Pricing Used Books on Amazon.com:

A Semiparametric Spatial Autoregressive Model of

Price Dispersion

Haoying Wang [*]

The Pennsylvania State University - University Park

E-mail: halking@psu.edu

Current Draft: June 22, 2013

---

[*]To whom correspondence should be addressed

Abstract

This paper investigates the pricing strategies of online used books sellers with a cross-sectional spatial autoregressive (SAR) model. In particular, this study focuses on the price competition among booksellers when they list used books for sale online. The used book condition categories (e.g., like new, very good, good, acceptable, and etc.) created by bookselling websites form a natural spatial context for booksellers to compete. The price competition among booksellers leads to significant endogenous spatial interactions in pricing strategies. The data used in this study is collected from the online used books listings at Amazon.com in July 2012, and the data contains all listings of the top 100 bestselling econometrics books and the top 100 bestselling environmental economics books according to the sales rank by Amazon.com. The model is estimated with the semiparametric adaptive estimator for SAR models proposed by Robinson (2010). The estimates are as efficient as the ones based on a conventional normal distribution function form of innovation terms, while they are more efficient than the pseudo maximum likelihood (ML) estimates with non-normal innovations. The more general though complicated estimation strategy is appropriate in this case given the potential unobserved characteristics of booksellers. The two different book subjects provide alternatives to test the stability of the model. The results show that there are significant spatial interactions among sellers' pricing strategies, and failing to account for that will lead to biased conclusions.

Keywords:

Spatial Autoregressive Model, Semiparametric Estimation, Internet Commerce, Price Dispersion

JEL Codes:

C01, C14, C21, M31

1. Introduction

The prediction that price competition and costless search in internet commerce would quickly reach perfect competition has been challenged in recent empirical literature on internet marketing and retailing. Persistent price dispersion has been observed in various internet retailing categories (Clay et al. 2001, 2002; Chevalier and Goolsbee, 2003; Baye et al. 2006), the cost advantage of online search (e.g., price comparison) does not seem to facilitate the realization of the "law of one price". As pointed out by Varian (1980) in his discussion of price dispersion, the challenge comes with the fact that price dispersion can persist in markets where at least some of shoppers behave rationally. This implies that consumers make their purchase decisions not only basing on price, but also other aspects of which sellers can take advantage to differentiate themselves and discriminate consumers. Pan et al. (2004) suggest that internet retailers can practice a random pricing strategy to discriminate against uninformed consumers but still occasionally attracting frequent shoppers. A retailer can create various non-price strategies to differentiate itself from other retailers with competitive prices, such as offering reviews, loyalty programs, more cost-efficient shipping option (e.g., the FREE Super Saver Shipping at Amazon.com), and other features facilitating online shopping experience. And non-price strategies do not necessarily related to product features. Clay et al. (2001) find that, Amazon.com is able to maintain a price premium relative to BarnesandNoble.com (BN.com) and Borders.com partially due to the fact that Amazon.com is successful in differentiating its books[1]. Chevalier and Goolsbee (2003) find that books at Amazon.com have a much lower price elasticity than BN.com. This also indicates that Amazon.com is better at differentiating its books with non-price strategies. Chevalier and Mayzlin (2006) suggest that the relative sales on a book between Amazon.com and BN.com are related to the differences in both the number of reviews and average star ranking that the book has received. Baye et al. (2007) provide a review documenting other research on online price dispersion[2].

Another possible explanation for failure of the search-price comparison-reduced price dispersion (i.e., the Frictionless Commerce) prediction is the existence of planned purchase. As docu-

---

[1]Barnes & Noble acquired Borders' trademarks and customer list in 2011.

[2]See Pan et al. (2004) for an early survey of analytical and empirical literature on online price dispersion.

mented in Charman-Anderson (2013) that: "nearly half of Amazon.com's book sales come from people who already know what they want and are simply using Amazon.com as a way to get it." And, "planned search by author or topic makes up a whopping 48 percent of all book choices at Amazon.com." This has broader implication than just about book prices. As found in Clay et al. (2001), Amazon.com does not necessarily offer competitive prices on books. It is possible that consumers choose shopping at Amazon.com for (new or used) books just because they value the quality of services which Amazon.com provides. For example, Amazon.com has a friendly return policy and an uniform and very organized listing format. These features can help to build outside awareness of Amazon.com and lead more intention to buy at Amazon.com, and which in turn attracts more potential booksellers to register and list at Amazon.com. Among other explanations, Ellison and Ellison (2009) suggest that retailers' engagement in obfuscation on price search can result in much less price sensitivity and potentially larger price dispersion in internet retailing. On the other hand, Brynjolfsson and Smith (2000) argue that the levels of price dispersion depends on how to measure prices and that dispersion in market share weighted prices is actually lower in internet channel than in conventional outlets.

Further, price dispersion is common to the asking prices of both heterogeneous goods and homogeneous goods (Stigler, 1961). Price dispersion does not exist only among different retailer websites, but also exists on one website and even on one webpage. A book, new or used (given its condition), is barely different from other copies listing for sale on the same website in terms of physical attributes. But why their sellers list them at different prices and these listing prices have no tendency to converge over time? This kind of price dispersion on the same website certainly can not be explained by the theory of searching cost. For example, the persistent price dispersion within the same condition category at Amazon.com (Figure 1) is frequently observed among popular books. Similar pattern can also be observed at other used book selling websites like Ebay.com and Half.com. The internet commerce has reduced the entry cost for some of the retailing business to almost zero, selling used books online is among which. As documented in Varian (2005), an estimated 23% of Amazon.com's sales are from used goods, and many of them are used books. However, the pricing strategies of these used books sellers who account for a popular and consider-

able business have not been well understood. Decomposing these strategies is the key to interpret the persistent price dispersion commonly observed in online used books retailing.

Amazon.com has established very clear and strict guideline on how to define book conditions for a seller to list used books for sale. Overstating an used book's condition or any other misleading information in the book condition notes section can cost the seller with forgone listing fee (as high as 15% for books, nonrefundable from Amazon.com if a book is sold) if returned and risk of getting negative seller ratings. Given used books' condition defined, booksellers basically face pricing decisions of nearly homogeneous goods to maximize long run profit. Therefore, they tend to seek other ways to justify their asking prices to potential buyers. To a large extent, the used book sellers are not differentiating the items they are selling, instead they are differentially themselves. In this paper, I argue it is sellers' self (service and non-service) differentiation which leads to price dispersion on used books. And any aspects where booksellers can signal their own creditability and the quality of their services can become drivers of price dispersion on used books. The paper is organized as following. Section 2 presents a discussion on the drivers of price dispersion in details. Measures of price dispersion and an empirical spatial model of price dispersion are described in Section 3. Data collection is briefly discussed in Section 4. Section 5 proposes a novel nonparametric estimation procedure based on Robinson (2010) to address issues with non-normal error terms. Estimation results are discussed in Section 6, and Section 7 concludes the paper.

2. Drivers of Price Dispersion

Internet used books retailing is certainly not a monopoly industry, while the Bertrand model predicts that all sellers will charge the same price and the price will fall to cost. It is clear that these standard economic theories fail to provide further insights to the drivers of observed price dispersion in internet used books retailing. Internet used book retailing is more like a contestable market (Baumol, 1982) rather than a perfectly competitive market, where entry is completely free and exit is also costless. Following Baumol (1982)'s characterization, contestability implies that there is no cost discrimination against entrants. A major exception to the contestability framework is that, the production (operational) cost of sellers is not revealed to each other in internet used books retailing.

5

Therefore, the observed price dispersion in Amazon.com used books listings does not persist as a consequence of competition structure. Instead, it is more likely derived from imperfect information and heterogeneity in sellers' cost structure. Another thing to note about internet used books retailing is that, sellers usually have little information on the demand side hence their pricing decisions do not necessarily have a strong supply-demand type structure ground. Instead sellers tend to make smaller price adjustment but more frequently than comparable offline retailers, which can be explained by smaller menu costs (Brynjolfsson and Smith, 2000) or the fact that sellers may be experimenting with different pricing and differentiation strategy combinations (Clay et al. 2002).

From a practical point of view, all drivers of price dispersion in internet retailing can be considered as violations of the Bertrand assumptions and the "law of one price", therefore motivation for either strategic pricing or simple markup pricing. The cost to a seller of ascertaining competitors' potential listing prices or reactions given his or her own price is a major source of price dispersion. The dispersion does not necessarily disappear even if sellers' cost of obtaining and assessing price information reduces to zero (Stigler, 1961). In Amazon.com used books listings, for example, some large professional sellers always automatically match their prices to the minimum price (or one cent lower) among other listings. Behavior like this limits the profitability of fierce price competition. Economic theory predicts that consumers' imperfect information about prices can lead to price dispersion. Uninformed consumers may not have access to relevant information on valuable characteristics of the goods, or it is just too expensive to attain and assess these information. Stigler (1961) suggests that advertising can help to deliver price information to consumers and establish certain creditability of sellers, which would narrow down the price dispersion. In Amazon.com used books retailing, however, the price information is costlessly available to consumers. And advertising (not including the general brand advertising on its retailing channel by Amazon.com) of any types, incurs overhead costs to sellers. Therefore, sellers may prefer not to competing directly on prices or by advertising their prices, instead they seek alternative ways to differentiate their listings (Clay et al., 2002).

Used books sellers at Amazon.com have two ways to signal the quality of services they provide with their books, condition notes and seller rating. Booksellers can use the condition notes space to

describe book conditions, highlight their service qualities, such as fast shipping, excellent customer service, and easy return policy. Listings with condition notes containing detail book description are more likely to be books with some writings, signs, or some degree of wearing/damage on. Clearly stating these conditions can reduce the probability of book return, while it may also lower the probability of getting sold. Seller rating is a more straightforward indicator of sellers' service quality to buyers. Amazon.com displays sellers' star rating and percentage of positive feedback within 12 months in the listings. It is natural to induce that sellers with higher seller rating have better chance to maintain a price premium if they want. Given all of these quality signals regarding either books or services, we can see that even in the same condition category used books are no longer perfect homogeneous products. This is particularly true if what buyers want is not only the lowest price available. Shipping option is another important non-price strategy which booksellers use to control operational cost and lower listing prices. Amazon.com has an uniform fee of $3.99 on standard domestic book shipping (via First Class Mail or Media Mail). Sellers can also use the shipping fulfillment provided by Amazon.com, which is usually employed by large professional sellers. If a book listing is affiliated with Amazon Fulfillment, then it is eligible for free shipping with at least $25 total purchase of eligible (for FREE Super Saver Shipping) products (not limited to books). Sellers will certainly be able to maintain a price premium if they can save shipping cost through Amazon Fulfillment. More expensive shipping options (e.g., expedited shipping) are also available at Amazon.com, but not likely to be frequently chosen.

Besides, sellers' risk taking ability can also affect their pricing decisions and be reflected into price dispersion. Even though Amazon.com has established clear guideline on used book condition definitions, it does not set any binding limits (the technical upper limit is $1,000,000,000.00) on the price level a seller can set. Booksellers have freedom to set their own prices which sometime may even not look rational. For example, in Amazon.com used book listings, overlapping of prices across different book condition categories is very common. In Figure 1, we can see that all four sellers in "Used - Good" category price their books higher than the seller #1 in "Used - Very Good" category. Other factors given, however, we expect that books in "Used - Good" are priced no higher than ones in "Used - Very Good". The seller #7 in "Used - Acceptable" even prices its book higher

7

than all books in "Used - Good" and "Used - Very Good". When a rational buyer pays higher price for a book of same condition as other copies (but lower prices), he or she would expect better services (e.g., fast shipping as some sellers signal in their condition notes, easier return process if unsatisfied with the book) are provided with the book. If a seller can not provide better services for the higher prices he asks, then he is at risk of bearing forgone listing fee and shipping cost. On the other hand, it might just be the seller #1's strategy to intentionally price lower to expect fast selling.

3. Empirical Models

3.1 Price Dispersion Measures

In the literature, price dispersion is generally considered as a result of market power and imperfect competition (Stigler, 1961; Brynjolfsson and Smith, 2000; Clay et al., 2001; Chevalier and Goolsbee, 2003). The empirical measures of price dispersion, however, are different from study to study. Price dispersion is most commonly measured by price range (max and min difference) and standard deviation (e.g., Brynjolfsson and Smith, 2000, Clay et al., 2002, Baye et al., 2007). Considering the heterogeneity among different goods, price dispersion can also be weighted with average or median prices. In Clay et al. (2002), price range and standard deviation are divided by average unit price. Brynjolfsson and Smith (2000) suggest that it is more reasonable to weight the prices by market shares in calculating price dispersion. In this paper, following the weighted price range measure the dependent variable is defined as:

$$Y_i = (P_i - \bar{P}_i) / \bar{P}_i \tag{1}$$

where $\bar{P}_i = \sum_{j \neq i}^{m} P_j / (m - 1)$, and $i = 1, ..., m$. P_i is the listing price of each used book i, and there are in total m listings within the same condition category under given book title. m varies by condition categories. Note that for each listing price P_i, the average price $\bar{P}_i$ from which it deviates is different because P_i itself is excluded from calculation. An implicit assumption here is that, when the seller of book i makes pricing decision all other listing prices are observed. This is a reasonable assumption under at least two cases: (1) the size of listings m is relatively large; (2) booksellers

frequently update their listing prices over time.

Even though internet used books selling is not a perfect competition market, direct price competition among sellers still co-exists with other non-price strategies. If there are significant spatial price competition among sellers, then non-spatial estimation methods may produce biased results. In this paper, a SAR model is proposed to model the pricing decision of booksellers. The goal is to identify important determinants of deviation in booksellers' pricing decision, hence have a deep understanding on what is driving the observed persistent price dispersion.

3.2 Spatial Autoregressive Models

The SAR models represent a set of spatial models depending on how the spatial weighting matrix is specified or estimated among endogenous variables, exogenous variables, and unobserved components, respectively. Among which, common ones are spatial Durbin model, spatial Durbin error model, SAR model, and spatial error model. Elhorst (2010a) provides a review on the relationships among different spatial dependence models for cross-sectional data. Depending on the context of problem being studied, three different types of spatial interaction effects can be used to explain why observed behavior or outcome associated with a specific location may be dependent on observations at other locations (Manski, 1993). The spatial locations here can be generalized to any irregularly spaced settings including the used books listings concerned in this study. Manski (1993) proposes a general model to capture these effects, which takes following form:

$$Y = \alpha \iota + \lambda W^Y Y + X\beta + W^X X\gamma + u \tag{2}$$

$$u = \rho W^u u + \varepsilon \tag{3}$$

where Y denotes a vector of observations on the dependent variable, ι is a vector of ones associated with the constant term parameter α, X denotes a matrix of k exogenous explanatory variables, with the associated parameters β and γ to be estimated. ε is a vector of disturbance terms which are usually assumed to be $i.i.d.$ with zero mean and variance σ^2. $W^Y Y$, $W^X X$, and $W^u u$ denote

9

the endogenous interaction among the dependent variables, the exogenous interaction among the independent variables, and the interaction effects among error components, respectively. Here W^Y, W^X, and W^u represent the spatial weighting matrices of the spatial units in the sample. λ is usually defined as the spatial autoregressive parameter, ρ the spatial autocorrelation coefficient. The model in (2) and (3) can fit with either cross-sectional data or panel data. If $\gamma = 0$ and $\rho = 0$, then the model in (2) and (3) reduces to a classical SAR model (which is also called spatial lag model in some cases):

$$Y = \alpha \iota + \lambda W^Y Y + X\beta + \varepsilon \tag{4}$$

SAR models have become popular in modeling both regularly and irregularly spaced data. In economics studies, irregularly spaced data is a more common context. The space involved is not limited to geographic settings, especially when it comes to economic and social behavior problems. In a cross-sectional setting (not necessarily regularly spaced), the interactions and interdependence among different units or individuals are often expected to be explained by some distance measures. In model (4), $\theta = [\lambda \; \beta^T]^T$ is often the focus of interest in estimation, and the spatial autoregressive parameter λ is conventionally confined to $(-1, 1)$. The matrix $[\iota \; X]$ is assumed to have full column rank in large sample.

The determination of spatial weighting matrix W^Y is always a practical matter. It is usually pre-specified exogenously basing on researchers' prior knowledge, or estimated in some ways. For example, Wang (2013) estimates the spatial weighting matrix in SAR model using a semiparametric instrument estimation. Qu and Lee (2012) propose a method of estimating the SAR model with an endogenous spatial weighting matrix. In this paper, the spatial weighting matrix W^Y is specified basing on the price competition structure among all booksellers. Further details on the specification of W^Y are discussed in Section 5.

4. Data

Internet bookselling has a history going back to 1980s. Some of the big players in this internet commerce today came into business in mid 1990s. Amazon.com, Barnes & Noble, and Borders went online in 1995, 1997, and 1998, respectively. Clay et al. (2001) provide a background review for internet bookselling business. The top 3 largest online booksellers in 2013 are Amazon.com, BN.com, and Powell's Books according to the ranking by TopTenREVIEWS.com[3]. This paper focuses on the largest online bookseller - Amazon.com, and its used books segment in particular. The data is collected from the online used book listings of Amazon.com in the week of July 2nd, 2012. The set of listings contains all of the top 100 bestselling econometrics books and the top 100 bestselling environmental economics books according to the sales ranking by Amazon.com. All books are uniquely identified by ISBN. Three parts of information are covered in the data: (1) condition category and price of used books listed; (2) bookseller's characteristics, including location, condition notes, size, availability of Amazon Fulfillment shipping, and availability of expedited shipping and international shipping options; (3) book information and book review score. Given the heterogeneity across different books, book fixed effects are used to control unobserved book characteristics in estimation. Therefore, all exogenous explanatory variables are constructed from booksellers' characteristics.

To eliminate the impacts from outliers, all listings with price higher than two times the lowest listing price of the new book with same ISBN are dropped. If the price of an used book is much higher than new books, then the used book price is not rational. To make sure there is enough price variation within each condition category, all condition categories (of a given book) with listing size less than five (i.e., $m < 5$) are excluded from the sample. For all duplicate listings (multiple copies of a book are listed by the same seller within a condition category), the listing with the lowest price is kept. Finally, we have two sub samples of 95 environmental economics books (3050 listings) and 80 econometrics books (1027 listings), respectively. Other than book characteristics, a major difference between two sub samples is the sample size.

Table 1 presents all of the variable definitions and summary statistics for two different samples.

---

[3]http://online-bookstores-review.toptenreviews.com/

Seller rating score is calculated using the percentage of positive feedback received within last 12 months, and is normalized between 0 and 1 due to the fact that most of original ratings are clustered around 90%. Seller size is approximated by the total number of seller ratings received within last 12 months. A larger bookseller is likely to have more books sold and therefore receive more buyer feedback and ratings. The seller size variable is also normalized to integer 0, 1, 2, 3, 4, as shown in Table 2. There are three dummy variables for shipping options. If the value is 1, it means the given shipping option is available from the seller. If a seller is willing to give specific and detail condition note on the listed book, it may reveal something about his or her pricing strategies. This is captured by the dummy variable for condition notes. New sellers are defined as ones with less than 50 seller ratings in their lifetime (not necessarily within 12 months). U.S. based sellers are defined as ones that ship books from a U.S. location (specified as a U.S. state name or general U.S. in the listing).

5. Estimation

When estimating the model in (4), the most popular approach is to build a Gaussian likelihood by introducing a normal distribution for the $i.i.d.$ error term ε. As pointed out by Robinson (2010), however, practitioners often lack a strong reason for choosing a particular parametric form for underlying distribution of error term ε. Mis-specification on ε can often lead to inconsistent estimates which are not necessarily more efficient than the Gaussian pseudo-ML estimates (Robinson, 2010). The mis-specification with small sample can produce even more serious estimation bias given that the sample sample properties of ML estimates is unknown.

An alternative, and more flexible approach is the adaptive estimation (e.g., Newey, 1988). The idea of adaptive estimation is to adapt the distribution form of ε, where the density function $f(\varepsilon)$ is nonparametric and estimated by some smoothing techniques. The key of adaptive estimation is the nonparametric estimation of score function $\psi(\cdot) = -f'(\cdot)/f(\cdot)$ instead of the density function $f(\cdot)$. One reason is that $f(\varepsilon)$ can become very small, which causes instability in estimating $f(\varepsilon)$ and then $\psi(\varepsilon)$ (Robinson, 2010). Beran (1976) proposes series estimation of $\psi(\cdot)$ in a time series autoregressive model, where $\psi(\cdot)$ is expanded by a sequence of functions (e.g., Fourier series). Let $\phi_l(\varepsilon), l = 1, 2, ..., L$, be a series of smooth functions, where L is a pre-chosen integer and increasing

slowly with sample size (Newey, 1988). Then the score function $\psi(\varepsilon)$ can be expressed as:

$$\psi(\varepsilon) = \bar{\phi}(\varepsilon)^T a \tag{5}$$

where $a = (a_1,...,a_L)^T$ is a vector of unknown expansion coefficients to be estimated. $\bar{\phi}(\varepsilon)$ is the vector of demeaned smooth functions, $\bar{\phi}(\varepsilon) = \phi(\varepsilon) - E\{\phi(\varepsilon_i)\}$, and $\phi(\varepsilon) = (\phi_1(\varepsilon),...,\phi_L(\varepsilon))^T$. Note that, by construction, we have $E\{\psi(\varepsilon_i)\} = 0$. Following Robinson (2010), a can be solved as:

$$a = \left[E\left\{\bar{\phi}(\varepsilon_i)\bar{\phi}(\varepsilon_i)^T\right\}\right]^{-1} E\left\{\phi'(\varepsilon_i)\right\} \tag{6}$$

Given a vector of observable ε, denoted as $\tilde{\varepsilon} = (\tilde{\varepsilon}_1,...,\tilde{\varepsilon}_n)^T$, then a can be empirically identified as:

$$\tilde{a} = \left[\frac{1}{n}\sum_{i=1}^{n}\Phi(\tilde{\varepsilon}_i)\Phi(\tilde{\varepsilon}_i)^T\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\phi'(\tilde{\varepsilon}_i)\right] \tag{7}$$

where $\Phi(\tilde{\varepsilon}_i) = \phi(\tilde{\varepsilon}_i) - \frac{1}{n}\sum_{j=1}^{n}\phi(\tilde{\varepsilon}_j)$. From (5) we can now empirically recover score function $\psi(\varepsilon_i)$:

$$\tilde{\psi}_i = \psi(\tilde{\varepsilon}_i, \tilde{a}(\tilde{\varepsilon})) = \Phi(\tilde{\varepsilon}_i)^T \tilde{a}(\tilde{\varepsilon}) \tag{8}$$

Further, following Robinson (2010), define the $n \times 1$ vector:

$$e(\theta) = (e_1(\theta),...,e_n(\theta))^T = (I - \lambda W^Y)Y - X\beta \tag{9}$$

and therefore,

$$\varepsilon = e(\theta) - E\{e(\theta)\} \tag{10}$$

Given an initial estimate $\tilde{\theta}$ (e.g., Gaussian ML estimate, various Instrument estimates) of θ, denote the $n \times 1$ vector $E(\theta)$ as the empirical proxy for ε, we have:

13

$$E(\tilde{\theta}) = e(\tilde{\theta}) - \iota_n \frac{1}{n} \sum_{i=1}^{n} e_i(\tilde{\theta}) \tag{11}$$

where ι_n is a vector of $n \times 1$ ones. Basing on (10) and (11), we can now estimate σ^2 as:

$$\tilde{\sigma}^2 = \tilde{\sigma}^2(\tilde{\theta}) = \frac{1}{n} E(\tilde{\theta})^T E(\tilde{\theta}) \tag{12}$$

Normalize ε to ε_0, where $\varepsilon_0 = \varepsilon/\sigma$, then $\tilde{\varepsilon}_0$ is given as:

$$\tilde{\varepsilon}_0 = \tilde{\varepsilon}/\tilde{\sigma} = E(\tilde{\theta})/\tilde{\sigma} \tag{13}$$

Combine (8) and (13), we can have:

$$\tilde{\psi}_i = \tilde{\psi}_i(\tilde{\theta}, \tilde{\sigma}) = \Phi(E_i(\tilde{\theta})/\tilde{\sigma})^T \tilde{a}(E(\tilde{\theta})/\tilde{\sigma}) \tag{14}$$

Now introduce the $(k+1) \times n$ matrix of derivatives of $e(\theta)$:

$$e'(\theta) = \left[\frac{\partial e(\theta)}{\partial \lambda}, \frac{\partial e(\theta)}{\partial \beta^T} \right]^T = \left[-W^Y Y, -X \right]^T \tag{15}$$

Denoting the $i-th$ column of $e'(\theta)$ as $e_i'(\theta)$ and defining following terms: $E_i'(\theta) = e_i'(\theta) - \frac{1}{n}\sum_{j=1}^{n} e_j'(\theta)$, $R = \sum_{i=1}^{n} E_i'(\theta)E_i'(\theta)^T$, $r(\theta, \sigma) = \sum_{i=1}^{n} \tilde{\psi}_i(\theta, \sigma)E_i'(\theta)$, with the information measure $\gamma = E\left\{ \psi^2(\varepsilon_i) \right\}$ being estimated as:

$$\tilde{\gamma}(\tilde{\theta}, \tilde{\sigma}) = \frac{1}{n} \sum_{i=1}^{n} \tilde{\psi}_i^2(\tilde{\theta}, \tilde{\sigma}) \tag{16}$$

Then the adaptive estimate of θ is given by (Robinson, 2010):

$$\hat{\theta}_A = \tilde{\theta} - \left[\tilde{\gamma}(\tilde{\theta}, \tilde{\sigma})R \right]^{-1} r(\tilde{\theta}, \tilde{\sigma}) \tag{17}$$

$$\sqrt{n}(\hat{\theta}_A - \theta) \xrightarrow{d} N(0, \frac{\sigma^2}{\gamma}\Omega^{-1}) \tag{18}$$

$$\Omega = \lim_{n \to \infty} \frac{1}{n} \begin{bmatrix} (GX\beta)^T \\ X^T \end{bmatrix} \begin{bmatrix} GX\beta & X \end{bmatrix} \tag{19}$$

where $G = W^Y(I - \lambda W^Y)^{-1}$, and the limit variance matrix of $\hat{\theta}_A$ is consistently estimated by $n(\tilde{\sigma}^2/\tilde{\gamma}(\tilde{\theta}, \tilde{\sigma}))R^{-1}$. The asymptotic normality and efficiency results are established in Robinson (2010). As pointed out by Robinson (2010), $\hat{\theta}_A$ achieves an efficiency improvement over Gaussian pseudo-ML estimates when ε_i is non-Gaussian. Further iteration of (17) can improve higher-order efficiency, as shown in Robinson (1988), and may be desirable in the context of practice. A special case to note is that, the above estimation procedure can not estimate λ in pure SAR models without any explanatory variables (Lee, 2004). The initial estimate employed in this paper, $\tilde{\theta}$ in (17), is the concentrated ML-estimate following from Elhorst (2010b) but reduced to cross-sectional case. The smooth functions $\phi(\cdot)$ in series expansion (in Equation (5)) use polynomial forms with $\phi_l(\varepsilon) = \varepsilon^l$, as in Newey (1988), Pinkse et al. (2002), and Robinson (2005, 2010).

The estimation results in (17) and (18) also rely on some restrictive assumptions imposed on the weighting matrix. The spatial weighting matrix W^Y should be a non-negative matrix of known constants, and the diagonal elements are zero (Lee, 2004). W^Y should also satisfy one of the following two conditions: (1) the row and column sums of the matrices, W^Y and $(I - \lambda W^Y)^{-1}$ without row normalizing (standardizing) W^Y should be uniformly bounded in absolute value as sample size n goes to infinity (Kelejian and Prucha, 1999); (2) the row and column sums of W^Y without row normalizing should not diverge to infinity at a rate equal to or faster than the rate of n (Lee, 2004). Robinson (2010) employs similar assumptions of uniform boundedness on W^Y and $(I - \lambda W^Y)^{-1}$.

In this paper, the spatial weighting matrix W^Y is specified as a symmetric block diagonal matrix. For a given book and condition combination which is defined as a group, the pricing of each listing equally depends on all other listings within the group. More precisely, for listing i and j where $i \neq j$, the corresponding element $w_{ij}(= w_{ji})$ in W^Y equals to 1 if i and j are within the same group. Otherwise, $w_{ij} = 0$. Therefore, the above assumptions imposed on W^Y are satisfied since W^Y is defined as a sparse matrix (i.e., each spatial unit has only a limited number of neighbours), and

the number of units (listings) in each block (book-condition group) is limited and not necessarily increasing with the sample size. Here the sample size increases mainly with the number of books included, and it is assumed that no spatial interactions exist among different books. Allowing for interactions among different books will certainly make a realistic direction for further research to explore. One thing to note on the W^Y employed in this paper is that, each group does not contain the same number of units (listings), which is different from Lee (2002) and Robinson (2010). Following the convention in the literature, W^Y is row normalized in the estimation procedure of this paper.

A key practical issue in nonparametric adaptive estimation is the choice of L, the number of smooth functions used in the series expansion. In this paper, the estimate of λ, the first element of θ, is found to be very sensitive to the choice of L. When L is relatively small, the adaptive estimator in (17) tends to overestimate λ. As the spatial autoregressive parameter, λ is confined to $[-1, 1]$. When L is chosen less than 10, however, estimate of λ goes below -1 with the data in this paper. For estimation with both samples in this paper, L is set to 25, around which estimates of L tend to be stable.

6. Results and Discussion

Two main concerns have been raised when estimating a parametric price dispersion model like $Y = \alpha \iota + X\beta + \varepsilon$ (i.e., the model in (4) with $\lambda = 0$) using least squares methods. The first concern is that non-spatial methods may lead to biased estimates when there are significant spatial interactions (price competition in this paper) among different cross-sectional units. Therefore, a SAR model is proposed to account for the spatial price competition among booksellers in this paper. Another concern is that unobserved effects in the model may violate commonly imposed Gaussian assumption on error terms, which can further lead to biased ML estimates. In this paper, an efficient semiparametric estimation is implemented on the SAR model and it does not require the Gaussianity assumption.

Table 3 and Table 4 presents ordinary least squares (OLS) estimates and fixed effects model estimates for two different samples. Basing on the results from these benchmark models, several

general observations can be summarized. (1) Sellers with higher ratings tend to price used books lower. Note that, it is possible that a higher rating score is partially due to the fact that sellers have been offering low prices. This leads to additional concern for endogeneity issues, but an instrument for seller ratings is not easy to find given the limited information observed on booksellers. (2) In both samples, larger sellers have significantly lower listing prices. This can be intuitively explained, for example, by the fact that larger sellers can achieve scale of economy in warehousing, shipping, and supply chain management in general. (3) As we expected, sellers who are willing to list specific and detail book condition notes are adopting low price strategy. This may just reveal the information that they are selling books with relatively less satisfied conditions compared to other books in the same condition category. Low price can potentially lead to fast selling and lower probability of getting book returned. (4) Obviously, sellers who are affiliated with Amazon Fulfillment shipping are able to price higher due to the saving from shipping expense. (5) Newly established sellers are willing to price lower, which indicates that they are seeking ways to build their reputation (i.e., getting more rating and positive feedback). Variables "exp_ship", "int_ship", and "us_seller" are expected to tell us something about sellers' risk taking abilities. "us_seller" is the only variable that has a consistent negative estimate between two samples. It implies that an international seller tends to demand a price premium to compensate the higher risk (e.g., high probability of package damage) involved. "exp_ship" and "int_ship" flip signs between two samples, which indicates they are probably not good proxies for seller's risk taking abilities.

Table 5 shows the ML estimates using the concentrated ML-estimation following from Elhorst (2010b). Left panel and right panel are for two sub samples, respectively. For each panel, left part reports results without book fixed effects, and right panel with book fixed effects. The ML estimates of SAR models are supposed to address issues with ignoring the spatial price competition effects. By comparing Table 3 and Table 4 with Table 5, the difference in $\hat{\theta}$ is pretty significant. However, the direct estimates of SAR model do not represent the true marginal effects of explanatory variables, which is further discussed below. Table 6 shows the semiparametric adaptive estimates of SAR model using the method proposed by Robinson (2010), and the results are reported in the same format as Table 5. The semiparametric adaptive estimation is supposed to address our second

concern regarding non-Gaussian error terms due to unobserved effects. Comparing between Table 6 and Table 5, we can see the significant correction the semiparametric adaptive estimation has made upon ML estimation, especially in the case of econometrics book sample where the sample size is relatively small. In the environmental economics book sample where sample size is larger, we can see that the ML estimates are close to the semiparametric adaptive estimates. Overall, the same conclusions regarding booksellers' pricing strategies still hold, but the magnitudes change as the estimation procedure gets improved.

Before having any quantitative interpretation on the estimates, as emphasized by LeSage and Pace (2009), it is necessary to obtaining an unbiased measure of the marginal effect of an explanatory variable in SAR models. As shown in Elhorst (2010a), the true marginal effect does not equal to the coefficient estimate of given explanatory variable. The true marginal effect should take into account both the direct effect and the indirect effects through the spatial interactions and simultaneous feedback in the SAR model. Basing on the model in (4), for the k_{th} explanatory variable in X, it follows that

$$\frac{\partial Y}{\partial X_k} = \beta_k (I - \lambda W^Y)^{-1} \tag{20}$$

The total marginal effects can be then obtained by taking average of all cross-section units:

$$n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial Y_i}{\partial X_{jk}} = \beta_k n^{-1} \left[\iota_n'(I - \lambda W^Y)^{-1} \iota_n \right] \tag{21}$$

where ι_n is an $n \times 1$ vector of ones as before. The direct effect is intuitively given by only averaging over the diagonal elements:

$$n^{-1} \sum_{i=1}^{n} \frac{\partial Y_i}{\partial X_{ik}} = \beta_k n^{-1} trace \left[(I - \lambda W^Y)^{-1} \right] \tag{22}$$

As shown in LeSage and Pace (2009), this direct effect is different from the coefficient estimate $\hat{\beta}_k$ because of the spatial autoregression process among dependent variables observed from different cross-section units. Table 7 presents the direct, indirect, and total marginal effects basing on the estimates using the proposed semiparametric adaptive estimation of SAR model with book fixed

effects (Table 6).

Taken two samples together, total marginal effects are generally smaller than direct effects (closer to original estimates $\hat{\theta}$) due to the significant indirect effects through spatial interactions. In environmental econ book sample, seller rating has a very large negative effects. If the seller rating is increased by 0.1 out of 1, the listing price is about 1.2% lower on average. The same effect is only a little over 0.1% in econometrics book sample. If the seller size goes up by one level as defined in Table 2, the listing price goes down somewhere in the range of 2% - 4% on average. Providing specific condition notes is associated with a price decrease effect of 4% - 7% on average, while affiliating with Amazon Fulfillment Super Free shipping program guarantees a price premium of 4% - 13% on average depending on the sample. New sellers on average price 6% - 9% lower than other sellers. U.S. based sellers also tend to price lower, on average between 1% - 6% depending on the sample. The results on the availability of expedited shipping and international shipping options are not conclusive between two samples.

7. Concluding Remarks

This paper has employed publicly available data on used books price and booksellers' characteristics at Amazon.com to estimate a spatial econometrics model of price dispersion. The results reveal very interesting linkages among booksellers' pricing strategy, non-pricing strategy, and characteristics. In general, it has been found that sellers with good reputation and large sales volume, newly established sellers, and U.S. based sellers (compared to international sellers) tend to price lower holding other factors same. An interesting negative relationship between the richness of condition notes and listing prices has been identified. And affiliation with Amazon Fulfillment Super Free shipping program guarantees a large price premium for sellers, which is likely due to the saving from shipping expense as we can expect. However, a solid conclusion can not be drawn without a clear understanding of the cost structure of the affiliated booksellers. On the other hand, better measures for booksellers' risk taking ability is necessary to improve the model estimated in this paper, which points to a fruitful direction of future research.

There are also some other worthwhile issues to be explored in further research. For example,

the data used in this paper is collected in a cross-sectional setting, while a dynamic understanding of booksellers' pricing strategy is definitely much more interesting and valuable to both business practitioners and the literature. This calls for a panel data set being collected over time. On the methodological side, as we have realized, the endogeneity issues (e.g, the seller rating variable) can potentially bias estimation results. To search for proper instruments for endogenous variables, we need to understand the behavior of internet booksellers well. Lastly, the model in this paper does not account for the demand side of used books. Of course, a structural model with both supply side and demand side will provide a better way to understand the online used books market.

Reference

1. Baumol, W. J. 1982. "Contestable Markets: An Uprising in the Theory of Industry Structure." American Economic Review 72(1): 1-15.

2. Baye, M. R., J. Morgan, and P. Scholten. 2006. Persistent Price Dispersion in Online Markets. In Jansen, D. W. (Eds) The New Economy And Beyond: Past, Present And Future: 122-143. Edward Elgar Publishing, Cheltenham, U.K.

3. Baye, M. R., J. Morgan, and P. Scholten. 2007. "Information, Search, and Price Dispersion." in Handbook of Economics and Systems (ed. by T. Hendershott), Amsterdam: North-Holland.

4. Brynjolfsson, E., and M. D. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." Management Science 46: 563-585.

5. Charman-Anderson, S. 2013. "Half of Amazon Book Sales are Planned Purchases." Forbes at <http://www.forbes.com>, February 20, 2013.

6. Clay, K., R. Krishnan, and E. Wolff. 2001. "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry." The Journal of Industrial Economics 49: 521-539.

7. Clay, K., R. Krishnan, E. Wolff, and D. Fernandes. 2002. "Retail Strategies on the Web: Price and Non-price Competition in the Online Book Industry." The Journal of Industrial Economics 50: 351-367.

8. Elhorst, J. P. 2010a. "Applied Spatial Econometrics: Raising the Bar." Spatial Economic Analysis 5(1): 9-28.

9. Elhorst, J. P. 2010b. Spatial Panel Data Models. In Fischer, M. M., Getis, A. (Eds.) Handbook of Applied Spatial Analysis: 377-407. Springer: Berlin Heidelberg, New York, U.S.A.

10. Ellison, G., and S. F. Ellison. 2009. "Search, obfuscation, and price elasticities on the internet." Econometrica 77(2): 427-452.

11. Goolsbee, A., and J. Chevalier. 2003. "Measuring Prices and Price Competition Online: Amazon vs. Barnes and Noble." Quantitative Marketing and Economics 1: 203-222.

12. Kelejian, H. H., and I. R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." International Economic Review 10(2): 509-533.

13. Lee, L. F. 2002. "Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models." Econometric Theory 18: 252-277.

14. Lee, L. F. 2004. "Asymptotic Distributions of Quasi-Maximum Likelihood Estimates for Spatial Autoregressive Models." Econometrica 72: 1899-1925.

15. LeSage, J., and R. K. Pace. 2009. Introduction to Spatial Econometrics. Boca Raton, FL: CRC Press.

16. Manski, C. F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." Review of Economic Studies 60(3): 531-42.

17. Manski, C. F. 2000. "Economic Analysis of Social Interactions." Journal of Economic Perspectives 14(3): 115-136.

18. Newey, W. K. 1988. "Adaptive Estimation of Regression Models via Moment Restrictions." Journal of Econometrics 38: 301-339.

19. Pan, X., B. T. Ratchford, and V. Shankar. 2004. "Price dispersion on the internet: a review and directions for future research." Journal of Interactive Marketing 18(4): 116-135.

20. Pinkse, J., M. E. Slade., and C. Brett. 2002. "Spatial Price Competition: A Semiparametric Approach." Econometrica 70(3): 1111-1153.

21. Qu, X., and L. F. Fee. 2012. "Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix." Working Paper at Department of Economics, The Ohio State University.

22. Robinson, P. M. 1988. "The Stochastic Difference Between Econometric Statistics." Econometrica 56: 531-548.

23. Robinson, P. M. 2005. "Efficiency Improvements in Inference on Stationary and Nonstationary Fractional Time Series." The Annals of Statistics 33: 1800-1842.

24. Robinson, P. M. 2010. "Efficient Estimation of the Semiparametric Spatial Autoregressive Model." Journal of Econometrics 157: 6-17.

25. Stigler, G. J. 1961. "The Economics of Information." The Journal of Political Economy 69: 213-225.

26. Varian, H. R. 1980. "A Model of Sales." The American Economic Review 70(4): 651-659.

27. Varian, H. R. 2005. "Reading Between the Lines of Used Book Sales." The New York Times at <http://www.nytimes.com>, July 28, 2005.

28. Wang, H. 2013. "A Semiparametric Analysis of the Spatial Structure of Farmland Values." The 52nd Southern Regional Science Association (SRSA) Annual Meeting, April 4-6, 2013. Metropolitan Washington, D. C.

Figures

$6.75	Used – Very Good	Seller: **#1**
+ $3.99 shipping	Minor wear, solid binding, ships within 24 hours	Seller Rating: ★★★★★ **100% positive** over the past 12 months. (4,444 total ratings)
		In Stock. Ships from MA, United States. Expedited shipping available. Domestic shipping rates and return policy.
$6.95	Used – Good	Seller: **#2**
+ $3.99 shipping	Buy with confidence. Excellent Customer Service & Return policy. We ship Daily including w/ends.	Seller Rating: ★★★★☆ **94% positive** over the past 12 months. (26,658 total ratings) In Stock. Ships from TX, United States. International & domestic shipping rates and return policy.
$7.49	Used – Good	Seller: **#3**
+ $3.99 shipping	Minor pencil markings on a few pages	Seller Rating: ★★★★★ **97% positive.** (211 total ratings) In Stock. Ships from NV, United States. Domestic shipping rates and return policy.
$7.89	Used – Good	Seller: **#4**
+ $3.99 shipping	Used but in good condition with some markings. No missing pages. May not include accessories such as CD, code or DJ. May be E… » Read more	Seller Rating: ★★★★★ **95% positive** over the past 12 months. (253,313 total ratings) In Stock. Ships from United States. Domestic shipping rates and return policy.
$7.95	Used – Good	Seller: **#5**
+ $3.99 shipping	Some highlighting, but otherwise the book is in good condition.. Will ship the next business Day.	Seller Rating: ★★★★★ **98% positive.** (41 total ratings) In Stock. Ships from CT, United States. Expedited shipping available. International & domestic shipping rates and return policy.
$7.99	Used – Very Good	Seller: **#6**
+ $3.99 shipping	Very good condition, paperback. Binding is intact. Clean read copy. Pages also in tact, and haven't any ear-markings or highl… » Read more	Seller Rating: ★★★★★ **100% positive** over the past 12 months. (50 total ratings) In Stock. Ships from CA, United States. Domestic shipping rates and return policy.
$8.00	Used – Acceptable	Seller: **#7**
+ $3.99 shipping	A number of underlinings and margin notes	Seller Rating: ★★★★★ **99% positive.** (204 total ratings) In Stock. Ships from MA, United States. International & domestic shipping rates and return policy.

Figure 1: A Sample of Amazon.com Used Books Listings

Tables

Table 1: Variable Definition and Summary Statistics

Variable	Definition	Obs	Mean	Min	Max	Obs	Mean	Min	Max
			Econometrics Sample			Environmental Econ Sample			
pd	dependent variable price deviation Y_i	1027	0.0197	-0.9238	3.2665	3050	0.0136	-0.9986	3.4458
rating	normalized seller rating score	1027	0.9236	0.0000	1.0000	3050	0.9117	0.0000	1.0000
size	normalized seller size	1027	1.9961	0.0000	4.0000	3050	1.7911	0.0000	4.0000
exp_ship	dummy for expedited shipping	1027	0.2726	0.0000	1.0000	3050	0.4882	0.0000	1.0000
int_ship	dummy for international shipping	1027	0.4382	0.0000	1.0000	3050	0.3423	0.0000	1.0000
note	dummy for detail condition notes	1027	0.2269	0.0000	1.0000	3050	0.4951	0.0000	1.0000
amazon	dummy for Amazon Fulfillment shipping	1027	0.0351	0.0000	1.0000	3050	0.0639	0.0000	1.0000
new_seller	dummy for new sellers	1027	0.1383	0.0000	1.0000	3050	0.1639	0.0000	1.0000
us_seller	dummy for U.S. based sellers	1027	0.8734	0.0000	1.0000	3050	0.9826	0.0000	1.0000

Table 2: Normalized Bookseller Size

Normalized Seller Size	Number of Ratings Received in 12 months
0	$[0, 100)$
1	$[100, 1000)$
2	$[1000, 10000)$
3	$[10000, 100000)$
4	$[100000, +\infty)$

Table 3: OLS and Fixed Effects Model Estimation Results (Econometrics Books)

Model	OLS		Fixed Effects (1)		Fixed Effects (2)	
Variable	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
rating	-0.0102	0.1995	-0.0251	0.2142	0.0540	0.2295
size	-0.0403***	0.0160	-0.0385**	0.0174	-0.0411**	0.0180
exp_ship	0.1441***	0.0330	0.1360***	0.0359	0.1284***	0.0376
int_ship	0.1091***	0.0323	0.1187***	0.0351	0.1323***	0.0366
note	-0.1070***	0.0331	-0.1466***	0.0380	-0.1730***	0.0415
amazon	0.0734	0.0745	0.0655	0.0796	0.0320	0.0863
new_seller	-0.1358***	0.0527	-0.1583***	0.0573	-0.1678***	0.0599
us_seller	-0.0018	0.0475	-0.0169	0.0523	-0.0146	0.0543
book fixed effects			Yes			
group fixed effects					Yes	
obs	1027		1027		1027	
R^2	0.06		0.08		0.08	

1. A group is defined as book and condition category combination;

2. Asterisks (*,**,***) indicate that the estimates are significantly different from zero at 10%, 5%, and 1% confidence level, respectively.

Table 4: OLS and Fixed Effects Model Estimation Results (Environmental Econ Books)

Model	OLS		Fixed Effects (1)		Fixed Effects (2)	
Variable	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
rating	-0.2417***	0.0932	-0.2574***	0.0963	-0.2647***	0.1004
size	-0.0869***	0.0094	-0.0894***	0.0100	-0.0957***	0.0104
exp_ship	-0.0981***	0.0190	-0.1019***	0.0199	-0.1104***	0.0208
int_ship	-0.0645***	0.0199	-0.0667***	0.0208	-0.0653***	0.0219
note	-0.0842***	0.0175	-0.0865***	0.0185	-0.1036***	0.0200
amazon	0.2715***	0.0364	0.2810***	0.0380	0.2932***	0.0398
new_seller	-0.1973***	0.0313	-0.2020***	0.0325	-0.1985***	0.0344
us_seller	-0.1298**	0.0665	-0.1526**	0.0718	-0.1655**	0.0752
book fixed effects			Yes			
group fixed effects					Yes	
obs	3050		3050		3050	
R^2	0.08		0.09		0.19	

1. A group is defined as book and condition category combination;

2. Asterisks (*,**,***) indicate that the estimates are significantly different from zero at 10%, 5%, and 1% confidence level, respectively.

Table 5: SAR Maximum Likelihood Estimation Results

Sample	Econometrics Books				Environmental Econ Books			
Variable	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
rating	-0.0228	0.1734	-0.0313	0.1773	-0.2210***	0.0883	-0.2349***	0.0896
size	-0.0378***	0.0139	-0.0341***	0.0144	-0.0816***	0.0088	-0.0834***	0.0092
exp_ship	0.1326***	0.0287	0.1186***	0.0297	-0.0897***	0.0178	-0.0928***	0.0183
int_ship	0.0984***	0.0280	0.1037***	0.0290	-0.0613***	0.0186	-0.0623***	0.0190
note	-0.0768***	0.0288	-0.1235***	0.0315	-0.0728***	0.0164	-0.0752***	0.0169
amazon	0.0680	0.0647	0.0617	0.0658	0.2557***	0.0341	0.2641***	0.0349
new_seller	-0.1005**	0.0458	-0.1237***	0.0474	-0.1835***	0.0294	-0.1850***	0.0298
us_seller	0.0062	0.0413	-0.0156	0.0433	-0.1155*	0.0629	-0.1271**	0.0647
book fixed effects			Yes				Yes	
obs	1027		1027		3050		3050	
λ(s.e.)	-0.9940 (0.0332)		-0.9980 (0.0340)		-0.9820 (0.0315)		-0.9830 (0.0306)	
log-likelihood	-457.5261		-443.0281		-1863.1000		-1855.3000	

1. Asterisks (*,**,***) indicate that the estimates are significantly different from zero at 10%, 5%, and 1% confidence level, respectively.

Table 6: SAR Semiparametric Adaptive Estimation Results

Sample	Econometrics Books				Environmental Econ Books			
Variable	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
rating	-0.0305***	0.0017	-0.0288***	0.0042	-0.2209***	0.0004	-0.2342***	0.0002
size	-0.0373***	0.0005	-0.0340***	0.0005	-0.0813***	0.0000	-0.0833***	0.0000
exp_ship	0.1371***	0.0010	0.1217***	0.0012	-0.0901***	0.0001	-0.0930***	0.0001
int_ship	0.0960***	0.0010	0.1092***	0.0011	-0.0614***	0.0001	-0.0625***	0.0001
note	-0.0803***	0.0010	-0.1327***	0.0012	-0.0726***	0.0001	-0.0752***	0.0001
amazon	0.0673***	0.0023	0.0745***	0.0025	0.2553***	0.0002	0.2636***	0.0001
new_seller	-0.1041***	0.0015	-0.1286***	0.0018	-0.1827***	0.0001	-0.1850***	0.0001
us_seller	-0.0001	0.0014	-0.0068***	0.0016	-0.1155***	0.0003	-0.1269***	0.0002
book fixed effects			Yes				Yes	
obs	1027		1027		3050		3050	
λ(s.e.)	-0.9778 (0.0181)		-0.9858 (0.0223)		-0.9742 (0.0012)		-0.9831 (0.0008)	

1. Asterisks (*,**,***) indicate that the estimates are significantly different from zero at 10%, 5%, and 1% confidence level, respectively.

Table 7: Marginal Effects Basing on Semiparametric Adaptive Estimation with Book Fixed Effects

Sample	Econometrics Books				Environmental Econ Books			
Variable	$\hat{\theta}$	Total Effects	Direct Effects	Indirect Effects	$\hat{\theta}$	Total Effects	Direct Effects	Indirect Effects
rating	-0.0288***	-0.0145	-0.0311	0.0166	-0.2342***	-0.1181	-0.2456	0.1275
size	-0.0340***	-0.0171	-0.0369	0.0198	-0.0833***	-0.0420	-0.0874	0.0454
exp_ship	0.1217***	0.0613	0.1317	-0.0704	-0.0930***	-0.0469	-0.0975	0.0506
int_ship	0.1092***	0.0550	0.1182	-0.0632	-0.0625***	-0.0315	-0.0655	0.0340
note	-0.1327***	-0.0668	-0.1437	0.0769	-0.0752***	-0.0379	-0.0789	0.0410
amazon	0.0745***	0.0375	0.0807	-0.0432	0.2636***	0.1329	0.2764	-0.1435
new_seller	-0.1286***	-0.0648	-0.1392	0.0744	-0.1850***	-0.0933	-0.1941	0.1008
us_seller	-0.0068***	-0.0034	-0.0073	0.0039	-0.1269***	-0.0640	-0.1331	0.0691

1. Asterisks (*,**,***) indicate that the estimates are significantly different from zero at 10%, 5%, and 1% confidence level, respectively.