sigma, \tag{18}$$

$$\overline{p}+\mu f^L \leq \underline{p}+\mu f^H - \mu\Delta u + \sigma + \omega. \tag{19}$$

Since each informed consumer has access to a list of base prices in the market, such a consumer has no incentive to visit any further firm in the proposed equilibrium, if he first visits the firm with the lowest base price, because any other firm has either the same price, fee, and quality, or higher price, higher fee and lower quality. In equilibrium, all firms with low base price must have a total price $\underline{P} = \underline{p} + f^H = p^* + c^H$,where $p^* = 2\sqrt{Fc}$ is the competitive base price equal to the minimum average cost of sale and shipping, and $\underline{p} \geq 0$ and $f^H \geq 0$. Technically, a firm could lower its base price below zero and increase its shipping fee accordingly and attract all informed consumers. To prevent this, we assume that negative prices are not feasible. If $\underline{P}$ was greater than $p^* + c^H$, a firm could lower its base price, $\underline{p}$, slightly and could attract all informed consumers and increase its profits, because informed consumers first visit the firm with the lowest base price. When all firms charge $\underline{P}$, deviating to a higher price causes a firm to lose all its customers. A firm would also never deviate and charge below $\underline{P}$, as the firm would be charging a price below its minimum average cost, resulting in a loss.

The partition of the total price $\underline{P}$ must be such that $\underline{p} = 0$, and $f^H = p^* + c^H$. Suppose that $\underline{p}$ was above zero. Then, any firm charging this price could lower its price slightly and raise its fee by an equal amount, and attract all informed consumers. If $\underline{p} = 0$, no firm deviates to raise its price, because then the firm would attract no informed consumers. We also assumed that prices below zero are not feasible. Then, the fact that $\underline{p} = 0$ and $\underline{P} = p^* + c^H$ implies that $f^H = p^* + c^H$.

No low price firm wants to deviate from high quality and offer low quality instead. A firm that deviates and offers a low quality incurs a shipping cost of c^L and can offer a consumer a surplus of at most $u^L - c^L$. Since $u^L - c^L < u^H - c^H$, no informed consumer prefers to purchase from this firm.

Now consider the high price firms. From (18), any price above $\underline{p} + \sigma$ cannot be an equilibrium price, because at that price an uninformed consumer prefers to become informed, and a firm deviating

[26] We assume that, when indifferent between being informed and searching randomly, an uninformed consumer chooses to search.

to $\underline{p} + \sigma$ could attract all such consumers. Therefore, $\overline{p} \leq \underline{p} + \sigma$. If $\underline{p} < \overline{p} < \underline{p} + \sigma$, then a firm charging $\overline{p}$ could deviate to $\underline{p} + \sigma$ without inducing its uninformed customers to become informed. It follows that $\overline{p} = \underline{p} + \sigma$. At this price, it is not profitable to deviate and raise or lower price. If a firm raises its price, it loses all its customers, as they prefer to become informed by (17) and (18). If a firm lowers its price, then the expected price in the market falls, but the minimum price, $\underline{p}$, does not change. Benefits from becoming informed on the left hand side of (17) then decreases. The purchasing behavior of the uninformed consumers then does not change. The behavior of the informed consumers does not change, either. Thus, the deviation is not profitable, as it only lowers profits. This implies $\overline{p} = \sigma$, since $\underline{p} = 0$. Therefore, from (19), we have $f^L \leq f^H - \Delta u + \frac{\omega}{\mu}$. It is easy to see that this must hold with equality, $f^L = f^H - \Delta u + \frac{\omega}{\mu}$, because, as in the case of base price, deviation from this fee in either direction results in a loss of profit.

Finally, consider the possibility of a high price firm deviating from low quality and offering high quality. If the deviant continues to charge $\overline{p}$ and f^L, the benefits to becoming informed on the left hand side of (19) go down, because $u^H > u^L$. Uninformed consumers continue to visit stores randomly, meaning that the deviant does not acquire any additional customers. But then the deviant makes less profit than other high price firms because its cost of shipping has increased from c^L to c^H. If it increases its price to compensate for this cost, (18) then implies that uninformed consumers want to become informed, and the deviant attracts no consumers. A similar argument applies to the possibility of increasing the shipping fee to compensate for the higher cost. Therefore, a high price firm does not want to deviate from low quality shipping and alter its price.

In equilibrium, low price firms sell to informed consumers and lucky uninformed consumers who happen to visit a low price firm. High price firms sell only to uninformed consumers. Therefore, the quantities sold by a low and a high price firm, respectively, are

$$\underline{x} = \frac{\phi}{\mu n} + \frac{1-\phi}{n}, \tag{20}$$

$$\overline{x} = \frac{1-\phi}{n}. \tag{21}$$

Free entry ensures zero profit for each firm in the market. For firms with low base price, the total price must equal the minimum average cost, and for firms with high base price, total price must equal their average cost. The expressions for prices, fees and the cost function then yield

$$\underline{p} + f^H = p^* + c^H = \frac{F}{\underline{x}} + c\underline{x} + c^H, \tag{22}$$

$$\overline{p} + f^L = p^* + c^H + \sigma - \Delta u + \frac{\omega}{\mu} = \frac{F}{\overline{x}} + c\overline{x} + c^L. \tag{23}$$

Given the expressions for $\underline{x}$ and $\overline{x}$ in (20) and (21), the equations in (22) and (23) jointly determine the number of firms, n, and the fraction of low price firms, μ. This finishes the characterization of the proposed equilibrium.

One final necessary condition for the proposed equilibrium to exist is that the fraction, μ, of informed consumers should not be too large and/or the declining portion of the average cost curve should not be very steep. If the market consists of primarily informed consumers (large μ), there will be a small market for firms with high base price, as they sell only to unlucky uninformed consumers. If the average cost curve is steep enough, there will not be enough high cost consumers to support even one high-price firm. The necessary condition is $(1-\mu) > \overline{x}/\underline{x}$, where $\underline{x} = \sqrt{F/c}$ is the quantity that yields the average minimum cost.

Note that if only σ is very large or both σ and ω are very large, $\overline{p}$ is the monopoly price, s, and high price firms offer higher quality and charge a monopoly fee of $f^H = u^H + \omega$. Thus, controlling for quality difference, we still have $\overline{p} > \underline{p}$ and $f^L > f^H$. If only ω is very large, then we have $\overline{p} = \sigma$ and $f^H = u^H + \omega$, i.e. high price firms charge monopoly price for quality, and their base prices are high enough to keep uninformed consumers purchasing randomly. In this case, too, we have $\overline{p} > \underline{p}$ and $f^L > f^H$, controlling for quality difference.

Finally, consider the case where becoming informed is equivalent to learning all base prices, shipping qualities and shipping fees in the market. Then, under proposed equilibrium, (7) is equivalent to

$$u^H - p_{(1)} - f_{(1)} \geq u^H - p_{(2)} - f_{(2)},$$

which implies that base price and shipping fee for low price firms are identified up to the sum $\underline{P} = p^* + c^H$. Also, (8), (10) and (12) imply

$$\overline{P} \leq \underline{P} + \frac{\sigma}{1-\mu} - \Delta u, \tag{24}$$

$$\overline{p} \leq \underline{P} + \sigma - (1-\mu)\Delta u - E[f], \tag{25}$$

$$\overline{P} \leq \underline{P} + \sigma + \omega - \Delta u. \tag{26}$$

Let $A = \min\{\underline{P} + \sigma + \omega - \Delta u, \underline{P} + \frac{\sigma}{1-\mu} - \Delta u\}$. Then, (25) allows identification of the high base price, $\overline{p}$. It is easy to verify that

$$\begin{aligned} \overline{p} &= \underline{P} + \sigma - (1-\mu)\Delta u - E[f], \\ f^L &= A - \overline{p}, \\ \overline{P} &= A. \end{aligned}$$

Note that $\overline{P} = \underline{P} + \min\{\sigma + \omega, \frac{\sigma}{1-\mu}\} - \Delta u$, so $\overline{P} > \underline{P}$ as long as $\min\{\sigma + \omega, \frac{\sigma}{1-\mu}\} > \Delta u$. Since $\underline{P} \geq \underline{p}$ for all choices of $\underline{p}$ by low base price firms, we have $\overline{p} > \underline{p}$, as long as $\sigma - (1-\mu)\Delta u - E[f] > 0$. Thus, controlling for quality difference, we have $\overline{p} > \underline{p}$, as long as $\sigma > E[f]$. We also have $f^L =$

$\min\{\omega, \frac{\mu\sigma}{(1-\mu)}\} - \mu\Delta u + E[f]$. Note that $E[f] = (1-\mu)f^L + \frac{1}{n}\sum_{i=1}^{n\mu} f_i^H$, where f_i^H is the shipping fee for the i-th firm with low base price, and the summation is taken over all firms with low base price. Therefore, $f^L = \min\{\frac{\omega}{\mu}, \frac{\sigma}{(1-\mu)}\} - \Delta u + \frac{1}{n\mu}\sum_{i=1}^{n\mu} f_i^H$. Thus, controlling for quality difference, f^L exceeds any f_i^H as long as $\min\{\omega, \frac{\mu\sigma}{(1-\mu)}\}$ is high enough. Therefore, when σ and ω are sufficiently high, we still obtain the result that high base price firms charge higher shipping fee, controlling for quality difference.

B Data collection

Our data were collected in two different points in time. The first wave was collected in a single week during October of 2003, and includes 133 different books sold by 21 book sellers. Many books are common across sellers. The second wave was collected in a single week during January of 2004, and includes 129 books sold by 21 book sellers. We avoided the holiday seasons in November and December for data collection, because most Internet sellers engage in unusual promotions in pricing and shipping during these periods. Furthermore, we kept the duration of data collection as short as possible (one week) to ensure sure that no substantial changes took place in seller composition, i.e. no entry and exit during data collection.

Book selection

To focus on as many common books as possible across sellers, we only selected relatively more popular books sold online. For this purpose, we chose two categories: bestsellers and technical books. The data for 133 books collected in October 2003 consist of 60 New York Times bestsellers and 73 Amazon.com technical bestsellers. The data for 129 books collected in January 2004 consist of 60 New York Times bestsellers and 69 Amazon.com technical bestsellers. Since all of these books are sold by many sellers, we obtain sufficient variation in prices and shipping fees across sellers. These books also belong to two well-defined product markets, bestsellers and technical books. Having two different categories of books is also important in demonstrating that the results apply to different categories in a similar way.

Seller selection

We only considered U.S. based sellers that sell books and possibly other items on the Internet. The data covers sellers that together account for a very large share of the all book sales made in the U.S., including all major sellers and many smaller sellers. We excluded all sellers that based their shipping fee schedule on weight of books sold, rather than the quantity of books, to make the structure of shipping policies as uniform as possible across sellers. 37 of the sellers provide regularly reported shipping options based on the quantity of books shipped, and we included all of them in the analysis of shipping fee-quality schedule. We ended up excluding many small sellers with no regular posted shipping fees. We also paid special attention to include a variety of sellers: general book sellers, specialized ones, and those that sell products other than books. This way our sample represents a broad set of sellers competing in the book market, not just those that are specialized in books. For the analysis of base and total prices, we focus on 21 sellers (out of 37 initially selected) for which price information on several books was available in the search engine used. For the remaining sellers, the search engine did not provide price quotes for a sufficient number of books.

Variables

Variables in our data include book prices and various seller characteristics (see Table 1 for variable descriptions), such as average delivery time, number of shipping options, age, scope, etc. We gathered price quotes through the search engine *www.fetchbooks.info* which allows comparison of book prices by more than 90 online booksellers. The large coverage provided by this search engine alleviates to some extent concerns about certain important sellers being strategically left out from the price list, which results in a selection problem. Shipping options and most of the other seller specific variables were collected directly from each bookseller's website. The proxy for firm size, the number of websites that have a link to a particular seller's website (SIZE), as well as the seller's number of years of operation on the Internet (AGE), was obtained from the information on websites listed on *www.alexa.com.* The only book-seller specific variable, availability (AVAIL), is obtained from sellers' websites using the expected time for a book to be shipment ready, as reported by the seller.

Variables	Description
PRICE	Base price of a book in dollars
SHIPTIME	Average delivery time in days for the slowest shipping option offered by a seller
SHIPFEE	Shipping fee for the standard, slowest option offered by a seller
PUBLIC	Dummy variable: 1 if a seller has made public offering, 0 otherwise
PHYSICAL	Dummy variable: 1 if a seller has physical stores, 0 otherwise
AGE	Number of years a seller has been selling on the Internet
SIZE	Logarithm of the number of links referring to a seller's website
SCOPE	Dummy variable: 1 if a seller offers goods other than books, 0 otherwise
GENERAL	Dummy variable: 1 if a seller offers many different categories of books, 0 if the seller specializes in a certain category
OPTION	Number of different shipping options offered by a seller
FREESHIP	Dummy variable: 1 if a seller offers free shipping for all purchases or only for purchases above $25, 0 otherwise
AVAIL	Average time in days for a seller to get a given book shipment ready
PHONE	Dummy variable: 1 if a sellers provides customer service by phone, 0 otherwise
RETURN	Dummy variable: 1 if a seller accepts returns without a restocking fee, 0 otherwise

Table 1. Description of variables used in the empirical analysis

	Fee Structure (Slowest Option)	Expected Delivery Range
FEDEX	$3.89 for 1 pound, $4.26 for 2 pounds*	1-5 business days
USPS	$1.00+$0.42 per pound**	2-9 business days
UPS	$3.86 for 1 pound, $4.26 for 2 pounds***	Depends on origin and destination
Amazon.com	$3.00 + $0.99 per book	3-7 business days
Barnes&Noble	$3.00 + $0.99 per book	3-8 business days
Walmart	$2.22 + $0.26 per book	3-7 business days
Powell's Books	$2.50 + $1.00 per book	7-14 business days

* Average of rates for 7 different delivery zones (Source: FEDEX ground and home delivery rates by service)

** Book rate (media mail) for all zones, upto 7 lbs. (Source: USPS domestic postal rates and fees)

*** Average of rates for 7 different delivery zones (Source: UPS rate and service guide, January 2004 edition)

Table 2. Shipping rates of 3 biggest shipping companies and selected sellers

		Shipping options for a single book			
		Option I	**Option II**	**Option III**	**Option IV**
quality (days)	**average**	7.91	3.24	1.63	1.03
	std	3.50	1.19	0.62	0.24
	cv	0.44	0.37	0.38	0.23
fee (dollars)	**average**	3.74	7.34	14.61	16.82
	std	1.42	3.41	7.97	2.19
	cv	0.38	0.46	0.54	0.13
	N	37	26	18	7

Table 3. Statistics for quality and fee for shipping options

	Results including all observations						Results excluding 2 observations with zero fee					
	Logarithm Specification			Level Specification			Logarithm Specification			Level Specification		
	I	II	III	I	II	III	I	II	III	I	II	III
SHIPTIME	-0.96***	0.91***	-1.14***	-3.22***	-3.12***	-3.39***	-0.71***	-0.72***	-0.77***	-3.16***	-3.05***	-3.32***
	(0.19)	(0.17)	(0.34)	(0.42)	(0.41)	(0.51)	(0.05)	(0.05)	(0.05)	(0.42)	(0.41)	(0.52)
SHIPTIME^2	-	-	-	0.15***	0.14***	0.16***	-	-	-	0.15***	0.14***	0.16***
				(0.03)	(0.03)	(0.04)				(0.03)	(0.03)	(0.04)
SIZE	-	0.22	-	-	0.24	-	-	0.02	-	-	0.11	-
		(0.15)			(0.30)			(0.03)			(0.30)	
AGE	-	-1.51	-	-	-0.24	-	-	-0.04	-	-	-0.144	-
		(0.98)			(0.29)			(0.16)			(0.30)	
GENERAL	-	0.28	-	-	-0.15	-	-	-0.06	-	-	-0.225	-
		(0.39)			(1.34)			(0.10)			(1.36)	
SCOPE	-	-1.06	-	-	0.54	-	-	0.04	-	-	1.327	-
		(1.02)			(1.20)			(0.09)			(1.09)	
PHYSICAL	-	-0.03	-	-	1.91	-	-	0.21**	-	-	2.19	-
		(0.46)			(1.37)			(0.10)			(1.39)	
PUBLIC	-	-0.14	-	-	-2.64*	-	-	-0.29**	-	-	-2.82*	-
		(0.43)			(1.41)			(0.12)			(1.46)	
Seller dummies	-	-	Y	-	-	Y	-	-	Y	-	-	Y
N	87	87	87	87	87	87	85	85	85	85	85	85
R^2	0.18	0.26	0.73	0.57	0.59	0.74	0.73	0.76	0.91	0.56	0.59	0.73

Notes: Robust standard errors in parantheses. (*),(**), and (***) indicate significance at 10%, 5%, and 1%, respectively for a two-sided test.

Table 4. Shipping fee regressions using alternative specifications

Seller	Estimated multiplier (θ_i) in the fee schedule, $f_{ij} = \theta_i q_{ij}^{-\beta}$
Overstock	6.57
Gene's books	11.19
Softdiscountbooks	11.72
Discountpcbooks	11.74
Buy.com	**13.16**
Booksamillion	14.28
Walmart	**14.28**
Booksfree	14.64
HPCB	15.04
Amazon	**15.06**
Biggerbooks	15.17
Strandbooks	15.29
Ecampus	15.69
A1books	16.10
Ecookbooks	16.30
Wordsworth	16.61
Barnes&Noble	**17.08**
TextbookX	17.36
MEDIAN	**17.48**
Powell's	17.60
Laissez Faire books	17.92
Totalcampus	17.99
Alibris	18.60
MEAN	**18.62**
Elephantbooks	18.69
Page1book	18.82
Digitalguru	19.26
Hamiltonbook	21.01
Wiley	21.98
Brain's books	22.15
Nerdbooks	22.27
Classbook	24.36
Booksmatter	25.59
Bookstrip	29.58
Motorbooks	31.72
Bargainbooks	33.18
Thebookcell	33.58

Table 5. Estimated multipliers for sellers' fee-quality schedules

	All books - October 2003					All books - January 2004					All books - Oct. 2003 and Jan. 2004 pooled				
	OLS I	OLS II	OLS III	LAD	LOG	OLS I	OLS II	OLS III	LAD	LOG	OLS I	OLS II	OLS III	LAD	LOG
SHIPTIME	0.58***	1.33***	1.84***	0.59***	0.25***	0.57***	1.59***	1.31***	0.51***	0.29***	0.56***	1.22***	1.19***	0.48***	0.30***
	(0.05)	(0.14)	(0.22)	(0.21)	(0.04)	(0.11)	(0.31)	(0.27)	(0.11)	(0.03)	(0.04)	(0.10)	(0.10)	(0.05)	(0.02)
SHIPFEE	2.63***	4.18***	6.09***	2.82***	0.42***	2.09***	3.28***	2.35***	2.03***	0.46***	2.29***	3.16***	3.19***	2.24***	0.46***
	(0.15)	(0.42)	(0.64)	(0.60)	(0.06)	(0.23)	(0.44)	(0.55)	(0.39)	(0.04)	(0.11)	(0.21)	(0.22)	(0.17)	(0.02)
FREESHIP	-	-2.84***	-3.99***	-2.75***	-0.13***	-	-0.52	2.25*	-0.14	-0.10***	-	-1.62***	-1.38***	-1.62***	-0.10***
		(0.34)	(0.46)	(0.52)	(0.02)		(0.67)	(1.18)	(0.46)	(0.01)		(0.24)	(0.37)	(0.29)	(0.01)
PUBLIC	-	0.78*	1.48***	-0.07	-0.02	-	0.65	0.63	-0.32	0.04*	-	0.16	0.39	-0.25	0.02***
		(0.46)	(0.52)	(0.49)	(0.02)		(0.82)	(0.82)	(0.44)	(0.02)		(0.29)	(0.27)	(0.17)	(0.00)
SIZE	-	0.23***	0.64***	0.02	0.01***	-	0.80**	0.29	-0.15	0.01	-	0.40***	0.24***	-0.06	0.006***
		(0.06)	(0.10)	(0.08)	(0.00)		(0.35)	(0.37)	(0.12)	(0.01)		(0.09)	(0.09)	(0.03)	(0.00)
PHYSICAL	-	0.75***	1.89***	-0.88	0.00	-	0.25	0.96	-0.31	0.07***	-	0.66***	-0.04	-0.09	0.02***
		(0.20)	(0.51)	(0.77)	(0.02)		(0.74)	(0.94)	(0.33)	(0.02)		(0.20)	(0.28)	(0.25)	(0.00)
AGE	-	-0.25***	-0.53***	0.19**	-0.01	-	-1.01***	-0.68*	0.08	-0.10***	-	-0.46***	-0.41***	0.14***	-0.06***
		(0.07)	(0.11)	(0.10)	(0.02)		(0.37)	(0.36)	(0.11)	(0.04)		(0.11)	(0.10)	(0.04)	(0.02)
GENERAL	-	-1.01**	-2.47***	-2.11***	-0.05***	-	-0.92	-3.17*	0.15	-0.09**	-	-1.33***	-0.81*	-1.78***	-0.03***
		(0.46)	(0.39)	(0.62)	(0.01)		(0.90)	(1.75)	(1.23)	(0.04)		(0.45)	(0.48)	(0.46)	(0.01)
SCOPE	-	1.67***	2.61***	0.42	0.05***	-	-1.35*	2.01	-1.45*	-0.04	-	0.91***	1.09***	0.47***	0.07
		(0.25)	(0.37)	(0.37)	(0.01)		(0.76)	(1.29)	(0.76)	(0.03)		(0.22)	(0.21)	(0.12)	(0.00)
OPTIONS	-	1.35***	1.27***	1.32***	0.13***	-	1.33***	0.71**	1.04***	0.10***	-	1.32***	1.28***	1.18***	0.13***
		(0.10)	(0.10)	(0.10)	(0.01)		(0.28)	(0.34)	(0.22)	(0.02)		(0.08)	(0.08)	(0.07)	(0.00)
AVAIL	-	-	0.32***	0.08	0.03***	-	-	0.12	0.04	0.04***	-	-	0.14***	0.09***	0.04***
			(0.06)	(0.05)	(0.01)			(0.09)	(0.05)	(0.01)			(0.04)	(0.04)	(0.00)
PHONE	-	-	0.86***	1.84***	0.06***	-	-	-2.44*	1.18**	-0.06**	-	-	0.22	0.74***	0.02**
			(0.28)	(0.49)	(0.01)			(1.27)	(0.54)	(0.03)			(0.32)	(0.25)	(0.00)
RETURN	-	-	-1.79***	0.91	0.01	-	-	2.30***	0.72	-0.01	-	-	1.19***	0.54***	0.009
			(0.50)	(0.60)	(0.02)			(0.85)	(0.44)	(0.02)			(0.28)	(0.16)	(0.00)
Book dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time dummies	-	-	-	-	-	-	-	-	-	-	Y	Y	Y	Y	Y
N	1729	1729	1729	1729	1729	1597	1597	1597	1597	1597	3326	3326	3326	3326	3326
R^2	0.921	0.938	0.942	0.811	0.956	0.952	0.954	0.954	0.843	0.967	0.972	0.974	0.974	0.853	0.973

Notes: Robust standard errors in parantheses. (*),(**), and (***) indicate significance at 10%, 5%, and 1%, respectively for a two-sided test. LAD regressions use 200 bootstrap replications to estimate standard errors. LOG refers to regression in logarithms of variables except for dummies.

Table 6. Base price regressions using alternative specifications

	All books - October 2003					All books - January 2004					All books - Oct. 2003 and Jan. 2004 pooled				
	OLS I	OLS II	OLS III	LAD	LOG	OLS I	OLS II	OLS III	LAD	LOG	OLS I	OLS II	OLS III	LAD	LOG
SHIPTIME	0.42***	-0.53***	-0.61***	-0.75***	-0.23***	0.64***	0.34	0.62***	-0.04	-0.04*	0.53***	-0.07	0.06	-0.42***	-0.11***
	(0.05)	(0.06)	(0.06)	(0.07)	(0.01)	(0.11)	(0.23)	(0.24)	(0.08)	(0.02)	(0.04)	(0.08)	(0.09)	(0.05)	(0.01)
PUBLIC	-	-6.10***	-5.75***	-4.99***	-0.25***	-	-3.82***	-2.87***	-3.42***	-0.16***	-	-4.95***	-4.27***	-4.16***	-0.20***
		(0.24)	(0.25)	(0.24)	(0.01)		(0.76)	(0.75)	(0.38)	(0.01)		(0.25)	(0.25)	(0.06)	(0.01)
SIZE	-	0.27***	0.07	-0.10*	0.01***	-	0.22	0.15	-0.28**	-0.02***	-	0.22**	0.05	-0.10*	-0.005***
		(0.06)	(0.07)	(0.06)	(0.00)		(0.30)	(0.34)	(0.14)	(0.01)		(0.09)	(0.10)	(0.06)	(0.01)
PHYSICAL	-	1.25***	-0.92**	-2.26***	-0.06***	-	1.68***	0.85	0.06	0.06***	-	1.49***	0.44	-0.71***	0.02**
		(0.22)	(0.38)	(0.42)	(0.01)		(0.63)	(0.80)	(0.33)	(0.02)		(0.20)	(0.31)	(0.24)	(0.01)
AGE	-	0.34***	0.41***	0.41***	0.07***	-	0.19	0.22	0.58***	0.26***	-	0.28***	0.29***	0.44***	0.15***
		(0.06)	(0.05)	(0.05)	(0.01)		(0.29)	(0.30)	(0.12)	(0.03)		(0.09)	(0.09)	(0.05)	(0.01)
SCOPE	-	-1.16***	-1.50***	-1.17***	-0.08***	-	0.78	1.33**	0.12	0.05***	-	-0.04	0.07	-0.46***	-0.002
		(0.21)	(0.22)	(0.19)	(0.01)		(0.63)	(0.60)	(0.35)	(0.01)		(0.25)	(0.24)	(0.13)	(0.01)
GENERAL	-	-2.77***	-1.16**	-2.23***	0.00	-	-2.95***	-1.84	0.38	-0.02	-	-3.02***	-1.51***	-2.20***	-0.01
		(0.46)	(0.47)	(0.61)	(0.01)		(0.84)	(1.34)	(1.37)	(0.03)		(0.44)	(0.44)	(0.66)	(0.01)
OPTIONS	-	0.67***	0.61***	0.59***	0.04***	-	0.67***	0.56**	0.45***	0.09***	-	0.68***	0.68***	0.61***	0.07***
		(0.11)	(0.11)	(0.08)	(0.01)		(0.23)	(0.25)	(0.10)	(0.01)		(0.10)	(0.09)	(0.06)	(0.01)
AVAIL	-	-	-0.14**	0.01	-0.01**	-	-	-0.38***	-0.28***	-0.06***	-	-	-0.24***	-0.04	-0.03***
			(0.07)	(0.05)	(0.01)			(0.11)	(0.08)	(0.01)			(0.06)	(0.03)	(0.01)
PHONE	-	-	1.06***	1.92***	0.06***	-	-	-1.05	1.80***	-0.04*	-	-	-0.22	1.33***	-0.005
			(0.27)	(0.28)	(0.01)			(0.95)	(0.41)	(0.02)			(0.27)	(0.18)	(0.01)
RETURN	-	-	2.17***	3.40***	0.13***	-	-	2.44***	2.03***	0.04**	-	-	1.75***	2.13***	0.06***
			(0.31)	(0.32)	(0.01)			(0.87)	(0.37)	(0.02)			(0.30)	(.21)	(0.01)
Book dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time dummies	-	-	-	-	-	-	-	-	-	-	Y	Y	Y	Y	Y
N	1729	1729	1729	1729	1729	1597	1597	1597	1597	1597	3326	3326	3326	3326	3326
R^2	0.866	0.913	0.916	0.788	0.946	0.950	0.951	0.952	0.832	0.959	0.965	0.970	0.970	0.832	0.962

Notes: Robust standard errors in parantheses. (*),(**), and (***) indicate significance at 10%, 5%, and 1%, respectively for a two-sided test. LAD regressions use 200 bootstrap replications to estimate standard errors. LOG refers to regression in logarithms of variables except for dummies.

Table 7. Total price regressions using alternative specifications

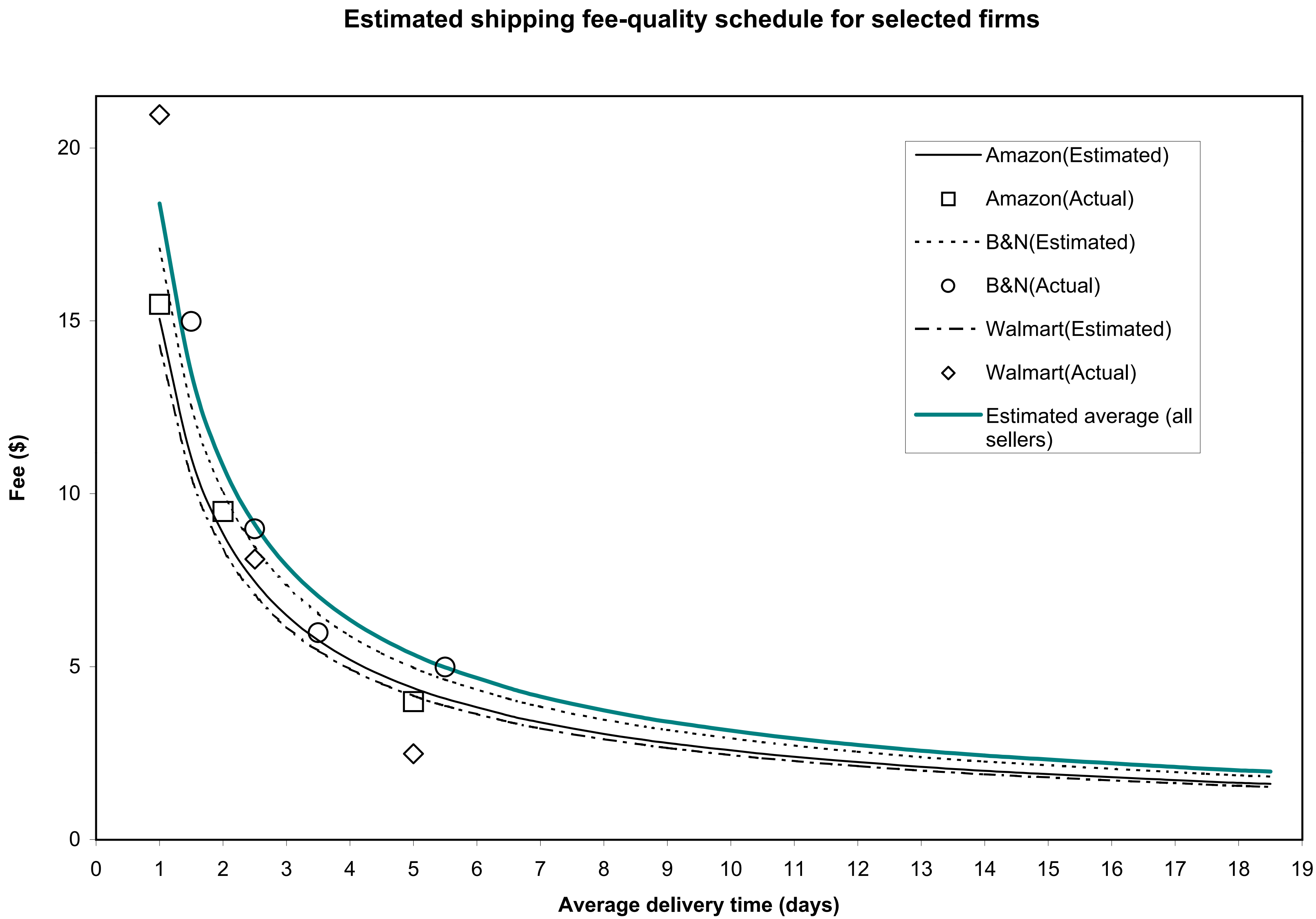


Figure 1. Estimated shipping fee-quality schedules for shipment of a single book

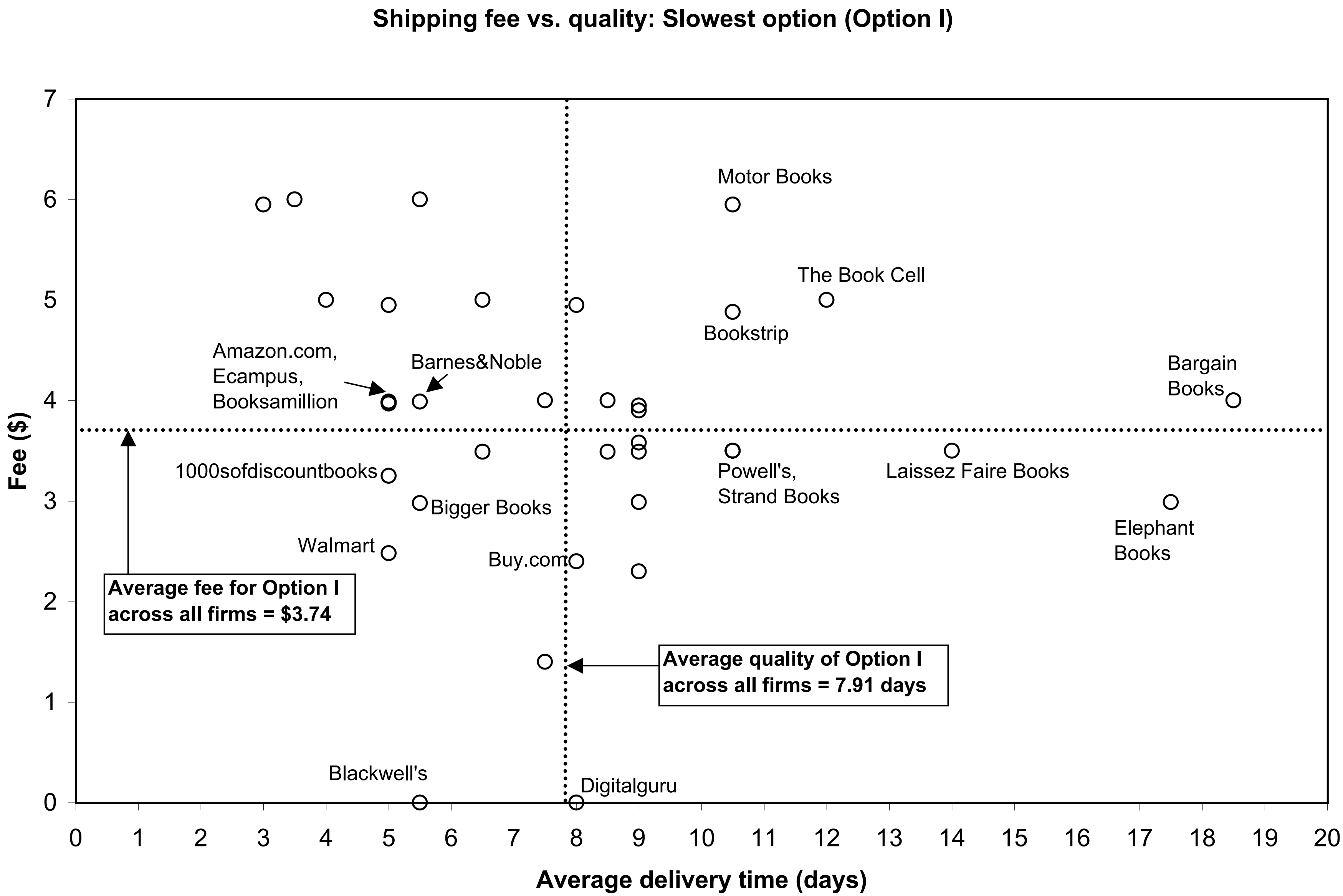
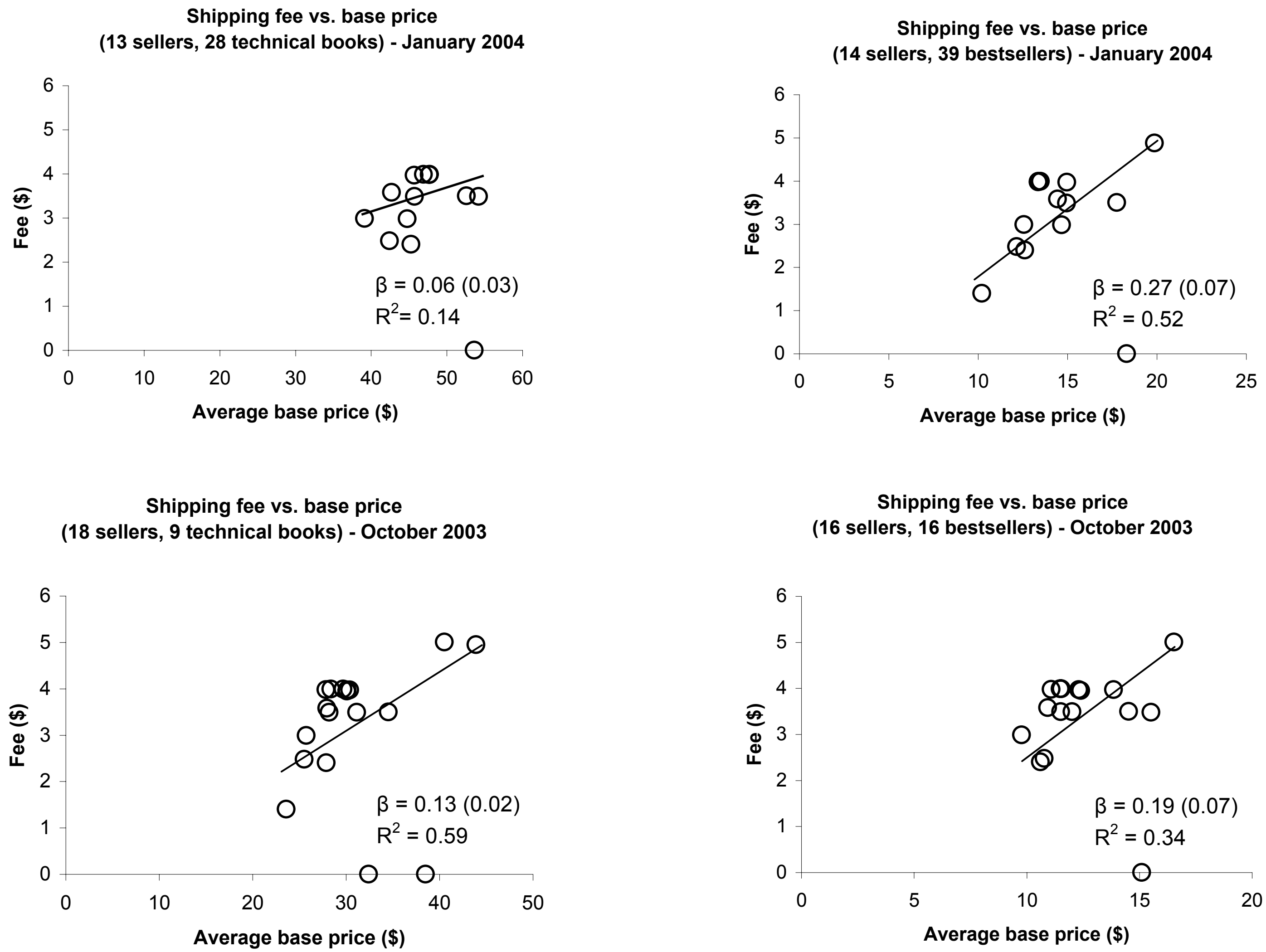


Figure 2. Shipping fee versus quality for the standard, slowest shipping options offered by booksellers

Figure 3. Fee-quality relationship for different shipping options offered by sellers



Note: Regression lines were obtained after dropping observations with zero fee. Standard errors in parantheses.

Figure 4. Bi-variate relationship between shipping fee and base price for different book types and time periods