NORTHWESTERN UNIVERSITY

Essays on E-Commerce and Consumer Demand

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Ting Wang

EVANSTON, ILLINOIS

June 2023

ABSTRACT

Essays on E-Commerce and Consumer Demand

Ting Wang

This dissertation consists of three chapters on theoretical and empirical industrial organization. The first chapter highlights a previously unnoticed property of commonly-used discrete choice models, which is that they feature parallel demand curves. The second chapter studies how a behavioral consumer preference with "price reference effect" can overturn the standard intuition of vertical integration. The last chapter identifies the effect of improved product information on demand and prices in the context of an online retail market.

Chapter 1 (which is joint work with Kory Kroft, René Leal-Vizcaíno, and Matthew J. Notowidigdo) highlights a previously unnoticed property of commonly-used discrete choice models, which is that they feature parallel demand curves. Specifically, we show that in additive random utility models, inverse aggregate demand curves shift in parallel with respect to variety if and only if the random utility shocks follow the Gumbel (Type 1 Extreme Value) distribution. Using results from Extreme Value Theory, we provide conditions for other distributions to generate parallel demands asymptotically, as the

number of varieties increases. We establish these results in the benchmark case of symmetric products, illustrate them using numerical simulations and show that they hold in extended versions of the model with correlated tastes and asymmetric products. Lastly, we provide a "proof of concept" of parallel demands as an economic tool by showing how to use parallel demands to identify the change in consumer surplus from an exogenous change in product variety.

In many settings, behavioral economists have documented a price reference effect: the fact that a consumer's willingness to pay for a good is affected by difference between the observed price and the reference price they rationally expect. In Chapter 2 (which is joint work with Junyan Guan), we first show theoretically that when this price reference effect is sufficiently large, it can overturn the standard textbook result that vertical integration improves joint profits. The key intuition is that the increase in quantity is dampened when consumers update their expectations. To test whether this force is large in a real-world setting, we develop a model of a downstream retailer who faces behavioral consumers and bargains with an upstream producer. We estimate this model using a novel dataset from a large online book retailer, where we observed retail prices, quantities sold and wholesale prices. Counterfactual simulations show that vertical integration would reduce joint profits by 11%. These findings highlight the importance of incorporating consumer expectations in the analysis of optimal pricing and vertical integration.

In E-Commerce, the shopping process is accompanied by numerous information of a single product. Researchers in Economics and Marketing have documented experimental evidence of how information affects consumers' behaviors. In Chapter 3, I first show a model with demand uncertainty and predict that the probability of returning products is decreased when consumers have less biased prior beliefs about product quality. To provide empirical evidence on this prediction, I use a novel data set from a large book retailer to identify the changes in return rates from an exogenous shock in the level of information provision. Estimation results suggest that the increased amount of information reduces return rates by 18%-19% and it is mainly driven by less popular books.

Acknowledgements

I would like to express my deepest gratitude to my advisors, Robert Porter, Gastón Illanes, and Matthew J. Notowidigdo for their invaluable advice, continuous support, and patience throughout the completion of this work. I would also like to extend my sincere thanks to Mar Reguant, Vivek Bhattacharya, David Dranove, and Kory Kroft for all their help and advice during my PhD study. Many thanks to my coauthors and all seminar participants at Northwestern University.

Thanks should also go to my friends Yi Qian Zhang and Bindan Zhang, as well as Ava Carn-Watkins. Without their tremendous understanding and encouragement, this endeavor would not have been possible. I had the pleasure of spending the past few years with my peers and many others in Evanston.

Finally, I would like to thank my parents for their unconditional love and unfailing support.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Parallel Inverse Aggregate Demand Curves in Discrete Choice Models

1.1. Introduction

This paper shows that some commonly-used discrete choice models satisfy a parallel inverse aggregate demand property – hereafter referred to as "parallel demands". Specifically, inverse aggregate demand curves shift vertically in parallel in response to an exogenous change in the number of varieties in a market. In this paper we show that this property holds for the Logit model and some of its Generalized Extreme Value (GEV) distribution variants. In additive random utility models (ARUM) featuring i.i.d. random utility shocks, this means that the random utility shocks are distributed according to the Gumbel (Type 1 Extreme Value) distribution. In fact, we show that the Gumbel distribution is both a necessary and sufficient condition for parallel demands in random utility models. As far as we know, this is a previously-unnoticed feature of this class of models, and as a result this paper focuses on characterizing this property theoretically and showing how it can be used in an economic application to identify the change in consumer surplus associated with an exogenous change in product variety.

In order to develop and build intuition, section 2 considers an additive random utility model with symmetric products and prices and an outside option. Theorem 1 establishes that the Gumbel distribution is necessary and sufficient for parallel demands. Next, we show that for a broad set of distributions of the random utility shock, inverse aggregate demand curves are asymptotically parallel; that is, the aggregate demand curves approach parallel demands as the number of varieties increases (Theorem 2). This result comes directly from Extreme Value Theory (EVT): when the random utility shocks are independent and identically distributed, the distribution of the maximum order statistic converges to a Gumbel distribution for a wide range of distributions. This means that assuming parallel demands may be a useful approximation in many markets featuring a large number of varieties. We illustrate the accuracy of this approximation result using numerical simulations, and we find that convergence happens fairly quickly.

In section 3, we extend the results in Theorems 1 and 2 in several ways. First, we extend the baseline model to allow for correlated tastes, which allows for differential substitutability within the market that has product variety, relative to the outside option. This extension allows us to accommodate the standard Nested Logit model as a special case (Cardell, 1997; McFadden, 1978). We show that in this extended model the Gumbel distribution is necessary and sufficient for parallel demands (Proposition 1). Second, we extend our results to allow for asymmetric products, since our baseline model assumes symmetric products and prices for simplicity. This extension allows us to accommodate a random utility model with unobserved product heterogeneity as in Berry (1994). The inverse aggregate demand curve is straightforward to define in the symmetric products model. When prices are asymmetric, however, we instead rely on the distribution of the maximal willingness-to-pay for any of the available varieties rather than the aggregate demand curve, and we provide necessary and sufficient conditions for when this distribution shifts in parallel, just as the inverse aggregate demand curve shifts in parallel in our baseline symmetric products model (Theorem 3). Theorem 4 extends the the asymptotic result of Theorem 2 to the asymmetric case.

Lastly, in section 4, we show how to use the parallel demands property to identify the change in consumer surplus from an exogenous change in variety. In our baseline model with symmetric products, graphically the change in consumer surplus is the area between the inverse aggregate demand curves before and after a change in variety. Thus, the change in consumer surplus – what we call the "variety effect" – is the area between these curves. Intuitively, a key feature of the parallel demands property is that identifying the "vertical gap" between the two inverse aggregate demand curves (at two different variety levels) at any one location on the demand curve is sufficient to identify the full area between the two demand curves. Proposition 3 provides a graphical representation of the identification of this vertical gap under parallel demands. It shows that several parameters are sufficient to calculate the variety effect. First, one needs to identify the sensitivity of demand to price, holding variety fixed. Second, one needs to identify the change in price and output in response to an exogenous change in variety. Jointly, under parallel demands, these two parameters are sufficient to identify the change in consumer surplus. Thus, the parallel demands property – which has a rigorous microfoundation based on the theoretical results in this paper – can be used to identify the change in consumer surplus stemming from a change in variety.[1] We next extend these results to cover the case of asymmetric products. When products are heterogeneous, we require an additional technical assumption that prices move uniformly after a change in variety. We

[1]One might speculate that since the assumed parallel shift in " aggregate demand" in an ARUM model amounts to assuming Logit demand, it is more direct to compute the effect on consumer surplus using the utility function directly. However, our results show that parallel demands are a good approximation for a larger set of distributions of the random utility shock beyond Logit.

show that under this assumption, a similar set of parameters identify the variety effect (Proposition 5). Since our approach to identifying changes in consumer surplus is based on aggregate demand, it is perhaps not surprising that we obtain identification by either assuming symmetric products or correlated prices – these are precisely the two scenarios highlighted in Nevo (2011) when discussing identification of aggregate demand and the problem of dimensionality.

This paper contributes to research that explores the theoretical properties of discrete choice models and the theoretical connections between these models and other economic properties. Perhaps most closely related to this paper is S. P. Anderson & Bedre-Defolie (2019) who consider a multi-product monopolist who chooses variety and price. They show that for asymmetric Multinominal Logit demand, the inverse demand shifts in parallel when the total variety increases and use this property to show that the monopolist chooses the socially optimal variety for a given total quantity. In terms of Spence's analysis of optimal quality provision (here phrased as product line length), the average and marginal consumer valuations coincide so that the monopolist chooses the right number of products under the Spence criterion of given total output. Another related paper is S. P. Anderson et al. (1987), which describes the formal connection between a Logit random utility model and an aggregate demand system featuring a representative agent with Constant Elasticity of Substitution (CES) preferences. This paper provides a formal connection between specific assumptions on the distribution of the shocks in additive random utility models and the resulting aggregate inverse demand curve that shifts in parallel with exogenous changes in product variety. Our theoretical approach makes use of Extreme Value Theory, which has been used in an additive random utility context in Gabaix et al. (2016) to show that there can exist high markups in large markets in equilibrium that are insensitive to the degree of competition. Our paper also relates to results in Kroft et al. (2021) who show that the parallel demands property is useful for identifying the love-of-variety from the passthrough of taxes under free entry. Lastly, our application of these theoretical results to identifying the benefits to consumers from greater variety relates to a large theoretical and empirical literature in international trade and industrial organization (see Arkolakis et al. 2008; Berry & Waldfogel 1999; Broda & Weinstein 2006; Dhingra & Morrow 2019; Dixit & Stiglitz 1977; Feenstra 1994; Mankiw & Whinston 1986; Romer 1994; Spence 1976a,b).

1.2. Parallel Demands: Symmetric Products

In this section, we consider a discrete choice model with symmetric products and derive necessary and sufficient conditions under which inverse market demand curves, evaluated at different levels of product variety, are exactly parallel. Next, we characterize a class of models where parallel demands is likely to be a good approximation.

1.2.1. Necessary and Sufficient Conditions

Consider a unit mass population of ex ante identical and independent consumers indexed by i. Consumers either choose to purchase a single product in the market $j \in \{1, ..., J\}$, where J is defined as the number of product varieties available, or choose the outside option $j = 0$.

Preferences. The indirect utility of individual i who purchases product j is given by:

$$u_{ij}(y_i, p_j) = \alpha(y_i - p_j) + \delta_j + \varepsilon_{ij} \tag{1.1}$$

where the scalar α is the marginal utility of income, y_i is consumer $i's$ income, p_j is the price of good j, δ_j is the quality of product j which captures vertical differentation and ε_{ij} is an idiosyncratic match value between consumer i and product j which captures heterogeneity in tastes across consumers and products and the degree of horizontal differentiation. The utility of individual i who chooses the outside option is given by $u_{i0} = \alpha y_i + \varepsilon_{i0}$.

Product-Level Demand. The indirect utility function in equation (1.1) generates demand for product j, $q_j(p_1, \ldots, p_J) : \mathbb{R}_+^J \to \mathbb{R}_+$, which we express as

$$q_j(p_1, \ldots, p_J) = \mathbb{P}\left(u_{ij}(y_i, p_j) = \max_{j' \in \{0, \ldots, J\}} u_{ij'}(y_i, p_{j'})\right) \tag{1.2}$$

Aggregate Demand. We express aggregate demand for all products excluding the outside good, when J varieties are available, as $Q(p_1, \ldots, p_J) : \mathbb{R}_+^J \to \mathbb{R}_+$, which takes the form

$$Q(p_1, \ldots, p_J) = \sum_{j=1}^{J} q_j(p_1, \ldots, p_J) \tag{1.3}$$

The share of the outside good is $q_0 = 1 - Q$. We now impose the following symmetry assumption.

Assumption 1: *We assume that (1) the random utility shocks $(\varepsilon_{ij})_{j=1}^{\infty}$ are continuously, independently, and identically distributed (i.i.d.), and are independent of the distribution of ϵ_{i0}, y_i, and $(\delta_j)_{j=1}^{\infty}$; (2) product qualities are symmetric, $\delta_j = \delta$.*

Assumption 1 implies that product prices will be identical in equilibrium ($p_j = p_k, \forall j, k \in \{1, ..., J\}$) under the additional assumption of identical production costs.[2] With symmetric prices, we can express the demand function as $q(p, J) : \mathbb{R}^2_+ \to \mathbb{R}_+$ and the aggregate demand function $Q(p, J) : \mathbb{R}^2_+ \to \mathbb{R}_+$ respectively as

$$q(p, J) = \mathbb{P}\left(u_{ij}(y_i, p) = \max_{j' \in \{0, ..., J\}} u_{ij'}(y_i, p)\right)$$

$$Q(p, J) = Jq(p, J)$$

Next, noting that $Q(p, J)$ is a strictly decreasing function with respect to p, we can invert it to obtain the inverse aggregate demand function $P(Q, J) : \mathbb{R}^2_+ \to \mathbb{R}_+$. We now introduce our definition of parallel demands with symmetric products.

Definition 1: The discrete choice model with symmetric products is said to give rise to parallel demands if for all J_0, $J_1 \neq J_0$, and Q

$$\frac{\partial P}{\partial Q}(Q, J_0) = \frac{\partial P}{\partial Q}(Q, J_1) \tag{1.4}$$

where $P(Q, J_t), t \in \{0, 1\}$ is the inverse aggregate demand function, and J_0 and J_1 are any numbers of product varieties. An equivalent definition of parallel demands that we will make use of below is $Q(p, J_0) = Q(p + d(J_0, J_1), J_1)$; in other words, there exists some index $d(J_0, J_1)$, such that output is the same when the price is p with J_0 varieties or the price is $p + d(J_0, J_1)$ with J_1 varieties.

We now state our first theorem using Definition 1 and Assumption 1.

[2] We do not explicitly model market equilibrium in this paper, but symmetric prices are achieved in equilibrium (Nash in prices) when firms have identical costs as shown in S. P. Anderson & Palma (1992).

Theorem 1: *Suppose that Assumption 1 holds, prices are symmetric and ε_{i0} follows a continuous distribution. Then a necessary and sufficient condition for parallel demands (Definition 1) is that the random utility shocks $(\varepsilon_{ij})_{j=1}^{\infty}$ in equation (1.1) follow a Gumbel distribution $G(x) = e^{-e^{-\frac{x-\mu}{\beta}}}$ for some location and scale parameters $\mu \in \mathbb{R}$ and $\beta > 0$.*

Proof. See Appendix.

As an illustration, in equation (1.1), if ε_{i0} is also Gumbel, then this model corresponds to a multinomial Logit model in which there are $J_0 + 1$ products including the outside option. For any $j \in \{1, ..., J_0\}$

$$q(p, J_0) = \frac{e^{\delta - \alpha p}}{1 + J_0 e^{\delta - \alpha p}}$$

Aggregate demand is equal to

$$Q(p, J_0) = \frac{J_0 e^{\delta - \alpha p}}{1 + J_0 e^{\delta - \alpha p}}$$

Thus, the inverse aggregate demand curve of the multinomial Logit model is given by

$$P(Q, J_0) = \frac{\delta}{\alpha} + \frac{1}{\alpha} \log J_0 - \frac{1}{\alpha} \log \left(\frac{Q}{1 - Q} \right)$$

We verify that $\frac{\partial P}{\partial Q}(Q, J_0) = -\frac{1}{\alpha}\frac{1}{Q}\frac{1}{1-Q} = \frac{\partial P}{\partial Q}(Q, J_1)$ and so Definition 1 is satisfied. Equivalently, note that $Q(p, J_0) = \frac{J_0 e^{\delta - \alpha p}}{1+J_0 e^{\delta - \alpha p}} = \frac{J_1 e^{\delta - \alpha(p + d(J_0, J_1))}}{1+J_1 e^{\delta - \alpha(p+d(J_0,J_1))}} = Q(p + d(J_0, J_1), J_1)$ for $d(J_0, J_1) = \frac{1}{\alpha} log \left(\frac{J_1}{J_0} \right)$.

1.2.2. Asymptotic Approximation as J Grows Large

The previous section showed that Gumbel random utility shocks are both necessary and sufficient for parallel demands in the case of symmetric products. Using Extreme Value Theory, we now show that there is a large class of random utility shocks beyond Gumbel that admit parallel demands asymptotically (as J grows large). The additive random utility models in this class have in common that the distribution of the maxima of the shocks is asymptotically Gumbel, which implies that the inverse aggregate demand curves are asymptotically parallel. We now define a class of models that admit this asymptotic approximation, and we provide a sufficient condition to show that a given additive random utility model is in this class.

Definition 2: Let (ε_{ij}) be i.i.d. distributed according to a continuous CDF F. Following Resnick (1987), F is in the domain of attraction of the Gumbel distribution if and only if there exist sequences (a_n, b_n) of real numbers such that $F^n(a_n x + b_n) \to G(x)$ for all x, where $G(x) = e^{-e^{-x}}$ is the standard Gumbel distribution.

Lemma 1: *Let x_0 be the supremum of the support of a CDF F that is twice continuously differentiable. If F satisfies $\lim_{x \to x_0} \frac{F''(x)(1-F(x))}{F'^2} = -1$ then F is in the domain of attraction of the Gumbel distribution.*

See Resnick (1987) for a proof of Lemma 1 and a full characterization of the domain of attraction of the Gumbel distribution. Although the characterization of the domain of attraction is outside the scope of the paper, it is worth mentioning the important result in statistics (the Fisher–Tippett–Gnedenko theorem) that plays a role akin to the Central Limit Theorem for Extreme Value theory. The result states that for a sequence of i.i.d. random variables X_i, letting $M_n = \max\{X_{1,}X_{2,}\ldots,X_n\}$ then if a sequence of real numbers (a_n,b_n) exists such that $\lim_{n\to\infty}\mathbb{P}\left(\frac{M_n-b_n}{a_n}\right) = F(x)$, where F is a non-degenerate CDF, then F is either Gumbel, Fréchet or Weibull. A useful intuition is that if the tails of the random utility shocks (ε_{ij}) are "very thin" the resulting converging distribution is Weibull, while if they are "heavy" the distribution of the maxima converges to Fréchet. Gumbel is the intermediate case that gives rise to parallel demands. For our purposes Lemma 1 is enough to show that some common distributions fall into the Gumbel domain of attraction.

The domain of attraction of the Gumbel distribution includes the Normal $N(\mu,\eta^2)$, Exponential, Lognormal, Gamma, Chi-square, and Weibull distributions, but does not include heavy-tailed distributions like the Cauchy, Fréchet, Pareto or Student distributions nor does it include short-tailed distributions like the Beta and Uniform distributions. The next theorem shows that inverse aggregate demands become "asymptotically" parallel as variety increases, for any additive random utility model with shocks in the Gumbel domain of attraction.

Theorem 2: *Let the random utility shocks (ε_{ij}) be i.i.d. and distributed according to F in the domain of attraction of the Gumbel distribution. Then for any $\epsilon > 0$, there exists large enough J_0 such that for all $J_1 > J_0$, there exists $d(J_0, J_1)$ such that for all $p \in \mathbb{R}_+$ we have*

$$|Q(p, J_0) - Q(p + d(J_0, J_1), J_1)| < \epsilon.$$

Therefore the inverse demands are approximately parallel $P(Q, J_1) \approx P(Q, J_0) + d(J_0, J_1)$ for all Q, for large enough J_0 and J_1.

Proof. See Appendix.

Later in the paper we assess the approximation result in Theorem 2 by numerically simulating different additive random utility models and considering the effect of an exogenous change in the number of varieties on consumer surplus, using the exact formulas for consumer surplus in additive random utility models and using a reduced-form approach that assumes demands are parallel.

1.3. Generalizations and Extensions: Correlated Tastes and Asymmetric Products

In this section, we generalize the model in 1.2.1 to a Logit model with correlated tastes, and we also consider a model with asymmetric products. While preserving the Extreme Value distribution of consumers' tastes within the inside market, the model with correlated tastes in many cases better captures the substitution patterns of products by allowing different substitutability within the variety market relative to the outside option and correlated tastes across products within the variety market. We show that in this model, we continue to obtain parallel demands when the distribution of random utility shocks satisfies the necessary and sufficient condition in Theorem 1. When we extend to asymmetric products, we are also able to obtain analogous results.

1.3.1. Logit Model with Correlated Tastes

Similar to the multinomial Logit model, we consider a population of statistically identical and independent consumers indexed by i of mass unity who choose to purchase a single product $j \in \{1, ..., J\}$ or the outside option $j = 0$. We extend the baseline model to allow preferences across products to be correlated within individuals.

Preferences. The indirect utility of individual i who purchases product j is given by:

$$u_{ij}(y_i, p_j) = \alpha(y_i - p_j) + \delta_j + (1-\sigma)\nu_i + \sigma\varepsilon_{ij} \tag{1.5}$$

where $(1-\sigma)\nu_i + \sigma\varepsilon_{ij}$ is the idiosyncratic match value between consumer i and product j, which captures heterogeneity in tastes across consumers and products, and correlation in tastes across products. When $\sigma = 1$ and ε_{ij} follows the Gumbel distribution, we obtain the Logit model and when $\sigma = 0$, consumer tastes for all products in the inside market are perfectly correlated. Thus, the parameter σ captures the degree of correlation in consumer preferences across products of the inside market. The utility of individual i who chooses the outside option is given by $u_{i0} = \alpha y_i + \varepsilon_{i0}$. Similar to the Logit model, we make the following assumption.[3]

Assumption 2: *We assume that (1) for $j \neq 0$, the random utility shocks $(\varepsilon_{ij}), j = 1...J$ are continuously, independently and identically distributed (i.i.d.) and independent of ε_{i0}, y_i, ν_i, and $\delta_j, j = 1...J$, but we allow ε_{i0} to be correlated with ν_i; (2) product qualities and prices are symmetric $\delta_j = \delta$ and $p_j = p$.* The next proposition extends the result in Theorem 1 to cover correlated tastes.

Proposition 1: *Suppose that Assumption 2 holds. Then, a necessary and sufficient condition for parallel demands (Definition 1) is that the random utility shocks (ε_{ij}) in equation* (1.5) *follow a Gumbel distribution.*

Proof. See Appendix.

[3] As in Section 2, we do not model the market equilibrium. Instead, we assume symmetric prices directly, which would be achieved as an equilibrium outcome in a Nested Logit demand model when firms have identical costs, following S. P. Anderson & Palma (1992).

The logic of the proof is the following: since the term $(1-\sigma)\nu_i$ does not vary across products, we can use a location normalization for utility and move this term into the outside option. Then, we can apply the same arguments in the proof of Theorem 1. This explains why it is not necessary to impose a specific functional form assumption on the distribution for $(1-\sigma)\nu_i$. While the proposition does not require a specific distribution, we can use the Nested Logit model as a special case of this model to illustrate this result.

In the Nested Logit model, the random utility shocks (ε_{ij}) in equation (1.5) are drawn from the Gumbel distribution, and $(1-\sigma)\nu_i$ has the distribution derived in Cardell (1997). In our case, there are two nests: one which includes $j = 1, ..., J$, and another nest which includes only the outside option $j = 0$.[4] In the Nested Logit model, product demand is:

$$q(p, J) = \frac{J^{\sigma-1}e^{\delta-\alpha p}}{1 + J^{\sigma}e^{\delta-\alpha p}}.$$

In turn, aggregate demand is equal to:

$$Q(p, J) = \frac{J^{\sigma}e^{\delta-\alpha p}}{1 + J^{\sigma}e^{\delta-\alpha p}}.$$

Inverting aggregate demand, the inverse aggregate demand curve is given by:

$$P(Q, J) = \frac{\delta}{\alpha} + \frac{\sigma}{\alpha}\log J - \frac{1}{\alpha}\log\left(\frac{Q}{1-Q}\right).$$

[4]See Cardell (1997) for the class of distributions, termed $C(\cdot)$ distributions, which makes the combined idiosyncratic shocks distributed Type I Extreme Value, and thus allows us to write the demand function in a closed form.

Thus, we see that the Nested Logit model (like the symmetric products Logit model above) satisfies Definition 1 since $\frac{\partial P}{\partial Q}(Q,J_0) = -\frac{1}{\alpha}\frac{1}{Q}\frac{1}{1-Q} = \frac{\partial P}{\partial Q}(Q,J_1)$ or equivalently $d(J_0,J_1) = \frac{\sigma}{\alpha} log\left(\frac{J_1}{J_0}\right)$.

1.3.2. Asymmetric Products

Assumptions 1 and 2 impose symmetric products and prices, which leads to clean results but may limit the generality of the model. We now extend our results by considering asymmetric products so that $\delta_j \neq \delta_k$ and $p_j \neq p_k$ for $j \neq k$, and we continue to allow for an outside option as in the previous sections. In order to characterize parallel demands in this general model, we impose a technical assumption that we use in Theorem 3 below.

Assumption 3: *We assume that (1) for $j \neq 0$, the random utility shocks $(\varepsilon_{ij})_{j=1}^{\infty}$ are continuously, independently and identically distributed (i.i.d.) and independent of ε_{i0} which has a continuous distribution; (2) $(\delta_j)_{j=1}^{\infty}$ is a deterministic sequence of real numbers, and there exists a real number $K > 0$ such that all the quality parameters are bounded: $\delta_j \in [0, K]$ for all j.*

In the case of symmetric products and prices considered above, we were able to invert the aggregate demand since there was a mapping from aggregate quantity to a single (uniform) price at a given level of product variety. This inverse aggregate demand curve corresponded to the distribution across consumers of their maximum willingness-to-pay (WTP) for any level of product variety. When prices and products are asymmetric it is no longer straightforward to characterize the inverse aggregate demand curve. Thus, with asymmetric products we instead state our results in terms of the distribution of WTP rather than in terms of aggregate demand. In particular we study the distribution of the random variable $\max_{j\in\{0,\ldots,J\}} wtp_{ij}$, where $wtp_{ij}(\delta_j) \equiv \frac{\delta_j+\varepsilon_{ij}-\varepsilon_{i0}}{\alpha}$. We now introduce the definition of parallel shifts in WTP.

Definition 3: Let $WTP_i(J) \equiv \max_{j\in\{1,\ldots,J\}} wtp_{ij}(\delta_j)$. The discrete choice model in equation with asymmetric products is said to give rise to parallel shifts in willingness-to-pay (WTP) if for all $J_1 \neq J_0$, there exists $d(J_0, J_1) \in \mathbb{R}$, such that for all $x \in \mathbb{R}$:

$$\mathbb{P}\left(WTP_i(J_0) \leq x\right) = \mathbb{P}\left(WTP_i(J_1) \leq x + d(J_0, J_1)\right).$$

In particular, when $J_1 > J_0$, if consumers value variety, then we expect that $d(J_0, J_1) > 0$.

With this definition of parallel WTP shifts, we can now state the theorem that generalizes Theorem 1 to the case of asymmetric products.

Theorem 3: *A discrete choice model with asymmetric products satisfying Assumption 3 gives rise to parallel shifts in WTP (Definition 3) for all models satisfying Assumption 3 if and only if the random utility shocks $(\varepsilon_{ij})_{j=1}^{\infty}$ follow a Gumbel distribution (independently of the distribution of ε_{i0}).*

Proof. See Appendix.

Note that Theorem 3 lets us reinterpret Definition 3 in terms of aggregate demand. Assuming Gumbel shocks, Theorem 3 implies that we also get parallel shifts in consumer surplus $\max_{j\in\{1,\ldots,J\}} wtp_{ij}(\delta_j - \alpha p_j)$ (by substituting $\hat{\delta}_j = \delta_j - \alpha p_j$) and so the shift $d(J_0, J_1)$ can be seen as either a horizontal shift in the CDF of $WTP_i(J_0)$ or a vertical shift of the following function:

$$Q(s) \equiv Q(p_1 + s, \ldots, p_J + s, J_0) = \mathbb{P}\left(\max_{j \in \{1,\ldots,J\}} wtp_{ij}(\delta_j - \alpha\,(p_j + s)) \geq 0\right)$$

which maps aggregate demand as a function of the price index s. Lastly, as in the symmetric case, we can also use Extreme Value Theory to show that there is a larger class of models that admit parallel WTP asymptotically.

Theorem 4: *Suppose Assumption 3 holds. Let $(\epsilon_{ij})_{j=1}^{\infty}$ be i.i.d. and distributed with CDF F in the domain of attraction of the Gumbel distribution. Furthermore, assume there exists (α_n, β_n) and a nondegenerate CDF H such that $\Pi_{j=1}^{n} F(\alpha_n x + \beta_n - \delta_j) \to H(x)$ for all x.*[5] *Then for any $\epsilon > 0$, there exists large enough J_0 such that for all $J_1 > J_0$, there exists $d(J_0, J_1)$ such that for all $x \in \mathbb{R}$*

$$|\mathbb{P}\left(WTP_i(J_0) \leq x\right) - \mathbb{P}\left(WTP_i(J_1) \leq x + d(J_0, J_1)\right)| < \epsilon.$$

Proof. See Appendix.

The technical result in Theorem 4, extends Theorem 2 to the maxima of non i.i.d. random variables. In the mathematics and statistics literature, it has proven difficult to extend the Fisher–Tippett–Gnedenko theorem to non i.i.d sequences of random variables. In particular, Kreinovich et al. (2015) show the impossibility of a simple generalization

[5]This second assumption is satisfied automatically for sequences where $(\delta_j)_{j=1}^{\infty}$ is non-increasing or non-decreasing. The condition may also be violated for alternating sequences. A counterexample can be constructed by taking $\delta = 0$ or $\delta = K$ for alternate periods of increasing length. Therefore this assumption constrains the variation in the vertical differentiation parameter of the new varieties that can enter the market.

of the Fisher–Tippett–Gnedenko theorem when random variables are not identically distributed and contrast it to the Central Limit Theorem where this is possible. In our particular case, we are able to show that when the sequence of random variables is composed of mean shifts of the same CDF in the domain of attraction of the Gumbel distribution, the asymptotic theorem obtains.

The results thus far demonstrate a connection between discrete choice models featuring Gumbel-type preferences and parallel demands. The next section provides an example where parallel demands are valuable as an economic tool.

1.4. Parallel Demands as an Economic Tool: Identification of the Variety Effect

In this section, we show how to use parallel demands to identify the change in consumer surplus from an exogenous change in variety. Measuring the change in consumer surplus due to a change in variety has been studied in many branches of economics, including international trade and industrial organization (see Berry & Waldfogel 1999; Broda & Weinstein 2006; Dhingra & Morrow 2019; Feenstra 1994). We begin with the standard definition of consumer surplus and derive the variety effect in the general case. When there are new varieties introduced into the market, the variety effect depends on all of the demands for the new goods. When there are many differentiated products, as is typically the case in economic applications, this is a high dimension problem with a large number of parameters to be estimated and we need to impose some form of dimension reduction. We consider two complementary approaches: symmetry and aggregation. First, we consider a symmetric product environment, as is typically assumed in models in macro and trade, and show that we can characterize the variety effect as the area between two inverse aggregate demand curves. Second, we allow for heterogeneity in demands and prices but assume that prices are correlated which allows us to aggregate; specifically, we assume that prices shift by the same amount after the introduction of new varieties. This result relates to Hicks (1936) that in order to aggregate goods into commodities, prices of the goods must be highly correlated. The advantage of aggregation is that it permits one to be more flexible on functional forms without having to specify underlying preferences. The disadvantage is that prices may not be highly correlated.[6]

1.4.1. Variety Effect

Consider the general discrete choice model in section 1.3.2 with J asymmetric products and prices which are denoted by the vector $\mathbf{p_J}$. There are no income effects which means that consumer surplus is a valid measure of welfare and we can avoid the problem of path dependence of price changes.

Definition 4: Let $Q_J(\mathbf{p})$ be the aggregate demand when there are J differentiated products and prices are given by $\mathbf{p_J} = (p_1, p_2, ..., p_J)$. In this case, consumer surplus is defined:

$$CS(\mathbf{p_J}, J) = \int_0^\infty Q_J(\mathbf{p_J} + s\mathbf{1}_{\boldsymbol{J}})ds \tag{1.6}$$

When new varieties are introduced into the market, there are two effects on consumer surplus. First, there is a "price effect" that arises since market prices may change when firms enter or exit the market. Second, there is a "variety effect" which captures how

[6]See discussion in Nevo (2011) for the dimensionality problem and alternative approaches to identifying demand.

much a new variety increases consumer surplus, holding prices constant. In this section, we focus on the "variety effect" which we define as follows.

Definition 5: Let $\mathbf{p_{J_0}} = (p_1, p_2, ..., p_{J_0})$ and $\mathbf{p_{J_1}} = (\mathbf{p_{J_0}}, p_{J_0+1}, \ldots, p_{J_1})$. The "variety effect" when the number of products goes from J_0 to J_1 (with $J_1 > J_0$) is defined as:

$$\Lambda = \int_0^\infty Q_{J_1}(\mathbf{p_{J_1}} + s\mathbf{1_{J_1}})ds - \int_0^\infty Q_{J_0}(\mathbf{p_{J_0}} + s\mathbf{1_{J_0}})ds \tag{1.7}$$

From Definition 5 we see identifying the variety effect requires identification of aggregate demand before and after the change in varieties. In order to make the problem more tractable, we focus on two special cases: symmetric products and asymmetric products with correlated prices (aggregation).

1.4.2. Symmetry

When all potentially existing products are symmetric, in the equilibium we have $p_j = p_k, \forall j, k$. Then we can use the definitions and foundations laid in Section 2 to simplify the expressions of consumer surplus and the variety effect as follows. First, consumer surplus is defined as the integral of aggregate demand:

$$CS(p, J) = \int_p^\infty Q(s, J)ds \tag{1.8}$$

Next, using the inverse demand $P(Q, J)$ we can adapt Definition 5 for the variety effect when variety in the market changes from J_0 to J_1 to:

$$\Lambda = \int_0^Q \big(P(s, J_1) - P(s, J_0)\big)ds \tag{1.9}$$

where instead of holding fixed prices, we are holding fixed quantity, as this will prove more useful in this section. The next result shows that, the variety effect can be calculated exactly in a simple form when we assume parallel demands.

Proposition 2: *Starting from equilibrium quantity Q_0 and price p_0, under the assumption of parallel demands (Definition 1), when variety changes from J_0 to J_1, the variety effect can be equivalently expressed as:*

$$\Lambda = Q_0 * d(J_0, J_1) \tag{1.10}$$

where $d(J_0, J_1)$ is such that $P(Q, J_0) + d(J_0, J_1) = P(Q, J_1)$.

Proof. See Appendix.

The price effect and variety effect are illustrated in Figure 1.1 which considers a reduction in product variety in the market from J_0 to J_1 . The price effect is represented by the area $efgh$ and the variety effect is given by the area $abcd$, so that $-\Delta CS = abcd - cdg + efgh$, where cdg is an adjustment that is second-order relative to $\Delta Q * \Delta J$. Intuitively, when the number of varieties is reduced, some consumers will no longer be able to purchase their most preferred option. Thus, the maximum willingness-to-pay for purchasing an inside good will be lower for these consumers. This is represented as a downward shift in the inverse aggregate demand curve. The area between the inverse aggregate demand curves $abcd$ before and after the change in variety up to initial quantity Q_0 corresponds exactly to the variety effect.

We can now state our next Proposition, which uses Definition 1.

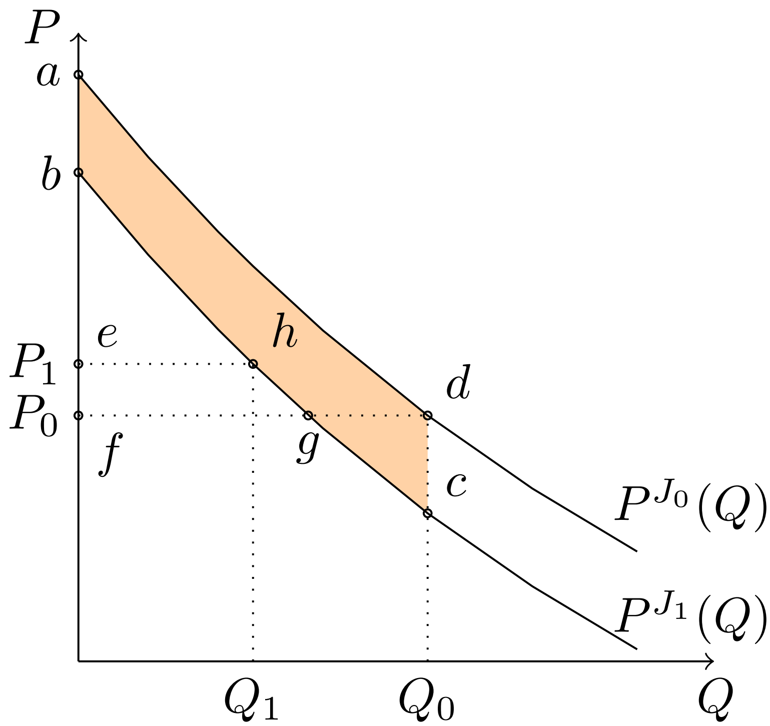


Figure 1.1. Variety Effect

Notes: This figure shows the result of a decrease in variety (from J_0 to J_1). The shaded area *abcd* between the two demand curves represents the variety effect.

Proposition 3: *Denote the equilibrium quantity Q_0 and market price p_0 when initial variety is J_0. Consider an exogenous increase in varieties from J_0 to J_1 and denote the new equilibrium quantity Q_1 and market price p_1. Under the assumption of parallel demands (Definition 1):*

$$d(J_0, J_1) = p_1 - P(Q_1, J_0) = \left(\frac{\frac{dp}{dJ}}{\frac{dQ}{dJ}} - \frac{\partial P(Q,J)}{\partial Q} \right) \frac{dQ}{dJ} \triangle J + O((\triangle J)^2) \tag{1.11}$$

where $\frac{\partial P(Q,J)}{\partial Q}$ denotes the slope of inverse demand when variety J is held fixed and $\frac{dp}{dJ}/\frac{dQ}{dJ}$ denotes the slope of inverse demand when J is variable.

Proof. See Appendix.

When variety changes from J_0 to J_1, prices change from p_0 to p_1. However, this is not sufficient to recover $d(J_0, J_1)$. This is because the counterfactual price $P(Q_1, J_0)$ is not directly observable since it depends on the market price that would prevail at the final level of output but on the original demand curve. To see how to recover an expression for $d(J_0, J_1)$, we note from Figure 1 that it must satisfy the following relationship $Q(p_1, J_1) = Q(p_1 - d(J_0, J_1), J_0)$. Thus, we can identify $d(J_0, J_1)$ as follows:

$$dQ = Q(p_1, J_1) - Q(p_0, J_0)$$

$$dQ = Q(p_1 - d(J_0, J_1), J_0) - Q(p_0, J_0)$$

$$dQ \approx \frac{\partial Q(p, J)}{\partial p}|_{p=p_0}(-d(J_0, J_1) + p_1 - p_0)$$

$$dQ = \left.\frac{\partial Q}{\partial p}\right|_{p=p_0}(-d(J_0, J_1) + \frac{dP}{dQ}dQ)$$

The first equality holds by definition. The second equality holds by Definition 1. The third approximation holds by doing a Taylor expansion of $Q(p, J)$ around p_0. The fourth equality holds by definition. Rearranging and solving for $d(J_0, J_1)$ yields:

$$d(J_0, J_1) \approx dp - \frac{\partial P}{\partial Q}dQ$$

In economic terms, $d(J_0, J_1)$ can be interpreted as the reduction in the willingness-to-pay for the marginal unit. Under Definition 1, it can further be interpreted as the change in willingness-to-pay for inframarginal units. In order to identify $d(J_0, J_1)$, two causal effects are required. First, one requires the effecs of an exogenous change in variety on prices $(\frac{dp}{dJ})$ and output $(\frac{dQ}{dJ})$. Second, one requires the effect of prices on demand, holding variety fixed, $\frac{\partial P}{\partial Q}$. Intuitively, when we multiply $\frac{\partial P}{\partial Q}\Delta Q$ we are implicitly calculating the counterfactual price that would hold when J_0 varieties are available and quantity is adjusted to Q_1.

1.4.3. Numerical Simulations

From Theorem 1 we know that when we have preferences in the form of (1.1) and the random utility shocks follow the Gumbel distribution, we can apply the parallel demands and compute $\frac{\partial P}{\partial J}(Q', J)$ for any Q' on the support of the aggregate demand function. This saves us from integrating over the whole support. Moreover, from Theorem 2 we have that as long as the random utility shocks (ε_{ij}) are distributed according to F in the domain of attraction of the Gumbel distribution, for any large enough varieties, we have parallel demands as an approximation. In this subsection we show these results in Monte Carlo simulations. We assess the parallel demands assumption by simulating a model of a large number of consumers with utility over products given by equation (1.1). We choose $\alpha = 1$ and $y = 1$ in the simulation, and we consider four different shock distributions (Gumbel, Normal, Exponential, and Pareto). For each distribution, we consider a hypothetical 20 percent increase in the number of products (from an initial value of J), and we compute the exact change in consumer surplus resulting from this change in variety by numerically integrating the increase in consumer surplus across each consumer. Then, we compute the change in consumer surplus implied by assuming parallel demands following equations (1.10) and (1.11).

The results in Figure 1.2 show that the bias that arises from assuming parallel demands is a function of the number of varieties in the market, measuring the bias as the

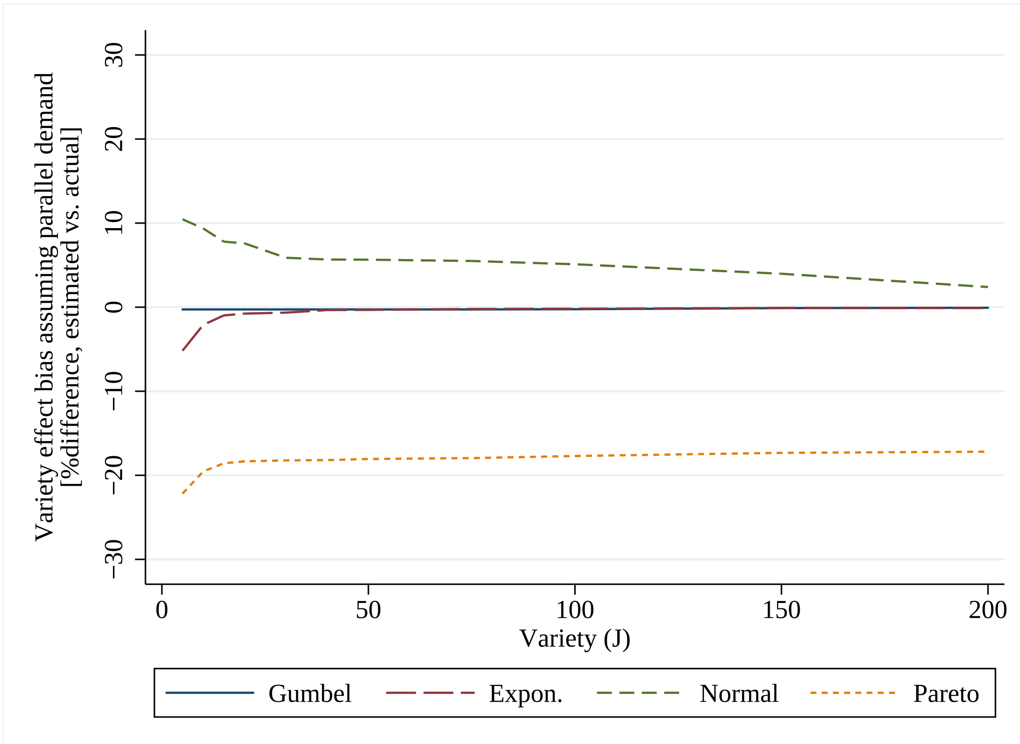


Figure 1.2. Approximate Parallel Demand Curves

Notes: This figure reports results from numerical simulations that are designed to evaluate the quality of the key approximation theorem (Theorem 2) in the main text. By simulating simple discrete choice models under different assumptions about the distribution of the i.i.d. error terms and increasing the number of varieties in the market, we calculate the (exact) variety effect numerically and compare it to the variety effect that we would infer from assuming parallel demands. Consistent with the result of Theorem 2, for distributions that satisfy assumptions of theorem, as J increases, the bias in the variety effect from assuming parallel demands approaches zero.

difference between the estimated and the exact change in consumer surplus. The benchmark distribution is Gumbel, where we know from Theorem 1 that the demand curves are exactly parallel, and therefore the bias is 0 for all initial values of J. For both the Normal and Exponential distributions, we find that the bias is small in magnitude and converges to 0 fairly quickly as the number of varieties increase. On the other hand, with a Pareto distribution, there is a bias of roughly 20 percent, which does not vanish as varieties increase. In this case, the change in consumer surplus from assuming parallel demands is a lower bound on the true change in consumer surplus, and it does not converge to 0 because the Pareto distribution is not in the domain of attraction of the Gumbel distribution.

1.4.4. Aggregation

The previous results focus on the special case symmetric products, which allows for a clear graphical representation since the inverse aggregate demand curve can be defined for a uniform (symmetric) price. We now consider the case of asymmetric products. The main objective in what follows is to give plausible and parsimonious sufficient conditions to identify the variety effect using reduced-form methods based on local information.

We first note that under the assumption of parallel shifts in WTP (Definition 3), there exists some price index $d(J_0, J_1)$ such that $Q_{J_1}(\mathbf{p_{J_1}} + s\mathbf{1}_{\boldsymbol{J_1}}) = Q_{J_0}(\mathbf{p_{J_0}} + (s - d(J_0, J_1))\mathbf{1}_{\boldsymbol{J_0}})$ for all $s \in \mathbb{R}$.[7] In other words, increase prices starting from $\mathbf{p_{J_0}}$ by some constant amount $d = d(J_0, J_1)$ until total quantity demanded equals quantity demanded when there are J_1 products in the market. Under this assumption, it follows that the variety effect can be expressed as:

$$\Lambda = \int_0^d Q_{J_0}(\mathbf{p_{J_0}} + (s - d)\mathbf{1}_{\boldsymbol{J_0}})ds.$$

Next, by the mean value theorem for integrals, there exists $d' \in [0, d]$ such that

$$\Lambda = Q(\mathbf{p_{J_0}} - d'\mathbf{1}_{\boldsymbol{J_0}}) * d.$$

[7]This is related to the price index in Feenstra (1994). However, in Feenstra (1994), the price index is defined as the (common) price change that would have to occur when there are J_0 goods in the market in order to give the same *utility* as when there are J_1 goods.

This leads to the following result.

Proposition 4: *Under the assumption of parallel shifts in WTP (Definition 3), when variety changes from J_0 to J_1 , there exists $d' \in [0, d(J_0, J_1)]$ such that*

$$\Lambda = Q_{J_0}(\mathbf{p_{J_0}} - d'\mathbf{1_{J_0}}) * d(J_0, J_1). \tag{1.12}$$

All that remains is to develop a method to identify $d(J_0, J_1)$. To do this, we introduce an additional technical assumption.

Assumption 4: *The prices of the existing products in the market ($j = 1, ..., J_0$) shift by the same amount after the introduction of new varieties, i.e. $p_j^1 - p_j^0 = p_k^1 - p_k^0$ for all products j, k available in both periods of time.*

With this assumption in hand, we can now state our main result for asymmetric products.

Proposition 5: *Suppose that the assumption of parallel shifts in WTP (Definition 3) and Assumption 4 holds. Let the post-entry equilibrium prices be $\mathbf{p}_{J_1}$ and define $\Delta P \equiv \rho \in \mathbb{R}$ to be the change in price of any of the existing products before and after entry of new varieties. Letting $\Delta Q = Q_{J_1}(\mathbf{p_{J_1}}) - Q_{J_0}(\mathbf{p_{J_0}})$, we have:*

$$d(J_0, J_1) = \left(\frac{\Delta P}{\Delta Q} - \left. \frac{dP}{dQ_{J_0}} \right|_{J_0} \right) \Delta Q + O\left((\rho - d)^2\right) \tag{1.13}$$

where $\left. \frac{dP}{dQ_{J_0}} \right|_{J_0} = \left. \left(\frac{dQ_{J_0}(\mathbf{p_{J_0}} + t\mathbf{1_{J_0}})}{dt} \right)^{-1} \right|_{t=0}$.

Proof. See Appendix.

Several features of Proposition 5 are worth highlighting. First, observe that the key step for the Proposition to hold is to be able to find a ρ and d such that $Q_{J_1}(\mathbf{p_{J_1}}) = Q_{J_0}(\mathbf{p_{J_0}} + (\rho - d)\mathbf{1_{J_0}})$. This requires both that all prices adjust uniformly after the introduction of the new varieties (Assumption 4) and that aggregate demands shift in parallel (Definition 3). Restricting prices to adjust in the same direction $\mathbf{1_{J_0}}$ as the vertical shift d allows us to identify d by a simple application of the Taylor approximation theorem.

Second, we interpret the directional derivative $\left.\frac{dQ_{J_0}}{dP}\right|_{J_0} = \frac{dQ(\mathbf{p_{J_0}}+t\mathbf{1_{J_0}})}{dt} = \sum_{j=1}^{J_0} \frac{\partial Q_{J_0}}{\partial p_j}$ as the short-run slope of aggregate demand in the direction of uniform price changes, that connects the interpretation of (1.13) with equation (1.11) in the symmetric model. Furthermore, if we observe the change in aggregate demand Q_{J_0} when all prices are increased simultaneously, one does not need to identify each partial derivative separately; instead it is sufficient to identify $\left.\frac{dQ_{J_0}}{dP}\right|_{J_0}$.

1.5. Conclusion

This paper highlights a previously-unnoticed feature of a class of discrete choice models, which is that they feature parallel demand curves. Specifically, we show that in additive random utility models, inverse aggregate demand curves shift in parallel with respect to variety if and only if the random utility shocks follow the Gumbel distribution. While it may seem that the parallel demands property is a special case, our theoretical results suggest instead that parallel demands are a general property of many discrete choice models. Specifically, using results from Extreme Value Theory, we provide conditions for other distributions to generate parallel demand asymptotically, as the number of varieties increases. We illustrate these results using numerical simulations and extend them to cover correlated tastes and asymmetric products.

Given the generality of our theoretical results, we provide an application and show that parallel demands are useful to identifiy the change in consumer surplus from a change in variety. In this application, parallel demands provide a straightforward identification approach – intuitively, identifiying the "vertical gap" at one point in the aggregate demand curve is sufficient for identifying the entire area between the inverse aggregate demand curves before and after the change in variety. Because of this, we view the parallel demands property as a tool that can potentially be used for both producing theoretical results on the value of variety (which can be an input into theoretical analysis of whether the equilibrium level of variety is socially optimal) as well as a tool for empirical work, where the parallel demands assumption may be used as an alternative "reduced-form" identification approach (instead of relying on specific structural models of consumer demand for identification).

We conclude by speculating that parallel demands may also be a useful property when studying other economic questions. Discrete choice models are widespread in economics, and our theoretical results may therefore be useful in other economic settings, such as the choice of neighborhood (Bayer et al. 2007; McFadden 1978), occupation (Hsieh et al. 2013), firm (Card et al. 2018; Chan et al. 2019; Lamadon et al. 2020), and school (Dinerstein & Smith 2014). In all of these settings, as long as the parallel demands assumption holds, the welfare effects corresponding to changes in the number of available choices (or "varieties") may be calculated using the approach described in this paper.

CHAPTER 2

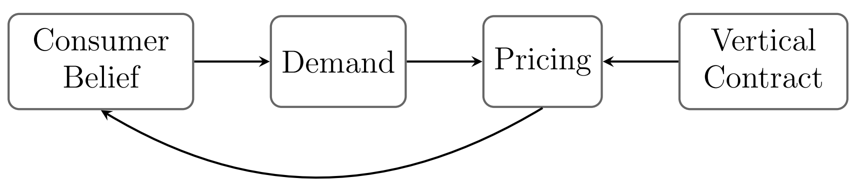
Price Reference Effects and Vertical Contracts in the Book Retail Market

2.1. Introduction

The theory of reference-dependent preferences, dating back at least to Kahnerman & Tversky (1979)'s work on prospect theory, captures a central intuition that outcomes are not experienced on an absolute scale, but rather are experienced relative to some point of reference. In many settings, behavioral economists have documented a price reference effect: the fact that a consumer's willingness to pay for a good is affected by difference between the observed price and the reference price they rationally expect (Gentry & Pesendorfer, 2021a). These findings raise the question of how firms might exploit the behavioral consumers (Ellison, 2006). In this paper, we study the impact of vertical contracting between an upstream producer and a downstream retailer when the consumer demand features price reference effects.

The textbook model of vertical integration suggests that it improves joint profits because of the elimination of double marginalization. However, this intuition may not hold when the retailer faces behavioral consumers with reference-dependent preferences in price. Figure 2.1 illustrates our conceptual framework. Consider a vertical integration between an upstream producer and a downstream retailer. The vertical integration incentivizes the retailer to lower retail prices because of the elimination of double marginalization. As a result, consumers' belief on prices will update in response to the lower prices as consumers form their expectations based on observed prices.[1] With price reference effects, consumers' willingness to pay would be lower when they expect lower prices. Thus, the increase in quantity due to lower prices would be smaller in the new equilibrium. If the price reference effect is sufficiently large, vertical integration can decrease joint profits as the decrease in price overshadows the increase in quantity. Ignoring the equilibrium channel of consumer belief may bias the counterfactual analysis. This paper aims to empirically test the existence of the price reference effect in a real-world setting and examines how it interacts with the vertical contracting.

Figure 2.1. Conceptual Framework



The book retail market is an ideal setting to study price reference effects and vertical contracts. In the online book retail market, retailers often give discounts for the books so consumers rarely purchase books at the list price. Since consumers form their expectations on book prices based on observed prices, the retailer's pricing decisions could influence consumers' expectations and their willingness to pay. Also, the retailer must negotiate wholesale prices with the upstream publisher. Recently, the rise of Amazon Publishing raises an important policy question of whether publishers and retailers should be allowed

[1]We use the phrase "consumer belief" or "consumer expectation", and "vertical integration" or "vertical merger" interchangeably.

to vertically integrate as Amazon receives a cut both as publisher and as retailer if a reader buys one of its titles. Incorporating price reference effects in the analysis of vertical integration could help us better understand how they interact in determining retail prices and joint profits.

We begin by presenting a model that predicts: (i) consumers' demand for books; (ii) retail prices set by the retailer; and (iii) wholesale prices negotiated between the publisher and the retailer. The model proceeds in three stages. In stage 1, the publisher and the retailer bargain over the wholesale price for the book being published; in stage 2, the retailer sets retail prices; and in stage 3, consumers with reference-dependent preferences in price make purchases. We start from the last stage and work backwards. To accommodate the price reference effect, we extend the discrete-choice demand model following the formulation of Kőszegi & Rabin (2006) and Gentry & Pesendorfer (2021a). In addition to the standard consumption utility, consumers experience gain-loss utility driven by differences between prices they rationally expect and prices they actually pay. To allow for comparison across books, we model the reference price as the ratio of retail price over list price that the consumer rationally expect for the book. Next, we analyze the equilibrium pricing decision of the retailer facing demand with the price reference effect. Following Gentry & Pesendorfer (2021b), we solve for the Bayesian Nash Equilibrium (BNE) in which the retailer maximizes its profit given consumers' belief, and consumers' belief is consistent with the retailer's pricing choices. Lastly, we model the bargaining between the publisher and the retailer as Nash bargaining over the wholesale price.

We use the model to illustrate how price reference effects may overturn the standard textbook intuition that vertical integration improves joint profits. We show theoretically that when the price reference effect is sufficiently large, vertical integration between an upstream producer and a downstream retailer can decrease joint profits. This is because the decrease in price overshadows the increase in quantity, resulting in the lower post-merger joint profits.

To test whether this force is large in a real-world setting, we estimate the model using a novel dataset from DangDang.com, the largest online book retailer in China. It had about 40% of market share in the online book retail market around 2018, and earned an annual revenue of \$1.5 billion. There were over 1 million unique books and 538 million copies sold on the platform between Jan 2017 and Aug 2019. DangDang is an ideal setting for investigating the impact of price reference effects and vertical contracts as it has data advantages. First, we observe both daily retail prices and the retail-list price ratio for each book on the platform. Therefore, we can separately identify the direct price effect and the indirect price reference effect using variations in consumers' reference expectations induced by variations in the retailer's costs over time. Second, we directly observe the daily quantity sold for each book, which is an important strength relative to other papers on online book retail markets that often have to overcome the problem of limited quantity data by imputing quantities from sales rankings. Third, we observe the bargained wholesale price for each book and use it directly in estimation rather than making separate assumptions to identify the bargaining parameter.

We estimate the consumer demand using a two-step procedure where the reference price distributions are estimated in the first step, and the remaining parameters are estimated in the second step. Our estimates document large price reference effects in this book retail market that we study, and the results are robust to including demand curvatures and alternative nested-logit specifications.

For the estimation of the supply side, we first use the equilibrium condition of the retailer's pricing problem to back out the marginal cost of the retailer. Then we use the Nash bargaining solution and observed book characteristics that correlate with the publisher's marginal costs to estimate the publisher's marginal cost and the bargaining parameter. We find that the publisher's marginal cost is around 40% of the list price which is consistent with industry sources. The retailer's marginal cost is estimated to be -30% of the list price. The negative marginal cost represents the value to the retailer of bringing in consumers beyond selling the book, e.g., loss-leader strategy or customer acquisition and retention (De los Santos & Wildenbeest, 2017). In fact, similar results are found in the e-book market, where Amazon frequently sets retail prices below the wholesale price (De los Santos et al., 2021). A number of empirical studies conclude that e-book and print prices are set below static profit maximizing levels on Amazon (Chevalier & Goolsbee, 2003, De los Santos et al., 2012, and Reimers & Waldfogel, 2017). In the estimation of bargaining parameters, we find that the retailer has most of the bargaining power, and that large publishers have more bargaining power than medium and small publishers. The predicted price distributions fit the general pattern of the observed price distributions.

In the counterfactual analysis, we compare equilibrium prices, quantities and firm profits under alternative specifications to empirically quantify the extent to which price reference effects and vertical contracts determine prices and firm profits. First, to examine the effect of consumer belief on equilibrium outcomes, we simulate a case where the consumers' belief is fixed, and the publisher and the retailer are allowed to re-optimize based on the fixed consumer belief (equivalently the demand curve). We find that the publisher and the retailer would negotiate a lower wholesale price, and the retailer would set lower retail prices. Joint profits would be higher given by the higher quantity sold. These results show that in the absence of price reference effects, the retailer would have additional incentives to lower the retail price since it no longer needs to consider the externality on consumer expectations in response to the lower prices.

Next, we simulate two alternative vertical contracts and examine how they interact with price reference effects. Table 2.1 summarizes the counterfactual results relative to the wholesale contract. First, we simulate a vertical merger between the publisher and the retailer. When consumers' belief is fixed, the post-merger joint profits are higher than the pre-merger joint profits as predicted by the standard model of eliminating double marginalization. However, when consumers' belief is in equilibrium, the opposite occurs. Second, we consider an alternative vertical contract of agency pricing where the publisher pays the retailer sales royalties to sell books at prices determined by the publisher. We find that joint profits would be lower when consumers' belief is fixed, while the opposite occurs when consumers' belief is in equilibrium.

In both alternative vertical contracts, the change in joint profits has the same sign as the change in retail price when consumers' belief is in equilibrium. This is because the change in quantity is smaller in magnitude when there exists the price reference effect as consumers' belief would update in response to the change in prices. Thus, the change in prices overshadows the change in quantities, resulting in the joint profits moving in the same direction as the prices. These findings highlight the importance of incorporating consumer expectations in the analysis of optimal pricing and firm profits.

Table 2.1. Summary of Counterfactual Results Relative to the Wholesale Contract

		Retail Price	Quantity	Joint Profits
Vertical Merger	Fixed Belief	−	+	+
	Equilibrium Belief	−	+	−
Agency Contract	Fixed Belief	+	−	−
	Equilibrium Belief	+	−	+

Notes: + indicates higher values and − indicates lower values.

Related Literature. This paper is related to three strands of literature. First, it relates to the literature on reference-dependent preferences that dates back to Kahnerman & Tversky (1979)'s work on prospect theory. Recent theoretical works have focused on the expectation-based paradigm due to Kőszegi & Rabin (2006, 2007, 2009) and Heidhues & Kőszegi (2008, 2014).[2] We join the growing body of empirical work, including in particular V. Crawford & Meng (2011), in supporting the expectation-based paradigm.[3] We follow Gentry & Pesendorfer (2021a) in extending the discrete-choice demand model in the spirit of Berry et al. (1995) to accommodate expectation-based reference-dependent price effects. We find large price reference effects in the book retail market. Also, we are the first, to the best of our knowledge, to empirically estimate the firm's pricing problem facing consumers with reference-dependent preference in price as derived theoretically in Gentry & Pesendorfer (2021b).

[2]See Bernheim et al. (2018) for a review in behavioral economics.
[3]There is a range of experimental studies including Abeler et al. (2011), Ericson & Fuster (2011), Gill & Prowse (2012) and Meng & Weng (2018). See also Mazumdar et al. (2005) for a review of quantitative marketing approaches that generally model reference prices as being formed on the basis of past purchases.

Second, our paper is related to the literature of vertical contracting in the book industry. The theoretical literature (Abhishek et al., 2016, Foros et al., 2017, and Johnson, 2017, 2020) focus on comparing retail prices and firm profits under wholesale and agency models. Empirically, De los Santos & Wildenbeest (2017) show that shifting from wholesale to agency pricing raises prices in the market for e-books.[4] To our knowledge, De los Santos et al. (2021) is the first paper to show that agency contracts can lead to higher or lower retail prices than wholesale contracts depending on the distribution of bargaining power. They structurally estimate a model with either contract form under Nash-in-Nash bargaining and find that bargaining explains the data better than an assumption of take-it-or-leave-it input contracts. We follow them in modeling the vertical contract as Nash bargaining between the publisher and the retailer over the wholesale price, but extend the demand side of the model to study the impact of vertical contracting with behavioral consumers.

Third, we contribute to the literature on bargaining and vertical integration. The Nash-in-Nash solution concept was first introduced by Horn & Wolinsky (1988). A number of empirical papers have used the Nash-in-Nash solution to study the wholesale model (G. S. Crawford & Yurukoglu, 2012, Gowrisankaran et al., 2015, and Ho & Lee, 2019) and the welfare effects of vertical mergers (G. S. Crawford et al., 2018, Diebel, 2018, Cuesta et al., 2019, and Sheu & Taragin, 2020). We contribute to this literature by comparing the bargaining equilibrium in both wholesale and agency contracts following,

[4]Related work by Li (2021) estimates a structural model to study publishers' optimal wholesale pricing strategy across online and offline channels. We focus on the print format in this study and abstract away from competition between print and electronic formats. Gilbert (2015) provides an overview of recent developments in the e-book industry.

and documenting the fact that whether vertical integration leads to higher or lower joint profits for the firms depends critically on the magnitude of the price reference effect.

This paper proceeds as follows. Section 2.2 presents our model in the content of online book retail market. Section 2.3 describes the background and data. Section 2.4 presents the empirical implementation, the parameter estimates, and the fit of our model. Section 2.5 evaluates prices and profits under alternative vertical contracts and consumer expectations. Section 2.6 concludes.

2.2. Model

In this section, we present a model that predicts: (i) consumers' demand for books; (ii) retail prices set by the retailer; and (iii) the wholesale price negotiated between the publisher and the retailer. The model proceeds in three stages. In stage 1, the publisher and the retailer bargain over the wholesale price; in stage 2, the retailer sets retail prices; and in stage 3, consumers make purchases. We start from the last stage and work backwards.

2.2.1. A Demand Model with Price Reference Effects

We define a product as a book sold on the retailer's platform. The utility consumer i derives from book j at time t is given by

$$u_{ijt} = X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \int \rho(R - \frac{p_{jt}}{l_j})dF_{jt}(R) + \varepsilon_{ijt} \tag{2.1}$$

where X_{jt}, p_{jt}, ξ_{jt} and l_j are the observed characteristics, price, unobserved characteristics and list price of book j at time t. Note that the list price l_j is printed on the back of the book and does not change over time. α and β are demand parameters, and ε_{ijt} is a consumer-book-time specific utility shock. We allow for an outside option with utility $u_{i0t} = \varepsilon_{i0t}$.

The key component of our demand model in Equation (2.1) is the term $\int \rho(R - \frac{p_{jt}}{l_j})dF_{jt}(R)$. It represents the price reference effect, i.e., the reference-dependent gain-loss utility the consumer derives from purchasing book j at price p_{jt} (Kőszegi & Rabin, 2006). Specifically, p_{jt}/l_j is ratio of retail price of book j over its list price and we call it the retail rate. $F_{jt}(R)$ denotes the distribution of consumers' rationally expected retail rates. We call each potential expected retail rate R the reference rate. Then $\rho(R - \frac{p_{jt}}{l_j})$ denotes the gain or loss that a consumer associates with comparing the actual retail rate p_{jt}/l_j to a reference rate R, where $\rho(\cdot)$ is a continuous, weakly increasing gain-loss function. Intuitively, if the actual retail rate is lower than what the consumer expects ($R - \frac{p_{jt}}{l_j} > 0$), the consumer would receive a higher utility, and vice versa. With multiple reference rates in the support of $F_{jt}(R)$, the reference effect term $\int \rho(R - \frac{p_{jt}}{l_j})dF_{jt}(R)$ equals the gain-loss in willingness to pay averaged with respect to the reference rate distribution $F_{jt}(R)$.

We assume that the reference rate distribution is the same across consumers.[5] If consumers expect a single retail rate R with certainty, $\int \rho(R - \frac{p_{jt}}{l_j})dF_{jt}(R)$ would simplify to the gain-loss willingness to pay $\rho(R - \frac{p_{jt}}{l_j})$ associated with a deterministic monetary gain or loss of size $(R - \frac{p_{jt}}{l_j})$.

Our main analysis follows Heidhues & Kőszegi (2008) and Gentry & Pesendorfer (2021a) in specifying the reference function $\rho(x)$ as piece-wise linear, with a potential

[5]Gentry & Pesendorfer (2021b) show that this assumption could be extended to allow for individual reference expectation F_i, accounting for the possibility that consumers weigh prices different from the retailer, as long as it is consistent with the retailer's pricing policy F. Also, it shows that all equilibria invlve the same aggregate reference expectation $F = E_i[F_i]$.

kink at zero to accommodate the fact that consumers may weigh gains and losses differently:

$$\rho(x) = \begin{cases} \delta^+ x & \text{if } x \geq 0 \\ \delta^- x & \text{if } x < 0 \end{cases} \tag{2.2}$$

where the non-negative parameters δ^+ and δ^- represent the changes in willingness to pay that consumers associate with perceived gains and losses, respectively. We say that consumers are **bargain hunters** if they weigh gains more than losses, $\delta^+ > \delta^-$, and **loss averse** if they weigh losses more than gains, $\delta^+ < \delta^-$.

Assuming that ε_{ijt} follows a standard Type I Extreme Value distribution, the market share of book j is

$$s_{ijt}(p) = \frac{\exp\left(X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \int \rho(R - \frac{p_{jt}}{l_j})dF_{jt}(R))\right)}{1 + \sum_{k=1}^{J} \exp\left(X_{kt}\beta - \alpha p_{kt} + \xi_{kt} + \int \rho(R - \frac{p_{kt}}{l_k})dF_{kt}(R))\right)}.$$

With aggregate market share data, the estimating equation can be based on a linear equation of the usual logit form:

$$\log(s_{jt}) - \log(s_{0t}) = X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \int \rho(R - \frac{p_{jt}}{l_j})dF_{jt}(R). \tag{2.3}$$

We will discuss how to estimate $F_{jt}(R)$, potential instrument for the prices, and identification of the reference effects in Section 2.4.

2.2.2. Retailer's Pricing Problem

In this section, we follow the analysis in Gentry & Pesendorfer (2021b). We assume that the cross-price elasticity of book $j's$ price p_{jt} on other books are negligible comparing to the own-price elasticity.[6] Then the retailer's pricing problem could be simplified to a single-product monopoly's pricing problem.[7] Denote π_{jt}^D the retailer's profit from book j at time t, where the superscript D stands for the downstream firm. Then

$$\begin{aligned}\pi_{jt}^D &= (p_{jt} - w_j - c_{jt}^D) \cdot s_{jt}(p) \\ &= (p_{jt} - w_j - c_{jt}^D) \cdot \left(s_{jt}^0 + s_{jt}^0(\log(s_{jt}) - \log(s_{jt}^0))\right)\end{aligned}$$

where w_j is the wholesale price, and c_{jt}^D is the retailer's marginal cost. We assume that the wholesale price does not change over time, but retailer's marginal cost could.[8] The second line of the equation follows from first-order approximation of the exponential function at the base share s_{jt}^0. Denote the retail rate $r_{jt} \equiv p_{jt}/l_j$.[9] Plug in Equation (2.3) and rewrite

[6]This is a non-standard assumption. However, in a logit model, the own price effect of product j is $\partial s_j/\partial p_j = -\alpha s_j(1 - s_j)$ while the cross price effect of on product k is $\partial s_k/\partial p_j = \alpha s_j s_k$. Our data includes over 1 million books and the top seller accounts for only 0.34% of the total quantity sold. Thus, the own-price elasticity is way larger than the cross-price elasticity. Also, it is computationally burdensome to solve the pricing problem given the large number of products. This assumption allows us to derive the retailer's pricing rule analytically and facilitates the estimation.

[7]The retailer does face other retailers selling the same book. Unfortunately, we do not have data from other online platforms or offline stores. Since the retailer in our empirical analysis is the largest online book retailer in the market that has about 40% of the market share and consumers face search costs, we assume that the retailer prices as a monopoly on its platform.

[8]We observe in the data that the wholesale price paid to the publisher does not change over time for a given book. However, the retailer's cost of warehouse, inventory, shipping, depreciation and advertising possibly will change over time.

[9]Since the list price is fixed for a given book, we use the phrase "retail rate" or "retail price" interchangeably. We will emphasize specifically if the distinction is important.

the profit function:

$$\pi_{jt}^D = l_j s_{jt}^0 \cdot (r_{jt} - \underbrace{\frac{w_j + c_{jt}^D}{l_j}}_{c_{jt}}) \cdot (\underbrace{\log(s_{0t}) + X_{jt}\beta + \xi_{jt} + 1 - \log(s_{jt}^0)}_{a_{jt}} - \underbrace{\alpha l_j}_{b_j} r_{jt} + \int \rho(R - r_{jt}) dF_{jt}(R))$$

where $c_{jt} \equiv (w_j + c_{jt}^D)/l_j$, $a_{jt} \equiv \log(s_{0t}) + X_{jt}\beta + \xi_{jt} + 1 - \log(s_{jt}^0)$, and $b_j \equiv \alpha l_j$. Assume that the base share s_{jt}^0 is fixed and omit the subscripts to simplify notation. The retailer's pricing problem is equivalently to solve

$$\max_r \quad \pi(r, F) = (r - c) \cdot D(r, F)$$

where

$$\begin{aligned} D(r, F) &\equiv a - br + \int_{\underline{r}}^{\bar{r}} \rho(R - r) dF(R) \\ &= a - br + \delta^+ \int_r^{\bar{r}} (R - r) dF(R) + \delta^- \int_{\underline{r}}^{r} (R - r) dF(R) \end{aligned} \tag{2.4}$$

with $\underline{r}$ and $\bar{r}$ denoting the infimum and supremum values in the support of the reference distribution F, respectively.

In solving for the retailer's equilibrium pricing distribution, the solution concept we consider requires all prices in the support of the pricing distribution to be optimal given the reference expectation F, so that there is no incentive for the retailer to deviate from the pricing plan. We can view this as a Bayesian Nash Equilibrium between the retailer and the consumers. Specifically, a Bayesian Nash Equilibrium is a retail rate distribution F, such that (i) the retailer maximizes profit given consumers' reference distribution F, and (ii) consumers' reference distribution is consistent with retailer's choice of F.

An alternative solution concept is the commitment solution in which the retailer can commit to a pricing policy ex ante. In this case, the retailer chooses a retailer rate distribution F to maximize expected profit, accounting for the effect that F has on consumers' reference expectations. Gentry & Pesendorfer (2021b) show that the commitment solution requires the retailer to commit ex ante to individually sub-optimal prices. However, it is not clear that there exist formal commitment devices for the retailer in the book retail market. Thus, we choose to use the BNE solution. Appendix B.1.1 compares the two solution concepts in detail.

We analyse the BNE of the pricing game between the retailer and consumers when consumers are bargain hunters who value gains more than losses, i.e. $\delta^+ > \delta^-$.[10] The characterization of the pricing equilibrium follows Gentry & Pesendorfer (2021b) closely. We first show that there does not exist a pure-strategy equilibrium. Thus, the retailer will not adopt a uniform-price policy facing bargain-hunting consumers. We prove this result by contradiction: suppose there exists r^* that maximizes retailer's profit. Optimality requires that the profit is non-decreasing as retail rates approach r^* both from the left and from the right:

$$\lim_{r \nearrow r^*} \partial\pi(r, F)/\partial r = a - br^* + (r^* - c)(-b - \delta^+) \geq 0$$

$$\lim_{r \searrow r^*} \partial\pi(r, F)/\partial r = a - br^* + (r^* - c)(-b - \delta^-) \leq 0.$$

[10]In the demand estimation, we do find that $\delta^+ > \delta^-$ in our data. Thus, we focus on the case that consumers are bargain hunters. When consumers are loss aversion ($\delta^+ < \delta^-$), Gentry & Pesendorfer (2021b) show that uniform pricing is optimal for the retailer. The fact that we observe fluctuations in retailer's pricing decision is suggestive evidence that $\delta^+ > \delta^-$, even though we do not impose this assumption in the demand estimation.

However, the inequalities hold only if $\delta^+ \leq \delta^-$, which contradicts the assumption that $\delta^+ > \delta^-$. Intuitively, it is because given any potential r^*, the retailer could always increase its profit by setting a retail rate slightly higher or lower than r^*.

We now characterize the mixed-strategy BNE. Denote $r_e \equiv \int_{\underline{r}}^{\bar{r}} R\, dF(R)$ the expected retail rate under the reference distribution F. We can rewrite $D(r, F)$ in Equation (2.4) as

$$\begin{aligned} D(r,F) &= a - br + \delta^+ \int_r^{\bar{r}} (R-r)dF(R) + \delta^- \int_{\underline{r}}^{r} (R-r)dF(R) \\ &= (a + \delta^- r_e) - (b + \delta^-) \cdot r + (\delta^+ - \delta^-) \cdot \int_r^{\bar{r}} (R-r)dF(R) \\ &= a' - b'r + \Delta \int_r^{\bar{r}} (R-r)dF(R) \end{aligned} \tag{2.5}$$

where $a' \equiv a + \delta^- r_e, b' \equiv b + \delta^-$, and $\Delta \equiv \delta^+ - \delta^-$. The second line is given by adding and subtracting $\delta^- \int_r^{\bar{r}} (R-r)dF(R)$ on the right-hand side of the equation.

For F to be a mixed-strategy equilibrium, all retail rates must yield the same profit for the retailer. In particular, every retail rate must yield the same profit as the supremum rate $\bar{r}$: $\pi(r, F) = \pi(\bar{r}, F)$, $\forall r \in [\underline{r}, \bar{r}]$. Dividing both profit functions by $r - c$, we obtain the equivalent indifference condition:

$$D(r,F) - \frac{(\bar{r} - c) \cdot (a' - b'\bar{r})}{r - c} = 0, \qquad \forall r \in [\underline{r}, \bar{r}].$$

Since the left-hand side is a constant that equals to zero over the support of r, we must have the derivative with respect to r equals 0 for all r in the support. Using Equation

(2.5), we have

$$-b' - \Delta(1 - F(r)) + \frac{(\bar{r} - c)(a' - b'\bar{r})}{(r - c)^2} = 0, \qquad \forall r \in [\underline{r}, \bar{r}].$$

Using the fact that $F(\bar{r}) = 1$, $F(\underline{r}) = 0$ and $\frac{\partial \pi(r,F)}{\partial r}\big|_{r=\underline{r}} = 0$, we can solve for the BNE retail rate distribution F:

$$F(r) = 1 - \frac{b + \delta^-}{\delta^+ - \delta^-}\left(\frac{(\bar{r} - c)^2}{(r - c)^2} - 1\right) \tag{2.6}$$

defined on the support $[\underline{r}, \bar{r}]$ where

$$\bar{r} = \frac{a + \delta^- r_e}{2(b + \delta^-)} + \frac{c}{2}, \quad \underline{r} = \frac{a + \delta^+ r_e}{2(b + \delta^+)} + \frac{c}{2} \tag{2.7}$$

and

$$r_e = \frac{a(\sqrt{b + \delta^+} - \sqrt{b + \delta^-}) + c((b + \delta^+)\sqrt{b + \delta^-} - (b + \delta^-)\sqrt{b + \delta^+})}{\delta^+\sqrt{b + \delta^-} - \delta^-\sqrt{b + \delta^+}}. \tag{2.8}$$

The detailed proof is in Appendix B.1.2. We say that the above equations characterize the unique BNE.

To sum up, we show that retailer's equilibrium pricing rule is a mix-strategy pricing distribution that depends on the demand parameters and retailer's costs. More importantly, it depends on δ^+ and δ^-, which represent the changes in willingness to pay that consumers associate with perceived gains and losses. In the limit when δ^+ and δ^- converge to zero, the retail rate distribution degenerates to a single limit price, which is the usual monopoly price solution of $\frac{a}{2b} + \frac{c}{2}$.

2.2.3. Bargaining over the Wholesale Price

Now we move to the first stage of the model where the publisher and the retailer bargain over the wholesale price by extending Gentry & Pesendorfer (2021b)'s analysis to allow for the bargaining between two parties. Retailer's profit from selling book j is

$$\pi_j^D = (p_j - w_j - c_j^D) \cdot s_j(p)$$

where p_j is the average retail price of book j, w_j is the wholesale price, c_j^D is the retailer's average marginal cost of book j, and $s_j(p)$ is the average market share of book j. The publisher's profit from selling book j to the retailer is

$$\pi_j^U = (w_j - c_j^U) \cdot s_j(p)$$

where c_j^U is the publisher''s marginal cost, and superscript U stands for the upstream firm. We assume that the wholesale price is determined through Nash bargaining between the upstream and downstream firms associated with book j. The Nash product is

$$\Omega_j(w) = (\pi_j^U)^\lambda (\pi_j^D)^{1-\lambda}$$

where $\lambda \in [0,1]$ is the bargaining weight of the upstream firm. Although we do not index λ to keep the notation simple, we will allow λ to vary with publisher sizes in the estimation. Also, we assume that the disagreement payoffs are zero. Since the wholesale price is bargained separately for each book and there is no cross-price effect in retailer's pricing problem, the Nash bargaining can be solved at the book level.

2.2.4. Numerical Examples

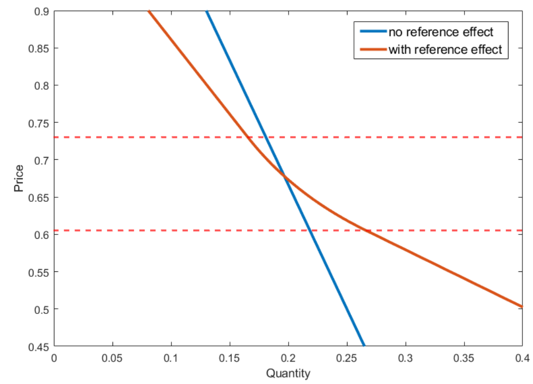
We illustrate the above model with numerical examples. In Figure 2.2, we consider the case in which consumers are bargain hunters, characterized by parameters $a = 0.4, b = 0.3, c = 0.4, \delta^+ = 1$, and $\delta^- = 0.2$. Figure 2.2(a) plots the demand curve in Equation (2.4) implied by a continuous reference distribution F which is the equilibrium reference distribution characterized by Equation (2.6). For comparison, it also plots the canonical linear demand curve $a - br$. Dashed lines in all sub-figures of Figure 2.2 correspond to the upper and lower bound of the retail rate distribution $\bar{r}$ and $\underline{r}$, respectively in Equation (2.7).

Figure 2.2(a) shows that the demand curve is more price-elastic than the canonical linear demand curve due to the price reference effect. The demand curve transitions smoothly from a slope of $-b - \delta^-$ at prices $r \geq \bar{r}$ to a slope of $-b - \delta^+$ at prices $r \leq \underline{r}$. The demand curve is convex since consumers are bargain hunters, namely $\delta^+ > \delta^-$. Figure 2.2(b) plots the density function of the optimal price distribution set by the retailer in equilibrium (density function of Equation (2.6)). The density function is decreasing and convex. Figure 2.2(c) plots the retailer's equilibrium profit function. It shows that when consumers have reference expectations described by the equilibrium reference distribution F, the retailer is indifferent between any retail rate in the support $[\underline{r}, \bar{r}]$. This confirms that the equilibrium price distribution is in fact a mix-strategy equilibrium.

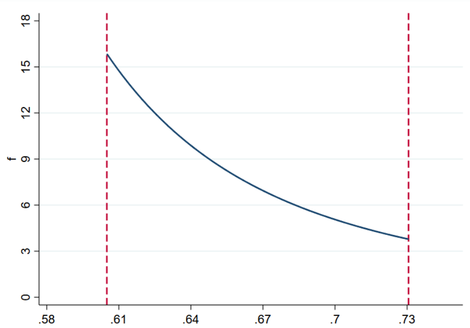
To illustrate how the bargaining parameter affects equilibrium outcomes, Figure 2.3 plots the equilibrium quantity, wholesale price, mean retail price, and profits as a function of the bargaining parameter λ. We set $a = 0.4, b = 0.3, c^U = 0.4, c^D = 0, \delta^+ = 1, \delta^- = 0.2$.

Figure 2.2. Demand, Retail Rate Density and Profit Function with Price Reference Effects

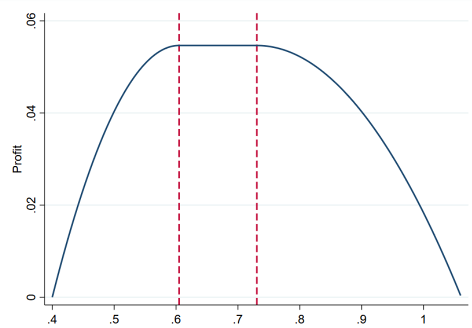
(a) Demand



(b) Retail Rate Density



(c) Profit Function



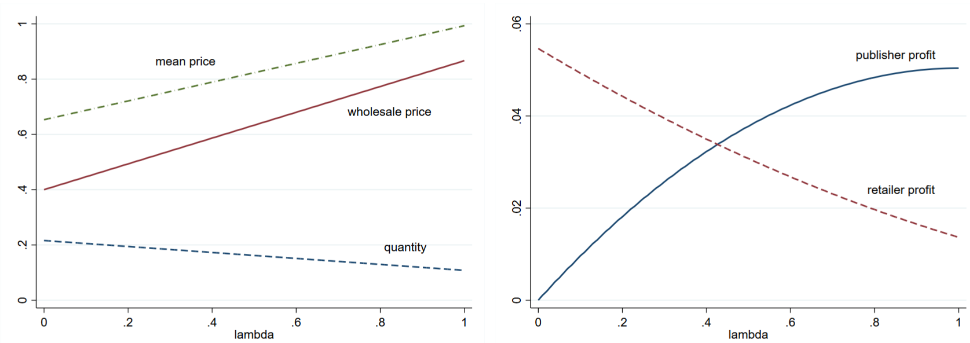
Notes: Equilibrium demand function, retail rate density and profit function when bargain-hunting consumers have equilibrium reference expectations. $a = 0.4, b = 0.3, c = 0.4, \delta^+ = 1,\ \delta^- = 0.2$.

Figure 2.3(a) shows that the more bargaining power the publisher has, the higher the negotiated wholesale price, which leads to higher mean retail prices. Also, the quantity sold decreases as the mean retail price increases. Figure 2.3(b) shows that publisher's profit is increasing in its bargaining power while the retailer's profit is decreasing in λ. Thus, the relative profits for the publisher and the retailer under Nash bargaining depend on their relative bargaining powers, similar to the canonical model without the price reference effect.

Figure 2.3. Quantity, Prices and Profits as a Function of the Bargaining Parameter

(a) Quantity and Prices

(b) Profits



Notes: Equilibrium quantity, prices and profits under different bargaining parameters when bargain-hunting consumers have equilibrium reference expectations. $a = 0.4, b = 0.3, c^U = 0.4, c^D = 0, \delta^+ = 1, \delta^- = 0.2$.

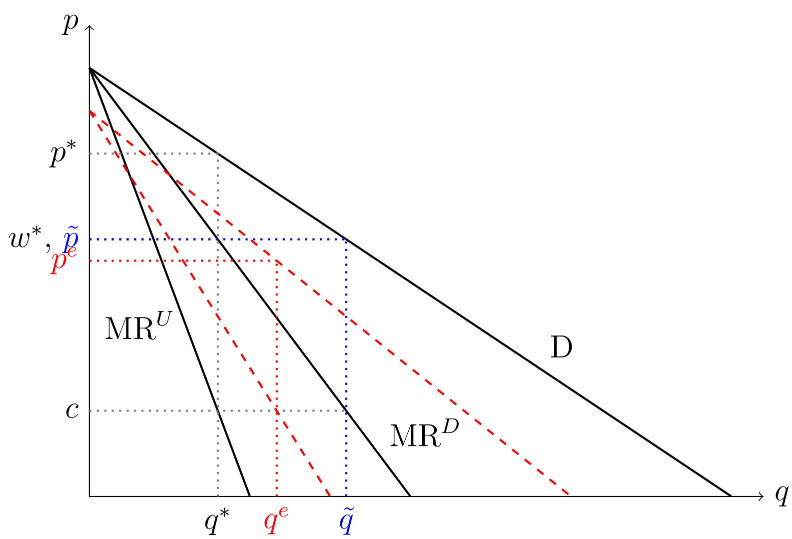
2.2.5. Analysis of Vertical Merger

We illustrate how vertical merger affects firms' joint profits when there exist price reference effects. Consider the case of vertical merger of bilateral monopoly. The merged firm maximizes the joint profits given the upstream and downstream marginal costs:

$$\pi_j^M = (p_j - c_j^U - c_j^D) \cdot s_j(p).$$

Figure 2.4 illustrates the price and quantity effects of vertical merger of bilateral monopoly, with and without the price reference effect. Assume that the upstream firm has a constant marginal cost c and the downstream firm's marginal cost is zero. First, we consider the case without price reference effects. The three downward-sloping solid lines represent the demand curve, the marginal revenue curve of the downstream firm, and the marginal revenue curve of upstream firm. The equilibrium quantity q^* is determined by the interaction of the marginal cost and marginal revenue curves of upstream firm. Then the equilibrium wholesale price and retail price are w^* and p^*, respectively. After the vertical merger, the marginal revenue curve of the downstream firm is now the marginal revenue curve of the joint firm. Thus, post-merger quantity and retail price are now $\tilde{q}$ and $\tilde{p}$.

Figure 2.4. Vertical Merger of Bilateral Monopoly



Notes: Illustration of price and quantity effects of the vertical merger of bilateral monopoly, without and with price reference effects.

The classic result of vertical merger states that joint profits would be higher after the merger due to the elimination of double marginalization. However, this result may not hold when there exist price reference effects. For illustration purpose, we ignore the mixed-strategy price distribution for now and focus on the average price and quantity. We later state the result formally with the mixed-strategy price distribution in Proposition 1. Consider the case with price reference effects. Since the retail price tends to be lower after the merger, consumers would update their belief and expect lower prices after the merger. As a consequence, the demand curve would shift downward. Then the retail price updates in response to the change of demand, and this process iterates until the new equilibrium is reached. In Figure 2.4, it could be the case that in the post-merger equilibrium, the demand and marginal revenue curves are the two downward-sloping dashed lines. Then quantity and retail price are q^e and p^e. In this case, the decrease in price overshadows the increase in quantity, resulting in the lower post-merger joint profits.

Formally, Proposition 1 shows that when consumers are pure bargain hunters ($\delta^+ > \delta^- = 0$), whether the joint profits are higher or lower after the merger depends critically on the magnitude of the price reference effect.[11]

Proposition 1. *When $\delta^- = 0$, there exist $\delta^{+^\star} > 0$ such that the post-merger joint profits are higher than the pre-merger joint profits when $\delta^+ < \delta^{+^\star}$, and the opposite is true when $\delta^+ > \delta^{+^\star}$.*

The proof is in Appendix B.1.3. To give a numerical example, Figure 2.5 plots the pre- and post-merger joint profits for different values of δ^+. When δ^+ and δ^- are both zero, the model degenerates to a linear demand model. In this case, the post-merger joint profits must be higher than the pre-merger joint profits as shown in Figure 2.5 when $\delta^+ = 0$. However, as δ^+ becomes significantly large, the post-merger joint profits would be lower than the pre-merger joint profits.

[11] We set δ^- to be zero for illustration purpose. δ^- could be positive in general. When $\delta^- > 0$, it is possible that the post-merger joint profits are lower than the pre-merger joint profits for all values of $\delta^+ > \delta^-$.

Figure 2.5. Pre- and Post-Merger joint profits with Price Reference Effects

Notes: Illustration of pre- and post-merger joint profits for different values of δ^+. $a = 0.4, b = 0.3, c^U = 0.4, c^D = 0, \lambda = 0.3, \delta^- = 0$.

2.3. Background and Data

2.3.1. Institutional Background

The online book retail market, as other digital goods, is growing rapidly in the recent decade. According to the industry report from Grand View Research (2020), the global online book retail market size is around \$17.7 billion in 2019, and is foreseeing an annual growth of 5.8% through 2027. The book retail market in China is witnessing a similar trend. Figure B2 shows that in the recent 10 years, the book retail market is growing at a rate of around 12% annually, and the increase comes entirely from the online channel. Back in 2010, the online channel accounts for only 10% of the sales. With the growth of secure online payment and shipping capacity, the online channel is growing on average at a rate of 30% annually and accounts for 62% of the sales in 2019.[12]

[12]During the same period, the list price of books in China is growing as well. Figure B3 shows the median list price of books published from 2010 to 2019. Compared with books published in the US, the list price

Established in 1999 as an online book retail website, DangDang.com is the largest online book retailer in this market. It has a steady 40% market share since 2007, with annual revenue of $1.5 billion in 2018. A typical transaction of books on DangDang.com follows three stages. First, DangDang negotiates the wholesale rate with each publisher for the majority of the books published in a year. The wholesale rate is the ratio of wholesale price over the list price. It can vary across different publishers and book categories, and is expected to be renewed annually. The wholesale rate does not change frequently because of the large workload and re-negotiation cost.[13] However, DangDang recognizes books to be published that have the potential to become a best seller. In such case, DangDang will attempt to negotiate a lower wholesale rate with the publisher, which could be different from the preset wholesale rate. After the negotiation, DangDang purchases the books and stores them in its warehouses.

The second stage is for DangDang to set retail prices. Prices are set for each book, but discounts can be applied to a broader sub-category of books through coupons. Best sellers may experience daily price changes as DangDang tries to capture short-term demand shocks. Figure B4 presents the daily retail rate of the best-selling book *To Live* from Aug 12, 2017 to Sept 2, 2019. It is clear that DangDang actively sets prices at a high frequency.[14]

of books in China shows a clear upward trend, even after adjusted for the Consumer Price Index (CPI). Our model accounts for this increase by using the retail rate as reference points instead of directly using the retail price.

[13]Books sold on DangDang.com are published from more than 500 publishers nationwide. The publication market is highly un-concentrated in China, with the top 10 publishers accounting for 24% of market share as shown in Table B3. For comparison, the top 5 publishers in the US account for over 80% of market share.

[14]DangDang may also respond to other retailers' prices. However, we are not able to model the price competition among the retailers due to the lack of data. Instead, we assume that DangDang behave as a monopoly facing residual demand when making the pricing decisions. This could be justified if consumers are royal to the platform and encounter search cost when they switch across platforms.

Lastly, consumers make purchases. They can search on DangDang.com by book title, author, publisher name or any other keywords included in the product information. Books are ranked by the matching quality. On the book specific webpage, more book information including a brief introduction and table of contents is provided. Consumers observe the list price, retail price, and retail rate of the book. After making payments, consumers typically receive the book within 3 days.[15]

2.3.2. Data

Our main data source is the administrative sales data from DangDang.com. We complement the sales data with book characteristics scraped from Douban.com, the largest book review and rating platform in China. Different from many traditional retailers that only provide the retail price of the product, the online book retailer DangDang.com chooses to present additional price information strategically. Figure 2.6 shows a webpage screenshot of the best seller *To Live* on DangDang.com. It shows a retail price of 19.3 CNY and a list price of 28 CNY. More importantly, it presents explicitly the retail rate of 69%. Recall that the retail rate is the term p_{jt}/l_j in Equation (2.1). Since consumers may care not only about the retail price, but also about the difference between what they are paying and what they expect to pay, we can separately identify the direct price effects from indirect price reference effects using variation in consumers' reference expectations induced, for example, by variation in firm's costs over time.

The full sample from DangDang.com contains 81 million title-date level observations for the period from January 2017 to August 2019. It covers more than 1 million unique

[15]Consumers would receive free shipping if the purchase amount is over 49 CNY, which is about the retail price of one and half books on average.

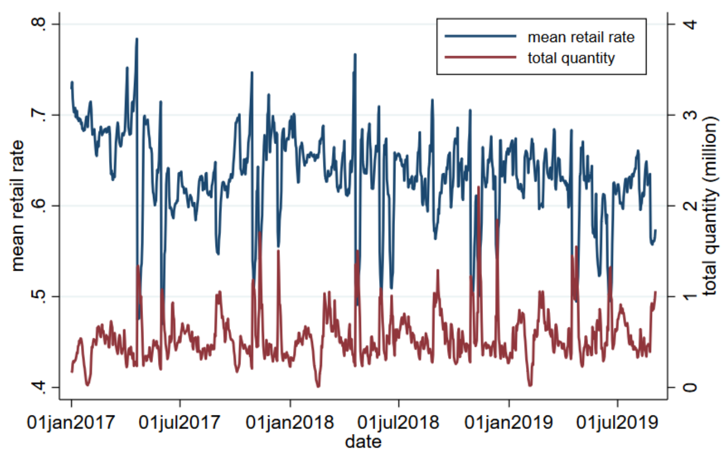
Figure 2.6. Webpage Example

Notes: Webpage example of a book (the best seller *To Live*) sold on DangDang.com. The screenshot was taken on Jan 21, 2021 from http://product.dangdang.com/25137790.html.

titles from over 500 publishers. For each book identified by the International Standard Book Number (ISBN), we observe the daily quantity sold and the retail price. This feature allows us to estimate demand with minimum measurement error in quantity as opposed to backing out quantity from sales rank data (Chevalier & Goolsbee, 2003). Also, we directly observe the bargained wholesale price for each title, which means we do not need to make additional assumptions to separately identify the bargaining parameter (De los Santos et al., 2021). Additionally, we observe the list price, name of the publisher, category and other book characteristics.

Figure 2.7 plots the daily mean retail rate and total quantity sold on DangDang.com from Jan 1, 2017 to Sept 2, 2019. The retail rate is on average 0.63 and around 0.55 million books are sold each day. There is substantial variation across dates during our sample period. The sharp dips in retail rate indicate large sales events. We observe a weak downward trend in the mean retail rate, and a strong negative correlation between the retail rate and the quantity sold. There is also seasonality in demand. For example, the low quantity sold around February in each year corresponds to the Chinese New Year. Note that Figure 2.7 plots the average retail rate across all books and is affected by the composition effect. To highlight the mixed strategy that DangDang is playing, Figure B4 presents the daily retail rate of the best-selling book *To Live*. It is clear that DangDang actively sets prices at a high frequency and there is substantial variation in the retail price across time.

Figure 2.7. Retail Rate and Quantity



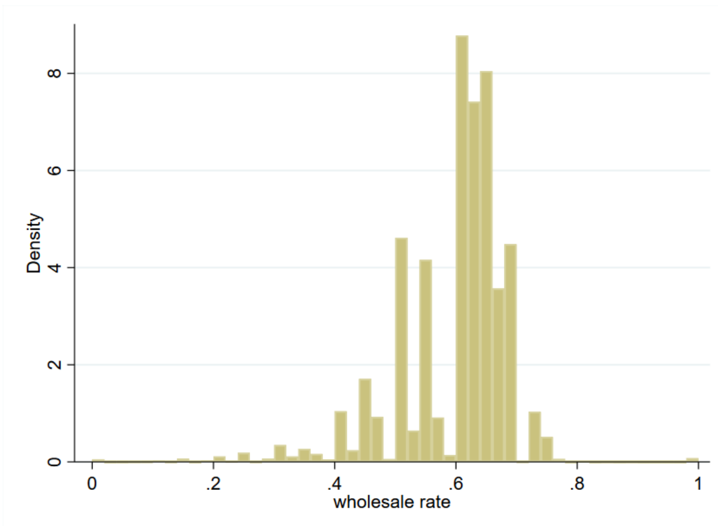
Notes: This figure plots the mean retail rate (left axis) and total quantity (right axis) of books sold on DangDang.com from Jan 1, 2017 to Sept 2, 2019. Retail rate is the ratio of retail price over the list price.

Figure 2.8 plots the histogram of wholesale rate of books sold on DangDang.com. There is large variation in the wholesale rate across books. DangDang is able to bargain a wholesale rate as low as 20% of the list price, while the highest wholesale rate is over 80%. For most of the books, DangDang pays between 50% to 66% of of list price to the publishers.

For the purpose of estimation, we focus on the best sellers that have more than 5,000 copies sold during the sample period.[16] In addition, we focus on books that were published

[16]As we have discussed in the institutional background, DangDang negotiates the wholesale rate with each publisher for the majority of the books published in a year. However, DangDang will negotiate a

Figure 2.8. Wholesale Rate



Notes: This figure plots the histogram of wholesale rate of books sold on DangDang.com from Jan 1, 2017 to Sept 2, 2019. Wholesale rate is the ratio of wholesale price over the list price.

after January 1, 2017, so that we are able to observe the sales from the first day since the book is published.[17] We exclude outliers that have a list price either less than 10 CNY or more than 500 CNY. The selection procedure leaves us with an estimation sample of 6,228 unique titles with 3 million daily observations. With 0.6% of the titles, these books account for 25% of total quantity sold on the platform.[18]

separate wholesale rate with the publisher for the potential best seller. Thus, our model assumption that the wholesale rate is negotiated for each book is valid for the best-selling books.

[17]There is no significant difference in the book characteristics for books published before and after 2017. The cutoff is chosen because we observe sales data since January 1, 2017.

[18]In our estimation sample, one-third of the books have an e-version. The e-version sales for those books are about 8% of sales of the print version at the median. In this paper, we abstract from the effect of e-version sales on print sales. In a separate paper, we study the retailer's pricing strategy given the substitutability or complementarity of the two versions.

Table 2.2. Summary Statistics

	Obs	Mean	S.D.	p25	p50	p75
	Title level					
Publish date	6,228	2017.98	0.67	2017.41	2017.92	2018.45
Quantity	6,228	21,716	50,430	6,551	9,725	17,821
List price	6,228	53.38	40.39	33.80	43.55	59
Retail rate	6,228	0.578	0.084	0.527	0.574	0.633
Wholesale rate	6,228	0.522	0.097	0.490	0.530	0.600
	Title-date level					
Sale date	3,133,844	2018.68	0.64	2018.20	2018.75	2019.22
Publish date	3,133,844	2017.74	0.57	2017.33	2017.58	2018.16
Quantity	3,133,844	43.16	280.2	4	11	28
List price	3,133,844	52.49	40.23	34	42	58
Retail rate	3,133,844	0.581	0.134	0.498	0.570	0.687
Wholesale rate	3,133,844	0.523	0.092	0.490	0.530	0.600

Notes: This table presents summary statistics of the estimation sample of books sold on DangDang.com from Jan 1, 2017 to Sept 2, 2019. List prices are in the unit of CNY. 1 USD $\approx$ 6.6 CNY in 2018. Retail and wholesale rates are the ratio of retail and wholesale prices over the list price.

Table 2.2 presents the summary statistics of the estimation sample.[19] The top panel shows that at the title level, the average book was published near the end of 2017, and has 22 thousand copies sold during the sample period. The distribution of quantity sold is right skewed—a few best sellers have significantly larger quantity sold than other books, which is the typical sales pattern in the book retail industry. The average list price is around 53 CNY ($\approx$\$8), and the mean retail rate is 58% of the list price. The wholesale rate is on average 52% of the list price, lower than the 59% in the full sample. This suggests that the retailer is able to negotiate a lower wholesale rate for the best-selling

[19]For comparison, Table B1 presents the summary statistics of the full sample. On average, the best sellers have lower list price, retail rate and wholesale rate. This is because the retailer usually tends to negotiate a lower wholesale rate with the publisher for potential best sellers, which leads to lower retail rates. Also, it is natural that books with lower prices are more likely to become best sellers.

books. The bottom panel of Table 2.2 presents the title-date level summary statistics. On average, 43 copies are sold for each title in each day, but the distribution of quantity sold is again right skewed. The list price, retail rate and wholesale rate are similar to those at the title level.

2.3.3. Discussion on Demand Dynamics

As illustrated above, our model predicts that the retailer utilizes a mix-strategy pricing policy, which can explain the fluctuation of retail prices in the data. However, one may wonder how our static reference-effect model compares to a dynamic demand model of search for low prices. Indeed, several papers in economics and marketing literature have documented demand dynamics due to consumer search and stockpiling behaviors (Pesendorfer, 2002, Hendel & Nevo, 2006a,b, Özer & Zheng, 2016, Zhang et al., 2018, and Chevalier & Kashyap, 2019). Notably, Hendel & Nevo (2013) show that the pricing strategy of regular prices with periodical sales is consistent with retailer's intertemporal price discrimination motive. In our case, the temporary price reductions could enable the retailer to discriminate between more price-sensitive consumers, who are more patient and willing to wait for low prices, and less price-sensitive consumers, who are not willing to wait.

To test for this potential dynamic demand explanation, we follow Hendel & Nevo (2013) in displaying the average quantity of books sold during sale and non-sale periods in Table 2.3. A sale is defined as any retail rate equal or below 0.5, which occurs 30% (23%) of the time at the daily (weekly) level.[20] At the daily level, the table shows that

[20]The results are robust to alternative definitions of a sale, e.g., cutoffs at 0.4 or 0.6.

Table 2.3. Average Quantity and Sale Events

Daily	$S_{t-1} = 0$	$S_{t-1} = 1$	
$S_t = 0$	39.6 (61.3%)	29.7 (8.5%)	37.5 (69.8%)
$S_t = 1$	50.5 (8.5%)	61.1 (21.6%)	56.3 (30.1%)
	40.9 (69.8%)	52.3 (30.1%)	
Weekly	$S_{t-1} = 0$	$S_{t-1} = 1$	
$S_t = 0$	258.7 (67.5%)	205.0 (9.5%)	256.5 (77.0%)
$S_t = 1$	331.9 (9.9%)	358.9 (13.1%)	345.7 (23.0%)
	268.1 (77.4%)	294.4 (22.6%)	

Notes: This table presents the average quantity of books sold using our estimation sample. A sale is defined as any retail rate equal or below 0.5. The numbers in paraphrases report the percentage of the time the sale and non-sale events (or their combination) occur.

during sales periods, the quantity sold is higher (56.3 versus 37.5, or 1.5 times more). However, we find that the quantity sold is not lower if a sale was held in the previous day (52.3 versus 40.9, or 28% higher). Similar results hold at the weekly level. Thus, the quantity pattern does not fit the wait-for-low-price story, which predicts that the quantity sold should be lower following a sale.

More importantly, the pricing pattern is also different from what is predicted by an intertemporal price discrimination model. Numbers in the parentheses in Table 2.3 report the percentage of the time the sale and non-sale events (or their combinations) occur. We find that prices are persistent at the daily level—a sale is more likely to be followed by a sale (21.6% versus 8.5% of the time), and a non-sale is more likely to be followed by a non-sale (61.3% versus 8.5% of the time), which is different from the periodical high-low pricing pattern in Hendel & Nevo (2013). Similar results hold at the weekly level. Figure B5 illustrates the difference between a periodical high-low pricing pattern and a persistent pricing pattern. We argue that the persistent pricing pattern could be explained by cost correlations across time, which motivates our instrument for prices (see more discuss in the demand estimation).

Our model also relates to the literature on consumer learning in both economics and marketing (G. S. Crawford & Shum, 2005 and Zhao et al., 2013). Broadly speaking, the learning models conceptualizes reference effects in terms of an adaptive reference point formed on the basis of past purchases. Thus, reference prices adjust when purchases are made, implying that a specific price can be counted as a gain or a loss depending on the prior purchase history. In contrast, we follow Kőszegi & Rabin (2006) in assuming that gain-loss expectations of a specific price are held fixed across time and are formed based on rational expectations of the equilibrium price distribution. Due to the lack of personal purchase history data, we are not able to distinguish between the two models. Instead, we argue that our model provides a static and computationally simple framework to capture consumers' behavioral motives. Gentry & Pesendorfer (2021a) show that the price reference effect generates improvements in the model's ability to predict out-of-sample choice behavior in a fully dynamic model that incorporates limited attention and forward-looking search. Accommodating dynamic motives is beyond the scope of the current paper and we will leave it for future research.

2.4. Estimation and Results

2.4.1. Demand Estimates

We estimate the demand side of the model using data on quantities, prices, and product characteristics. Since we assume that the model has a logit structure, we estimate

(2.9)

$$\log(s_{jt}) - \log(s_{0t}) = X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \delta^{+} \int\limits_{R \geq \frac{p_{jt}}{l_j}} (R - \frac{p_{jt}}{l_j}) dF_{jt}(R) + \delta^{-} \int\limits_{R \leq \frac{p_{jt}}{l_j}} (R - \frac{p_{jt}}{l_j}) dF_{jt}(R).$$

Equation (2.9) follows directly from Equation (2.3) and the piece-wise linear assumption in Equation (2.2). The demand estimation follows a two-step procedure. The reference rate distribution $F_{jt}(R)$ is estimated in the first step, then the remaining parameters are estimated in the second step using Equation (2.9) where $F_{jt}(R)$ is replaced with the first-step estimate $\hat{F}_{jt}(R)$.

Our model has the feature that the price of the book p_j enters into the utility both directly through αp_{jt} and indirectly through p_{jt}/l_j in the reference effects. Gentry & Pesendorfer (2021a) show that when choices under $K \geq 3$ distinct reference rate distributions $F^1_{jt}, \ldots, F^K_{jt}$ are observed, the reference effects can be separately identified from the direct price effect. This identification argument holds when consumer preferences are constant but the price distribution changes over time.[21] In our setting, when marginal cost of the retailer changes over time, consumers are exposed to different reference distributions. Since an individual consumer has a negligible effect on the supply side pricing decisions, each price distribution F^k_{jt} is determined in part as a random process driven

[21]Note that the list price does not change over time for a given book. The variation of reference distributions come solely from the change in prices across time.

by random cost realizations that are exogenous from the perspective of the consumer. Therefore, the variation of reference distributions does not lead to endogeneity concerns when we estimate demand.

Empirically, for each choice environment $k = 1, \ldots, K$ faced by the consumers, our data contain many observations of prices and choices, so that we can consistently estimate both the reference distribution F_{jt} and the choice probability s_{jt}. The empirical distribution of prices will provide consistent estimates of the reference distribution, while the empirical frequency distribution of choices will then provide consistent estimates of choice probability.

Following Gentry & Pesendorfer (2021a), we partitioned prices of each book into $K = 3$ equal-length reference periods to allow for reference expectations to vary over time. The results are similar when we partition prices into four periods, or optimally chosen periods based on the price volatility. We estimate the empirical reference rate distribution $\hat{F}_{jt}(R)$ and then calculate the reference effect terms for each book. Figure B6 shows the distribution of reference effect terms pooled across all books. The positive part corresponds to the term that multiplies δ^{+} while the negative part corresponds to the term that multiplies δ^{-}.

In the second step, we estimate the remaining parameters of demand. Table 2.4 presents the demand side estimates. In all specifications, we include book fixed effects and date fixed effects, and cluster standard errors at the book level. In specification (1), we do not control for price endogeneity, which means that we estimate the demand using OLS. The price coefficient is highly significant but the demand is relatively inelastic—the average own-price elasticity is -1.138. This is due to the fact that we estimate demand using data at the daily level and consumers are relatively less responsive to price changes in the short run relative to the long run.[22]

Table 2.4. Demand Estimates

	(1) OLS	(2) IV	(3) OLS	(4) IV
Price	−0.0367***	−0.0315***	−0.0138***	−0.0112***
	(0.00103)	(0.00122)	(0.000753)	(0.00119)
δ^{+}			2.556***	2.678***
			(0.0845)	(0.0975)
δ^{-}			1.659***	1.804***
			(0.0676)	(0.0759)
Book FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Mean Own-price elas	−1.138	−0.977	−1.141	−1.106
First-stage F Stat		3909.36		629.80
R-squared	0.610		0.618	
Obs	2,217,222	2,214,612	2,217,222	2,214,612

Notes: This table reports the estimated coefficients of the model in Equation (2.9). The unit of observation is a book-date. The dependent variable is log of quantity sold. All specifications include book and date level fixed effects. The IV specifications use the average lagged prices in the previous week as an instrument for price. Standard errors are clustered at the book level and presented in parentheses below the coefficients. *p<0.1, **p<0.05, ***p<0.01.

The unobserved characteristic ξ_{jt} in demand captures unobserved quality, which is likely to be correlated with the price. Since all of our specifications include book-level fixed effects, the book-specific variation in unobserved quality that does not vary over time is captured by the book dummies. However, the book-level fixed effects will not pick up variation in prices due to changes in unobserved quality over time. For instance, a

[22]The literature has found that Amazon also faces relatively inelastic demand for books, ranging form −0.4 to −1 (Chevalier & Goolsbee, 2003, Reimers & Waldfogel, 2017, and De los Santos & Wildenbeest, 2017).

favorable review on national television may lead to an increase in both demand and retail prices. To deal with the unobserved quality changes over time, we estimate the model by two-stage least squares.

Ideally, we would like to use the BLP-type instruments for the prices. Unfortunately, the BLP-type instruments are difficult to apply in this context since book attributes do not explain much of the variance in sales and demand. Hausman instruments are not suitable either since there is no regional price variation in this market. However, an instrument that is available in this context is the lagged price (Villas-Boas & Winer, 1999). This instrument has been used in other markets in which it is difficult to use traditional instruments such as the market for console video games (Shiller, 2013). We argue that lagged price is a valid instrument in this setting because first, the lagged price is correlated with current price due to cost correlation across time. Even though the wholesale price is fixed for a given book, other costs of the retailer are persistent across time, e.g., the cost of warehouse, inventory, shipping, depreciation and advertising. Second, the persistent cost shocks are likely to be uncorrelated with the transitory demand shocks. For example, the cost of warehouse does not change because of a temporary increase in demand. In specification (2), we use the average price of the previous week as instrument for the current price and the first stage F-statistic is large. The estimated price coefficient decreases in magnitude when using the lagged price as instrument.

Next, we include price reference effects in specifications (3) and (4). We find that including price reference effects reduces direct price coefficients by around 63% and the coefficients of the reference effects are statistically significant. This suggests that reference effects are picking up a substantial fraction of variation in indirect utility due to the direct price effect. The magnitude of reference effects is sizable. A one standard deviation (0.06) increase in the reference expectation would lead to a 16% (0.06×2.678) increase in the quantity sold on average. Also, our estimates suggest that consumers are bargain hunters ($\delta^+ > \delta^-$). In other words, consumers react strongly and asymmetrically to perceived price savings, while being less so to perceived price losses. This is consistent with the finding of Gentry & Pesendorfer (2021a), which suggests that in the context of small repeated purchases, consumers may act as bargain hunters who perceive small gains as more salient than small losses.

For robustness checks, Table B2 presents additional demand estimates that include squared price or nested shares of a book in its own category. The results of the reference effects are similar to our main estimates, implying that the reference effects are not just capturing curvature of the demand curve and are robust to alternative substitution assumptions.

To explore the heterogeneity of reference effects across different book categories, we also estimate the demand incorporating interactions of reference effects and book category dummies. Figure B7 plots the estimated δ^+ and δ^- separately for each category. There is moderate level of heterogeneity in the price reference effect for books in different categories, but δ^+ is larger than δ^- on average and the estimates are similar to our main results.

2.4.2. Supply Estimates

Taking the estimated demand parameters as given, we next describe how to estimate the supply side. Our approach first uses the equilibrium condition of the retailer's pricing problem to back out the marginal cost of the retailer. Then we use the equilibrium condition of the bargaining model and observed book characteristics that correlate with the publisher's marginal cost to get estimates of the publisher's marginal cost and the bargaining parameter.

We assume that the average observed retail rate equals to the expected equilibrium retail rate given by Equation (2.8). Thus, we can back out the marginal cost c_{jt}^D of the retailer for book j for each period given demand estimates. This approach is feasible because we observe the wholesale price. Also, the identification of the retailer's marginal cost follows directly from the equilibrium condition of the model.[23]

Moving backwards to the bargaining between the publisher and the retailer, given the bargaining parameters, we are able to back out the publisher's marginal cost c_j^U from the wholesale price w_j using the equilibrium condition of the bargaining model. This allows us to estimate a linear marginal cost equation in which c_j^U depends on observed product characteristics and an error term, i.e., $c_j^U = Z_j\gamma + \eta_j$. We minimize the sum of squared residuals of this regression across books to solve for the bargaining parameters. In the empirical estimation, we use the list price as the observed product characteristic.[24] Also, we allow the bargaining parameter to vary across publisher sizes. This is because larger publishers may have higher higher bargaining power.

[23]Alternatively, we could jointly estimate demand and supply sides and to use covariance restrictions to deal with endogeneity concerns (MacKay & Miller, 2019 and De los Santos et al., 2021).

[24]Naturally, the list price is highly correlated with the publisher's marginal cost. Industry sources show that publishers usually use the cost-pricing method to set the list price such that the marginal cost is around 40% of the list price (https://www.douban.com/note/653885569/). We have tried other characteristics such as number of pages, number of weeks since release, number of reviews and star ratings. However, they are not as highly correlated with the marginal cost as the list price.

Table 2.5. Supply Estimates

	No Reference Effect		With Reference Effect	
	Mean	S.D.	Mean	S.D.
Cost estimates				
Publisher's mc, c^U/l	0.398	0.132	0.427	0.116
Retailer's mc, c^D/l	−0.751	0.404	−0.294	0.073
Margins				
Publisher margin, $(w - c^U)/l$	0.125	0.080	0.095	0.062
Retailer margin, $(p - w - c^D)/l$	0.807	0.429	0.350	0.040
Bargaining parameters				
Large publishers	0.170		0.094	
Medium publishers	0.133		0.063	
Small publishers	0.137		0.065	
Average	0.147		0.074	
Number of Observations	6,228		6,228	

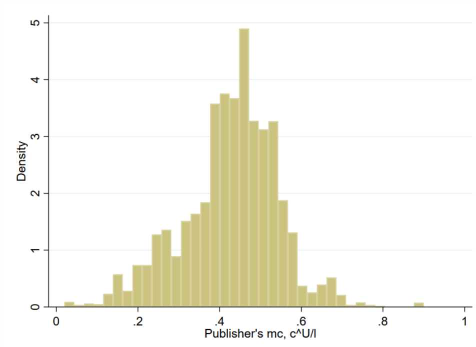
Notes: Estimates of publisher and retailer marginal costs, margins and bargaining parameters for two demand specifications (without and with reference effects).

We use the IV estimates in column (2) and (4) of Table 2.4 to estimate the supply side. Table 2.5 gives the estimates of marginal costs, margins and bargaining parameters for the two specifications without and with reference effects. The estimated publisher's marginal cost is on average around 40% of the list price for both specifications. This is consistent with industry sources that the list price is usually set such that the book's marginal cost (printing cost) should be around 40% of the list price.

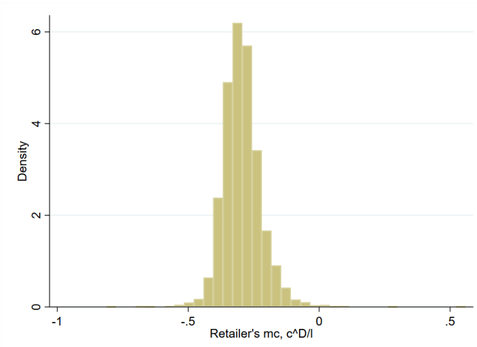
The retailer's marginal cost is estimated to be −29% of the list price using the reference effect demand estimates. The negative retailer marginal cost may seem odd. Empirically, it comes from the retailer setting lower prices than the case with zero or positive marginal cost given the relatively inelastic demand. Similar results are found in the e-book market, where Amazon frequently sets retail prices below the wholesale price (De los Santos et al.,

Figure 2.9. Marginal Costs and Margins

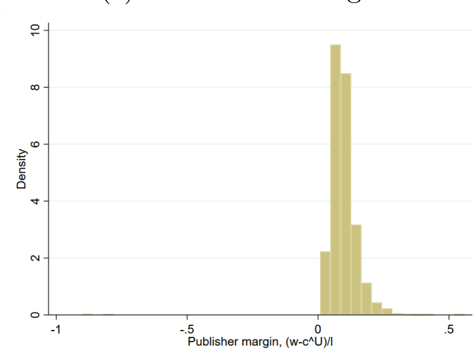
(a) Publisher's Marginal Cost



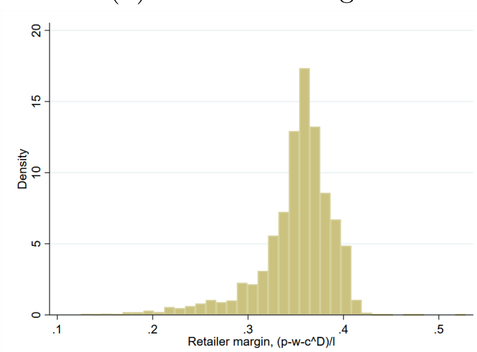
(b) Retailer's Marginal Cost



(c) Publisher's Margin



(d) Retailer's Margin



Notes: Estimates of the distribution of publisher and retailer marginal costs and margins. The distributions are plotted across 6,228 books.

2021). A number of empirical studies conclude that e-book and print prices are set below static profit maximizing levels on Amazon (Chevalier & Goolsbee, 2003, De los Santos et al., 2012, and Reimers & Waldfogel, 2017). There are several explanations for why retailers set lower prices, including loss-leader pricing and behavioral explanations. De los Santos & Wildenbeest (2017) suggest that negative retailer marginal cost represents the value to the retailer of bringing in consumers beyond selling the book, e.g., loss-leader strategy or customer acquisition and retention. By increasing consumer satisfaction, retention, and repeated business, this strategy is focused on revenue growth and cash flow instead of margins. In contrast, the estimated retailer's marginal cost is -75% of the list price using no reference effect demand estimates, which is even higher (in absolute value) than the wholesale rate (on average 52% of the list price). Thus, including reference effects in demand estimation gives more reasonable supply estimates. Given the cost estimates, we find that the publisher's profit margin is on average 9.5% of the list price and retailer's profit marginal is on average 35% of the list price. Figure 2.9 plots the marginal costs and margins distributions across books estimated from the reference effect model.

In the estimation of bargaining parameter λ, we allow λ to be different across publisher sizes. Publisher size is defined as the log of total quantity sold in the sample period for books published by the publisher. Large, medium and small publishers are defined as the top 20, top 20 to 60, and over top 60 publishers, respectively, such that large, medium and small publishers each published roughly one-third of the books in our estimation sample.[25] The average bargaining parameter across publishers is 0.074, which suggests that the retailer has most of the bargaining power. This is consistent with the fact that publishers' profit margin is much lower than the retailer's profit margin. Additionally, large publishers have higher bargaining power than medium and small publishers.

To further decompose the profit margins across publisher sizes, Table 2.6 presents the average profit margin, wholesale price and marginal cost for large, medium and small publishers. It shows that the average marginal costs of the publishers are similar. However,

[25] Allowing the bargaining parameter to vary at the publisher level is computationally infeasible due to curse of dimensionality as there are over 300 publishers. We show in the model fit that the R-squared of the linear regression of publisher' marginal cost is already over 0.9 with only three levels of bargaining parameters.

Table 2.6. Margins across Publisher Sizes

	$(w - c^U)/l$	w/l	c^U/l	Obs
Large publishers	0.110	0.542	0.432	2,114
Medium publishers	0.088	0.510	0.422	2,011
Small publishers	0.086	0.514	0.428	2,103

Notes: Estimates of average profit margin, wholesale price and marginal cost for large, medium and small publishers.

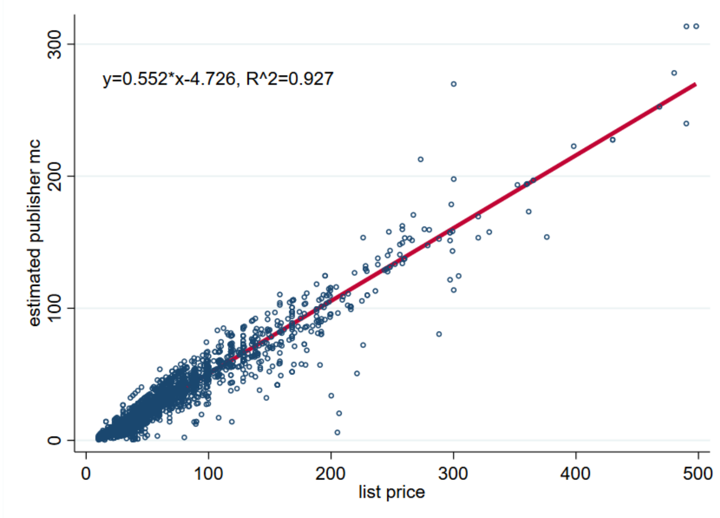
large publishers negotiate higher wholesale price and have higher profit margin, which is consistent with the fact that larger publishers have higher bargaining power.

2.4.3. Model Fit

We report the model fit first by showing that the linear regression we use to solve for bargaining parameters fits well with the data. Figure 2.10 shows the scatter plot with fitted line of the estimated publisher's marginal cost versus the list price of the books in our estimation sample. The objective function when calculating the bargaining parameters is to minimize the sum of squared residuals of the regression, which is equivalent to maximizing the R-squared of the regression. Figure 2.10 shows that the maximized R-squared is 0.927, which means that even with only three bargaining parameters, we are able to fit the data well.

We have found that the bargaining parameters are relatively small for the publishers. The average bargaining parameter is estimated to be 0.074, comparing to an average of 0.21 in the literature when the big 5 publishers bargain with Amazon in the e-book market in the US (De los Santos et al., 2021). This difference is due to the fact that publishers in China are relatively small in size comparing to the publishers in the US. Table B3 reports

Figure 2.10. Model Fit of Estimated Publisher's Marginal Cost



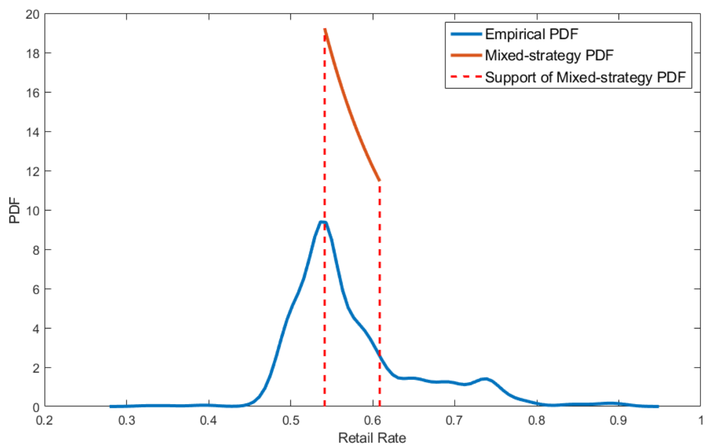
Notes: The figure shows the scatter plot of the estimated publisher's marginal cost and the list price across 6,228 books. The solid line is the fitted line.

the number of quantity sold, number of unique titles, and total revenues of the top 10 publishers in our data ranked by quantity sold as the percentage of the total quantity sold of all publishers. It shows that the top 10 publishers on average account for less than 3% of quantity and revenue, and less than 0.7% of the titles. Therefore, the publishers have much smaller market share comparing to the retailer who has 40% of market share in the downstream market.

Another finding in our estimates is that larger publishers negotiate higher wholesale price and have higher profit margin. We confirm this finding by regressing wholesale rates on publisher sizes. Table B4 reports regression results of the wholesale rate on the publisher size controlling for different book-level fixed effects. Across all specifications, the publisher size is positively correlated with the wholesale rate and the coefficients are statistically significant. The coefficients imply that the publisher can negotiate 1%

(0.00562/0.522) higher wholesale rate if it is 1 log point (2.7 times) larger in size. To put in perspective, the top 1 publisher is about 3 times the size of the top 20 publisher and 8 times the size of the top 60 publisher.

Figure 2.11. Model Fit of Price Distribution



Notes: This figure plots the density of observed prices and the density of predicted optimal prices calculated based on the demand estimates for the best seller *To Live*.

Lastly, we examine the fit of the predicted price distribution from the mix-strategy BNE in Equation (2.6). Figure 2.11 plots the density of observed prices and the density of predicted optimal prices calculated based on the demand estimates for the best seller *To Live*. Both distributions involve randomized prices. The predicted price distribution captures the fact that lower prices have more mass than higher prices. However, the observed price density has larger support than the theoretically optimal one. It suggests that there are "too many" low and high price observations compared to what is implied by the optimal pricing equilibrium. This could come from the fact that we only use the average retailer marginal cost and demand estimates to simulate the pricing equilibrium. In reality, there could be unobserved cost shocks that change retailer's marginal cost and pricing decision. Also, the observed price distribution could represent the mix of mixed-strategy distributions due to unobserved cost and demand shocks.

2.5. Counterfactual Analysis

In the counterfactual analysis, we compare equilibrium prices, quantity and firm profits in three alternative scenarios: (i) fixed consumer belief, (ii) vertical merger, and (iii) agency pricing. These counterfactuals are interesting because they allow us to empirically quantify the extent to which price reference effects and vertical contracting determine the optimal pricing and firm profits.

We use the mean parameter estimates in the demand and supply estimation to simulate the counterfactuals. Specifically, we consider a book with average list price, publisher and retailer marginal costs, bargaining parameter, price elasticity and reference effect parameters. We first simulate the baseline case where the publisher and retailer bargain over the wholesale price and allow consumer belief to respond to retail prices. The result is shown in the first row of Table 2.7. The wholesale rate is negotiated to be 50.4% of the list price and the mean retail rate is 55.9% of the list price. Publisher and retailer profits are calculated as the product of profit margin, quantity and list price. $\pi^M = \pi^U + \pi^D$ is the joint profit of the two firms.

Fixed Consumer Belief. To examine the effect of consumer belief on equilibrium outcomes, we simulate the case where consumers' belief is fixed at the level in row (1) and allow the publisher and retailer to re-optimize based on the fixed consumer belief, or equivalently the demand curve. In this way, we shut down the equilibrium channel that consumer belief affects the pricing in Figure 2.1.

Table 2.7. Counterfactual Simulations

	w/l	τ	r_e	q	π^U	π^D	π^M	CS
(1) Wholesale model, equilibrium belief	0.504	–	0.559	0.987	4.007	18.224	22.231	10.60
(2) Wholesale model, fixed (1) belief	0.455	–	0.508	1.145	1.682	21.092	22.773	13.45
(3) Vertical merger, equilibrium belief	–	–	0.496	1.025	–	–	19.653	11.38
(4) Vertical merger, fixed (1) belief	–	–	0.495	1.188	–	–	22.802	14.26
(5) Agency pricing, equilibrium belief	–	0.564	1.194	0.605	2.990	31.036	34.026	3.97
(6) Agency pricing, fixed (1) belief	–	0.197	0.748	0.525	4.826	12.277	17.103	3.05

Notes: Simulated wholesale rate, royalty rate, mean retail rate, quantity, publisher's profit, retailer's profit, joint profits and consumer surplus under different cases. Simulations are based on mean demand and supply estimates in Table 2.4 and Table 2.5.

The second row of Table 2.7 presents the results. It shows that when the demand curve is fixed, the publisher and retailer would negotiate a lower wholesale price and the retailer would lower the retail price. Quantity sold would be 16% higher as a result of the lower prices. Consumer surplus is higher due to the lower price and higher quantity. Even though quantity goes up, publisher's profit decreases as the quantity increase is not enough to compensate for the lower wholesale price received by the publisher. However, retailer's profit and joint profits would be higher due to the higher quantity sold.

These results show that in the absence of price reference effects, the retailer would have additional incentive to lower the retail price since it no longer needs to consider the negative externality on consumer expectation in response to the lower prices. With price reference effects, the retailer is unwilling to lower retail prices because lower prices would change consumer expectations. In this sense, the updated consumer belief disincentives the retailer to lower prices and results in higher prices than the case with fixed belief. The distorted retail price also leads to lower joint profits for the firms.

Vertical Merger. We have illustrated in Section 2.2.5 that whether vertical merger results in higher or lower joint profits depends critically on the magnitude of the price reference effect. When there is no price reference effect, the post-merger joint profits must be higher than the pre-merger joint profits. This is confirmed by comparing row (4) to row (2) in Table 2.7, where the consumer belief, and equivalently the demand curve is fixed. The vertical merger leads to lower prices and higher quantity as predicted by the canonical model of eliminating double marginalization.

However, when there exist price reference effects, it is possible that the post-merger joint profits are lower than the pre-merger joint profits as illustrated in Figure 2.4 and Figure 2.5. Row (3) in Table 2.7 presents the results when consumer belief is allowed to update based on prices. Comparing with row (1), retail price is lower after vertical merger, but the increase in quantity is much smaller than the case of fixed demand in row (4). As a consequence, the joint profits are 11% ($= 1 - 19.653/22.231$) lower than the non-integrated case in row (1). As being predicted in Figure 2.4, consumers update their belief and expect lower prices after the merger. Then demand curve shifts downward since the reference prices are lower. In the new equilibrium, the quantity increase is relatively small such that the post-merger joint profits are lower than the pre-merger joint profits. This result highlights the importance of the interaction of consumer expectation and vertical merger in determining retail prices and firm profits. Nevertheless, the change in consumer surplus is similar to the case with fixed consumer belief.

Agency Pricing. Lastly, we consider an alternative vertical contract between the publisher and the retailer. Under agency pricing, the publisher pays the retailer sales royalties to sell books at prices determined by the publisher. Agency arrangements are prevalent in online markets. For example, third-party sellers on Amazon set the retail prices for their products, while Amazon receives a percentage of the revenue. Other examples include Apple App Store and Taobao Marketplace in China.

In the agency model, the publisher and the retailer bargain over the royalty rate τ. Retailer's profit from selling book j is

$$\pi_j^D = (\tau_j p_j(\tau) - c_j^D) \cdot s_j(p(\tau))$$

where τ_j is the royalty rate given to the retailer. The publisher's profit is

$$\pi_j^U = ((1 - \tau_j) p_j(\tau) - c_j^U) \cdot s_j(p(\tau)).$$

The Nash product is

$$\Omega_j(\tau) = (\pi_j^U)^\lambda (\pi_j^D)^{1-\lambda}.$$

De los Santos et al. (2021) show that in the absence of price reference effects, whether prices will be higher or lower under agency model than under wholesale model depends on the relative bargaining power of the firms. Similar to their findings, comparing row (6) to row (2) in Table 2.7 shows that when the demand curve is fixed, the retail price is higher under agency contract than under wholesale contract. However, joint profits are lower under agency contract due to the decrease in quantity.

To examine the case with price reference effects, we simulate the equilibrium prices and profits under the agency model. Row (5) in Table 2.7 presents the results with updated equilibrium belief. It shows that the retailer would charge a 56.4% royalty. Prices would be on average 119.4% of the list price and quantity is much smaller comparing to the wholesale model in row (1).[26] Even with the much higher prices, the publisher has lower profit due to the high royalty and low quantity. However, the retailer would receive much higher profit than that under the wholesale model. The joint profits move in the same direction as the price as the increase in prices overshadows the decrease in quantity.

2.6. Conclusion

In this paper, we study the impact of price reference effects on optimal pricing and firm profits under different vertical contracts. We develop a structural model where demand features reference-dependent utility in price, and the upstream and downstream firms bargain over wholesale prices. We examine the Bayesian Nash Equilibrium pricing solution of the downstream firm, and the Nash bargaining between upstream and downstream firms. We show theoretically that whether the joint profits are higher or lower after vertical merger depends critically on the magnitude of price reference effects.

Using a novel dataset from a large online book retailer, we estimate the structural model and find significant bargain-hunting behavior of consumers. The estimates suggest that the retailer has most of the bargaining power when negotiating with the publishers, and larger publishers have higher bargaining power and negotiate higher wholesale price than smaller publishers.

Counterfactual simulations show three results: (i) prices would be lower if consumers' expectation is fixed; (ii) joint profits could be lower after vertical merger if consumers' expectation is updated; and (iii) prices would be higher if the vertical contract switches from

[26]The simulation puts no restriction on the prices that the publisher could set. However, it is common practice that the retail price should not exceed the list price in the book retail market. Thus, we also simulate the results under the agency model while prohibiting the retail price to be higher than the list price. We find that under this retail price restriction, the retail price is still higher than that under the wholesale model.

wholesale contract to agency contract. These findings highlight the effects of consumer expectation on optimal pricing and firm profits.

Our paper is one of the first structural analysis of the price reference effect in industrial organization. We believe that the methods developed in this paper can be applied to other settings in which reference-dependent preference is a potential concern. For future research, we think that it is important to study other types of markets in order to understand the impact on consumer welfare and possible policy interventions.

CHAPTER 3

Demand Uncertainty and Information Provision: Evidence from Chinese Online Book Market

3.1. Introduction

As e-commerce industry has been taking off, a lot of changes have been brought to consumer's shopping and purchasing process. Quarterly U.S. retail E-commerce sales as a percent of total retail sales has been growing from 5% in 2012 to 14.7% by the end of 2022 (*U.S. Census Bureau, E-Commerce Retail Sales as a Percent of Total Sales [ECOMPCTSA]*, 2003). Compared to using traditional methods such as trying on clothes in person before purchasing them, consumers now rely much more on the detailed product information such as fabric type, texture, and color range, for more informed shopping decisions when they shop online. This raises an important question to economics researchers: how does information provision affect demand.

Improved consumer information about horizontal aspects of products of similar quality leads to better consumer matching but also to higher prices, so consumer surplus can go up or down (S. P. Anderson & Renault, 2009). However, there has been little direct evidence on the effect of product information provided by e-commerce platforms on consumer surplus in the field of economic research. This is mainly because of data constraint. In marketing science, researchers have studied consumer learning and product experience experimentally (Hoch & Ha, 1986 and van Osselaer & Alba, 2000), but it is difficult to relate their qualitative results to consumer surplus. In this paper, I use a novel daily sales data set from a dominant book retailer in China to study the effect of product information provision on demand, in the context of online retail market. In the data set, in addition to daily prices and sales that can be quantified to compute consumer surplus, I also observe a direct indicator of inefficient purchasing decisions, which is the return of books.

The online book retail market is an ideal setting to study the effect of information provision. Different from a local book store, consumers rely on the web-page information on book attributes and contents to decide whether a book is worth buying. The economic rationale for return policies of these "experience goods" is that customers do not fully know their preferences for the products until after they gain some experience with them. Therefore, with the option to return, a consumer who has learned that she does not like a product can nullify her purchase by simply returning it (Che, 1996). In this paper, the online book retailer experienced an exogenous shock on the amount of information provided to customers. This event will serve as a direct approach to study the quantitative effect on demand and prices.

I begin by presenting a conceptual framework of a change in consumer's prior beliefs of product quality that predicts: (i) The change of ex-ante bias in prior beliefs of product quality can be a mixture of more than one scenario. Three most commonly expected scenarios lead to ambiguous predictions on total sales. (ii) Regardless of how the observed outcomes are mixed from three scenarios, the probability of returning products is decreased when consumers have more accurate prior beliefs about product quality. (iii) Consumers with more accurate prior beliefs also make better matches with products by missing fewer products that give them positive utilities.

I use online book sales data from DangDang.com, a leading online book retailer in China for decades. During my sample period of 2017 to 2019, DangDang.com introduced exogenous shocks to the amount of book content preview, while keeping other product information unchanged. These changes come from the introduction of e-books. Whenever the e-version of a physical book is available on DangDang.com, the amount of content preview increases from one page to 2% - 8% of the book. I use event study analysis, which is a widely used approach in empirical industrial organization (Schwert, 1981 and Rose, 1985), to identify the effect on consumer demand and prices caused by increased amount of product information, where I define an event as the release of the e-version on DangDang.com. In the selected sample in which books are more subject to readers' tastes, results suggests that the increased amount of information provision reduces the return rate by 18%-19% on average. In the full sample, although results show a robust negative point estimation without statistical significance.

Related Literature. This paper is related to three strands of literature. First, it relates to the literature on demand uncertainty and return policies. Researchers have studied demand uncertainty by various outcome variables that can be observed in data. Moretti (2011) looks at the learning process of movie quality using sales data. M. Anderson & Magruder (2012) and Oestreicher-Singer & Sundararajan (2012) study consumer's learning by review and recommendation data. How return policies in e-commerce market affect consumers' uncertainly about products is seen in literature. However, these studies have been mainly focusing on theories in design, including Marvel & Peck (1995), Che (1996), and Johnson & Myatt (2006). Some studies use survey data (Pei et al., 2014). The majority of the reason is the lack of return data. This paper is one of the first analyses that use directly observed return data to empirically estimate and find strong evidence on the effect of information on demand uncertainty.

Second, this paper focuses on a new information supply channel that has not been studied by consumer learning literature. Information provided by online platforms is hard to be quantified and converted into analysable data. Previous literature has been focusing on channels that are common in offline markets (S. P. Anderson & Renault (2009) looks at comparative advertising; Cheng & Liu (2012) looks at free trials; and Bell et al. (2018) looks at offline showrooms). In the context of online markets, the widely used variables are review and rating data (M. Anderson & Magruder, 2012 and Oestreicher-Singer & Sundararajan, 2012). In this paper, I examine a new channel from which consumers can receive information on good quality, which is the product attributes supplied directly by websites.

Third, this paper generally relates to the literature of platform's market power. The rise of the platform economy has brought researchers' attention to a lot of ways in which the platform can exercise market power. Lee & Musolff (2021) evaluates the problem of firms that operate platforms matching buyers and sellers, and influence market outcomes by platform-guided search. Relatedly, Barach et al. (2020) and Farronato et al. (2023) investigate the self-steering behaviors by by search recommendation and ranking. Bamberger & Lobel (2017) concerns about platforms' market power from their multi-sided network. This paper studies how platforms can potentially grow market power by the tool of cost-less information, since improved information comes with higher prices (S. P. Anderson & Renault, 2009).

This paper proceeds as follows. Section 3.2 presents a simplified model and graphical illustration in the context of online book retail market with demand uncertainty and returns to motivate the empirical work. Section 3.3 describes the background and data. Section 3.4 presents the empirical implementation. Section 3.5 discusses more work that could be done on this topic and conclude.

3.2. Graphical Illustration

In this section, I present a simplified model with graphical illustration that shows: (i) The change of ex-ante bias in prior beliefs of product quality can be a mixture of more than one scenario. Three most commonly expected scenarios lead to ambiguous predictions on total sales. (ii) Regardless of how the observed outcomes are mixed from three scenarios, the probability of returning products is decreased when consumers have more accurate prior beliefs about product quality. (iii) Consumers with more accurate prior beliefs also make better matches with products by missing fewer products that give them positive utilities.

3.2.1. A Simplified Model with Demand Uncertainty

For simplification and graphical illustration, I assume that there is a mass of one consumers indexed by i and one book in the market. For each representative individual consumer i, the utility from reading the book is given by

$$U_i = \alpha_i^0 + \alpha_i^1$$

where α_i^0 is the true quality of the book and α_i^1 is the ex-ante bias in prior beliefs held by individual consumers. I further assume that the book has a price p, and that individual consumers have dis-utility from purchasing the book at price βp. I define $B_i = -\beta p$, so that the consumer i will purchase the book when

$$\alpha_i^0 + \alpha_i^1 > B_i$$

I assume that after purchasing the book, consumer i will fully reveal the true quality of the book α_i^0. In addition, she can choose to return the book for free. She will choose to return the book when

$$\alpha_i^0 < B_i.$$

Using these inequalities, I can define three outcome variables that are affected by the change of information provision and examine in which direction these variables change.

3.2.2. The Effects of the Changes in Information Provision

I first define three outcome variables that I can quantify from the model as the following.

I define "purchasing" to be the probability of buying the book after learning the price p. In my data set, it is observed by the sales data. I define "return" to the probability of returning the book, after revealing the true quality of the book α_i^0. It is also observed in data set by the number of returned copies. Last but not least, I define "inefficient no purchase" to be the probability of consumer's missing the book that gives positive utility, because of being too pessimistic about the quality. This outcome is not observed in data set, but still important in quantifying the consumer surplus.

$$Purchase\,and\,no\,return = \Pr[\alpha_i^0 + \alpha_i^1 > B_i, \alpha_i^0 > B_i],$$

$$Purchase\,and\,return = \Pr[\alpha_i^0 + \alpha_i^1 > B_{ij}, \alpha_i^0 < B_i],$$

$$Inefficient\,no\,purchase = \Pr[\alpha_i^0 + \alpha_i^1 < B_i, \alpha_i^0 > B_i].$$

For visual illustration, I normalize $B_i = 0$ and assume normal distributions $\alpha_i^0 \sim U(a_0, a_1)$, $\alpha_i^1 \sim U(b_0, b_1)$.

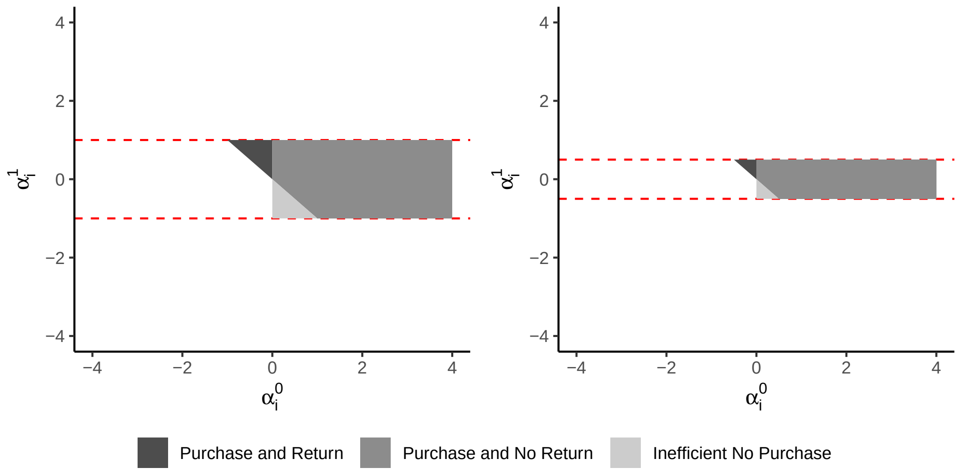
Figure 3.1 (a) shows the scenario in which there is a reduced ex-ante bias in prior beliefs. In the figure, it is depicted by the narrowed support of α_i^1 from the left figure to the right. In this case, the probability of returning the book, which is the darkest triangle above the forty-five degree line, is decreasing. This is because when consumers have more precise prior beliefs around the true quality, there is a smaller probability to be over pessimistic. In addition, there is also a reduction of the probability of "inefficient no purchase" (indicated by the lightest triangle), which comes from fewer too pessimistic consumers. The probability of buying the book is the area above the forty-five degree line. When there is no change in the bias mean, this probability does not change.

Figure 3.1 (b) and (c) show the other two scenarios, where there is no change in bias variance, but there is an upward (downward) shift in bias mean. In this case, over pessimistic (optimistic) consumers adjust their prior beliefs about the book quality by being given more information, and converge to efficient purchasing decisions. Too optimistic consumers make fewer returns, as they buy fewer books whose true quality is below the utility threshold. In the opposite direction, too pessimistic consumers make less “inefficient no purchases”.

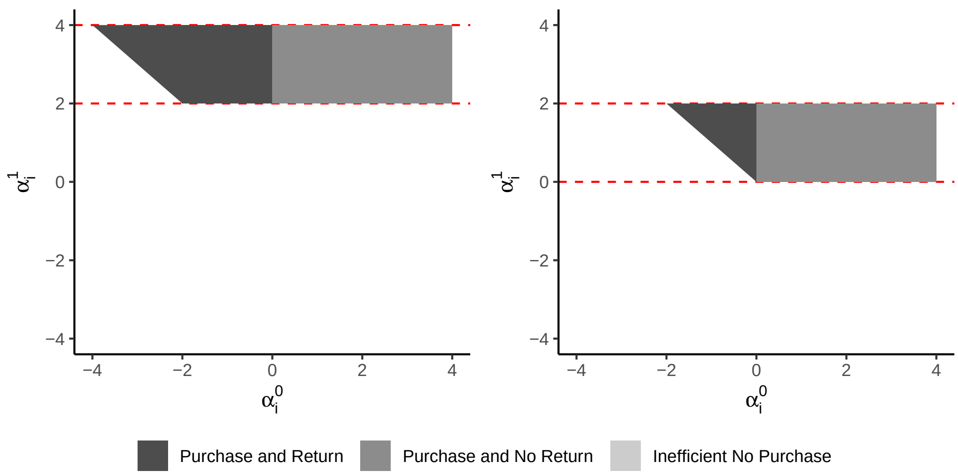
To summarize, when the real-world problem has a mixture of these three cases, the probability of returning the book is predicted to be decreasing, given more information about product quality. There is ambiguity in the probability of buying. “inefficient no purchase” is predicted to be decreasing as well, but it is not directly observed in DangDang.com data set.

Figure 3.1. Changes in Prior Beliefs on Sales and Returns

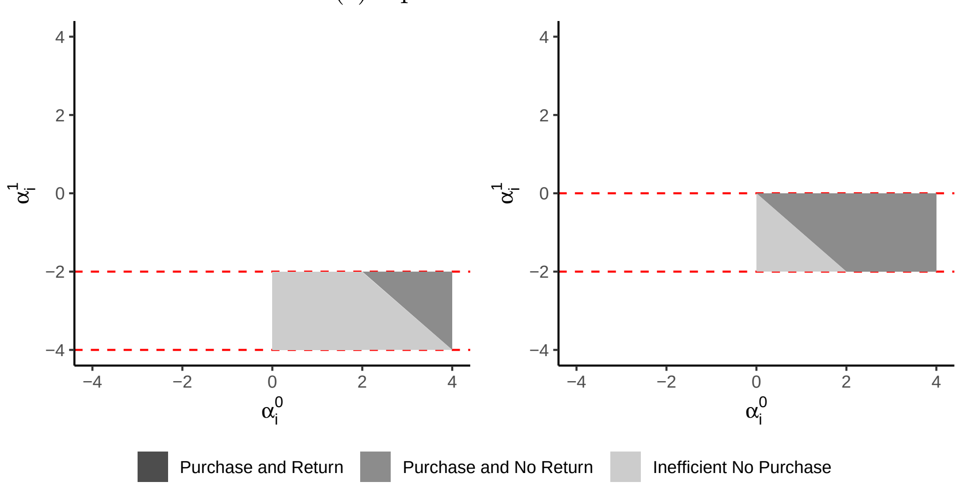
(a) Reduced Variance



(b) Downward Shift Mean



(c) Upward Shift Mean



3.3. Background and Data

This section introduces the institutional background of the online book retail market with a focus on the information provision shocks. Two screenshots are presented in Section 3.3.1 to distinguish different levels of product information supply before and after the shock. Section 3.3.2 summarizes the data set.

3.3.1. Institutional Background

The general institutional background of global and Chinese online book retail markets, as well as the online book retailer DangDang.com, has been introduced in Section 2.3.1. In this section, I discuss the chapter specific context in which DangDang.com makes exogenous decisions to change the amount of book information on the web-page. I use this change as a shock to estimate the effect on demand and prices.

Figure 3.2 shows a web-page screenshot of the best seller *To Live* on DangDang.com. This book has about 400 pages. In red color, there is a button called "Online Preview". After clicking this button, at no cost, readers are able to preview the book content by one page.

Figure 3.3 shows the web-page screenshot of another best seller *Brothers* from the same author and publisher on DangDang.com. This book has a similar number of pages. Different from *To Live*'s web-page, there are two more buttons for the e-version. At the bottom, the blue button called "Buy E-Book" allows readers to access the web-page of the e-version on DangDang.com. The e-version web-page will then provide similar information as the physical version web-page and let readers decide whether to buy the e-version. Readers can also view the price of the e-version on the physical version web-page, next to the physical version price. If and only if the e-version of a book is available to be bought from DangDang.com, these changes will take place. The most important change is the "Online Preview" button. Once the e-version is available, the number of pages accessible for content preview will be increased.[1]

Figure 3.2. Webpage Example without E-Version and Increased Preview

Notes: Webpage example of a book (the best seller *To Live*) sold on DangDang.com. The screenshot was taken on May 29, 2021 from http://product.dangdang.com/25137790.html.

[1]Increases are between 2% to 8%, according to DangDang.com. The reason is that there is very little cost associated with adding content preview to web-page when the e-version is online.

Figure 3.3. Webpage Example with E-Version and Increased Preview

Notes: Webpage example of a book (the best seller *Brothers*) sold on DangDang.com. The screenshot was taken on May 29, 2021 from http://product.dangdang.com/25147357.html.

I use the increases of the amount of content preview as the shocks to estimate the effect of information provision on demand and prices. There are limitations in this approach. First, although DangDang.com claims to put e-books on their website once they are published, the decision of publication can be correlated with consumer's prior beliefs and that brings biases in the estimation. For example, when a physical book gets more popular over time, publishers may decide to publish the e-version, which introduces more content preview to DangDang.com. At the same time, consumers may receive information about the popularity and update their prior beliefs on the book quality. Such examples will contaminate the results, potentially in both directions. I will discuss how to solve them in the following sections.

3.3.2. Data

I use administrative sales data from DangDang.com during 2017 to 2019, which has been introduced in Section 2.3.2. In addition to the baseline sample, I make the following changes to the data set to exclude outliers and eliminate biases.

I aggregate sales data to the weekly level, instead of the daily level. By averaging prices and quantities across time, I am able to avoid abnormal time periods (such as holidays and big sales) and draw overall inferences for the event of interest.

I keep the six biggest categories on DangDang.com to exclude abnormal books that are in minor categories with different demand curves. These six categories are novel, management, inspirational&investment, non-fic, history, and computer. Books from these categories account for 80% of the market shares on DangDang.com throughout the sample period. I keep books with positive sales records in more than half of the weeks and also and in more than 3 weeks after the publication date to avoid sparse data bias. I only keep books with both physical and e- versions available on DangDang.com, so that the sampled books are comparable. These selections leave 8,565 titles.

Table 3.1 summarizes the full sample in the pre-event period. Each week there are 81 copies are sold per title, at the price of 36 CNY ($\approx$ 6 USD) on the average. There is one copy being returned per week title. From the distribution of returns, the majority of returns are concentrated in a small group of books, presumably popular book with large sales quantities. This observation motivates the weighting strategy in the following section.

Table 3.1. Summary Statistics of Full Sample

	mean	p10	p25	p50	p75	p90	max	sd
Book Week Level:								
quantity	81.350	2	4	12	42	140	12,241	345.303
return	.731	0	0	0	0	1	4,572	21.469
price (in CNY)	36.001	18.448	22.586	29.977	41.3	57.413	616.2	26.120
Observations				56,218				
Book Level:								
week_since_publish	34.423	10.5	20	31.542	45	64.5	94.5	19.869
Observations				2,763				

Notes: This table presents summary statistics of the estimation sample of books sold on Dang-Dang.com from Jan 1, 2017 to Sept 2, 2019. Prices are in the unit of CNY. 1 USD $\approx$ 6.6 CNY in 2018.

To have heterogeneity in results, I continue to select three categories (novel, history, non-fic) out of the six and call them the selected sample in the pre-event period. I argue that these categories are more subject to readers' tastes. Table 3.2 summarizes the selected sample. Compared to the full sample, the selected sample has more sales per week and a higher return rate on the average.

3.4. Empirical Methhod and Results

3.4.1. Empirical Approaches

The approach that I use to identify the changes on demand and prices caused by increased amount of content preview is the event study analysis. This is a widely used method in industrial organization when there are exogenous shocks happening over time (Schwert, 1981 and Rose, 1985). Using the full (selected) sample of books, I define an event as the

Table 3.2. Summary Statistics of Selected Sample

	mean	p10	p25	p50	p75	p90	max	sd
Book Week Level:								
quantity	100.473	2	5	16	56	180	12,161	400.067
return	.975	0	0	0	0	1	4,572	27.656
price (in CNY)	32.781	17.542	21.05	26.657	34.486	48.95	616.2	29.390
Observations				30,333				
Book Level:								
week_since_publish	33.121	10.571	19.5	30	43.5	60.188	94.5	18.815
Observations				1,445				

Notes: This table presents summary statistics of the selected estimation sample (novel, history, non-fic) of books sold on DangDang.com from Jan 1, 2017 to Sept 2, 2019. Prices are in the unit of CNY. 1 USD $\approx$ 6.6 CNY in 2018.

release of the e-version on DangDang.com, given that the physical version already exists on the website.

The baseline two-way fixed-effect specification is

$$y_{jt} = \alpha + \beta I\{postevent\}_{jt} + \xi_j + \delta_t + \phi_{r_{jt}} + \epsilon_{jt}$$

where y_{jt} is an outcome variable observed in the data set for book j in week t. β is the key parameter of interest that indicates the effect on the outcome variable. $I\{postevent\}_{jt}$ is the indicator of whether the observation has experienced the event. $I\{postevent\}_{jt} = 1$ when the e-version is available for book j in week t, and zero otherwise. ξ_j and δ_t are the standard book and week fixed effects. r_{jt} is the number of weeks, and $\phi_{r_{jt}}$ is the number of weeks since the publication of the physical version fixed effect for book j in week t. It helps to control for the shared product demand cycle of books. I also extend the baseline analysis to including other fixed effects and time trends as the robustness checking, which will be discussed next.

3.4.2. Empirical Estimates

This section presents the estimation results of the information provision effect on demand. As predicted by the model in Section 3.2, the probability of return the book is expected to drop when more content preview is available on the web-page. In this section, I focus on the return rate, which is defined by the number of copies returned over the number of copies bought per title per week. Table 3.3 shows the estimation results of the effect on the return rate using the full sample.

Table 3.3. Results of Full Sample: Returns

VARIABLES	(1) return	(2) return	(3) return	(4) return	(5) return	(6) return
post_event	-0.0771*	-0.0632	-0.0709	-0.0745	-0.0140	-0.0361
	(0.0452)	(0.0518)	(0.0457)	(0.0480)	(0.169)	(0.157)
pre_event_mean	0.899	0.899	0.899	0.899	0.899	0.899
Observations	147,747	147,747	147,747	146,399	147,747	147,747
R-squared	0.041	0.059	0.042	0.085	0.059	0.078
Book Time Trend	N	Y	N	N	N	Y
Cate Time Trend	N	N	Y	Y	N	N
Cate Week FE	N	N	N	Y	N	N
Pre Sale Weight	N	N	N	N	Y	Y

Notes: This table reports the estimation results from the event study analysis. The unit of observation is a book-week. The dependent variable is return rate (from 0 to 100). All specifications include book and week level fixed effects, as well as week since publication fixed effects. Standard errors are clustered at the book level and presented in parentheses below the coefficients. *p<0.1, **p<0.05, ***p<0.01.

In Table 3.3, column (1) presents the baseline specification results. There is a 7% reduction in return rate from the increased amount of content preview with weak statistical significance. However, considering each title may have specific demand cycle in a calendar year,[2] I add book specific time trend to the baseline specification and show the results in column (2). Column (3) and (4) control for category fixed effects and category specific time trend as analog to the first two columns. Since returns are observed for a small group of books, in column (5) and (6), I weight observations by the average sales in the pre-event period, so that books that are more likely to have returns will drive the results. Based on the full sample estimation results, I cannot reject that the information provision has no effect on the return rate.

Table 3.4. Results of Selected Sample: Returns

VARIABLES	(1) return	(2) return	(3) return	(4) return	(5) return	(6) return
post_event	-0.191***	-0.182**	-0.193***	-0.184***	-0.157	-0.166
	(0.0655)	(0.0736)	(0.0667)	(0.0709)	(0.183)	(0.189)
pre_event_mean	0.970	0.970	0.970	0.970	0.970	0.970
Observations	75,026	75,026	75,026	74,375	75,026	75,026
R-squared	0.048	0.068	0.048	0.094	0.074	0.098
Book Time Trend	N	Y	N	N	N	Y
Cate Time Trend	N	N	Y	Y	N	N
Cate Week FE	N	N	N	Y	N	N
Pre Sale Weight	N	N	N	N	Y	Y

Notes: This table reports the estimation results from the event study analysis. The unit of observation is a book-week. The dependent variable is return rate (from 0 to 100). All specifications include book and week level fixed effects, as well as week since publication fixed effects. Standard errors are clustered at the book level and presented in parentheses below the coefficients. *p<0.1, **p<0.05, ***p<0.01.

[2]For example, fictions may have very different demand cycles from computer textbooks. The former may be more popular during holidays, while the latter may be more popular during semesters.

Table 3.4 presents the estimation results using the selected sample. Results are negative and robust when the observations are not weighted by their sales in the pre-event period. This suggests that the increased amount of information provision reduces the return rate by 18%-19%, and it is statistically significant. However, when the observations are weighted by their sales in the pre-event period, the variance increases and that the information provision has no effect on the return rate cannot be rejected. Since books with larger sales quantities in the pre-event period have more weights in these results, it suggests that the significant average effect on the return rate in the first four columns is mainly driven by less popular books, whose sales depend more on the web-page information.

I also look at the effects of increased amount of content preview on prices and quantities in Table C1 to C4. As predicted by the model in Section 3.2, these effects are ambiguous. Indeed, based on the results from the event study, I do not observe robust statistically significant results on sales and prices. The empirical approach does not help distinguish price effects and information effects that can come in opposite directions and affect sales. The model on the demand side does not support the analysis on prices. I will discuss more work that can be done on this topic in Section 3.5.

3.5. Conclusion and Discussion

This paper studies the effect of information provision on demand in a setting of Chinese online book retail market. It uses a novel data set with detailed daily sales and prices, and more importantly, daily returns that have been rarely used in literature. A demand model with uncertainty on product quality predicts that the probability of returning products is decreased when consumers have more accurate prior beliefs about product quality. To provide empirical evidence, this paper uses the event study analysis to identify the changes in return rate and sales from an exogenous change in the level of information provision. Estimation results suggest that the increased amount of information reduces the return rate by 18%-19% and it is mainly driven by less popular books.

My paper is one of the first empirical analyses that studies the effect of information supplied directly by the online retailer. I believe the richness of this novel data set will be useful on other research questions that are related to e-commerce and consumer demand.

For future research, it is important to amplify the analysis on heterogeneous effects from empirical aspect, and develop a structural model from theoretical aspect. First, as suggested by Table 3.4 column (5) and (6), the significant average effect on the return rate in the first four columns is mainly driven by less popular books. However, the pre-event sales data is not the most ideal indicator of popularity, since sales are correlated with returns. For future research, it will be helpful to examine the heterogeneous effects based on reviews and ratings data. It will also be helpful to look at heterogeneous effects over time, as it suggests the learning process of consumer's.

Second, developing a structural model that incorporates a random prior bias and the option to return will give more insights on the research question. Although the simplified model predicts the direction in which the return rate changes, it is important to characterize the bias and quantify the changes in order to draw any conclusion on consumer surplus. Moreover, with current estimation, I cannot distinguish the information effect from the price effect on changes of sales. With a structural model, this will be feasible and the paper will be able to answer a question: compared to prices, how important the product information is in the online marketplace.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, *101* (2), 470–92.

Abhishek, V., Jerath, K., & Zhang, Z. J. (2016). Agency selling or reselling? channel structures in electronic retailing. *Management Science*, *62* (8), 2259–2280.

Anderson, M., & Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, *122* (563), 957–989.

Anderson, S. P., & Bedre-Defolie, O. (2019). Variety provision of a multiproduct monopolist. *Working Paper*.

Anderson, S. P., & Palma, A. D. (1992). The logit as a model of product differentiation. *Oxford Economic Papers*, *44*, 51–67.

Anderson, S. P., Palma, A. D., & Thisse, J.-F. (1987). The CES is a discrete choice model? *Economics Letters*, *24* (2), 139–140.

Anderson, S. P., & Renault, R. (2009). Comparative advertising: Disclosing horizontal match information. *The RAND Journal of Economics*, *40* (3), 558–581.

Arkolakis, C., Demidova, S., Klenow, P. J., & Rodríguez-Clare, A. (2008). Endogenous variety and the gains from trade. *American Economic Review: Papers & Proceedings*, *98* (2), 444–450.

Bamberger, K. A., & Lobel, O. (2017). Platform market power. *Berkeley Tech. LJ*, *32*, 1051.

Barach, M. A., Golden, J. M., & Horton, J. J. (2020). Steering in online markets: the role of platform incentives and credibility. *Management Science*, *66* (9), 4047–4070.

Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, *115* (4), 588–638.

Bell, D. R., Gallino, S., & Moreno, A. (2018). Offline showrooms in omnichannel retail: Demand and operational benefits. *Management Science*, *64* (4), 1629–1651.

Bernheim, B. D., DellaVigna, S., & Laibson, D. (2018). *Handbook of behavioral economics-foundations and applications*. Elsevier.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, *25*(2), 242–262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Berry, S., & Waldfogel, J. (1999). Free entry and social inefficiency in radio broadcasting. *The RAND Journal of Economics*, *30*(3), 397–420.

Broda, C., & Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly Journal of Economics*, *121*(2), 541–585.

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, *36*(1), 13–70.

Cardell, S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, *13*(2), 185–213.

Chan, M., Kroft, K., & Mourifie, I. (2019). An empirical framework for matching with imperfect competition. *Working Paper*.

Che, Y.-K. (1996). Customer return policies for experience goods. *The Journal of Industrial Economics*, *44*(1), 17–24.

Cheng, H. K., & Liu, Y. (2012). Optimal software free trial strategy: The impact of network externalities and consumer uncertainty. *Information Systems Research*, *23*(2), 488–504.

Chevalier, J., & Goolsbee, A. (2003). Measuring prices and price competition online: Amazon and barnes and noble. *Quantitative Marketing and Economics*, *1*, 203–222.

Chevalier, J., & Kashyap, A. K. (2019). Best prices: Price discrimination and consumer substitution. *American Economic Journal: Economic Policy*, *11*(1), 126–59.

Crawford, G. S., Lee, R. S., Whinston, M. D., & Yurukoglu, A. (2018). The welfare effects of vertical integration in multichannel television markets. *Econometrica*, *86*(3), 891–954.

Crawford, G. S., & Shum, M. (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica*, *73*(4), 1137–1173.

Crawford, G. S., & Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television markets. *American Economic Review*, *102*(2), 643–85.

Crawford, V., & Meng, J. (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review*, *101*(5), 1912–32.

Cuesta, J. I., Noton, C., & Vatter, B. (2019). Vertical integration between hospitals and insurers. *Working Paper*.

De los Santos, B., Hortaçsu, A., & Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, *102*(6), 2955–80.

De los Santos, B., O'Brien, D. P., & Wildenbeest, M. R. (2021). Agency pricing and bargaining: Evidence from the e-book market. *Working Paper*.

De los Santos, B., & Wildenbeest, M. R. (2017). E-book pricing and vertical restraints. *Quantitative Marketing and Economics*, *15*(2), 85–122.

Dhingra, S., & Morrow, J. (2019). Monopolistic competition and optimum product diversity under firm heterogeneity. *Journal of Political Economy*, *127*(1), 196–232.

Diebel, A. S. (2018). Vertical integration in the us health care market: An empirical analysis of hospital-insurer consolidation. *Working Paper*.

Dinerstein, M., & Smith, T. (2014). Quantifying the supply response of private schools to public policies. *SIEPR Working Paper*.

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, *67*(3), 297–308.

Ellison, G. (2006). Bounded rationality in industrial organization. *Econometric Society Monographs*, *42*, 142.

Ericson, K., & Fuster, A. (2011). Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *Quarterly Journal of Economics*, *126*(4), 1879–1907.

Farronato, C., Fradkin, A., & MacKay, A. (2023). *Self-preferencing at amazon: evidence from search rankings* (Tech. Rep.). National Bureau of Economic Research.

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *American Economic Review*, *84*(1), 157–177.

Foros, Ø., Kind, H. J., & Shaffer, G. (2017). Apple's agency model and the role of most-favored-nation clauses. *The RAND Journal of Economics*, *48*(3), 673–703.

Gabaix, X., Laibson, D., Li, D., Li, H., Resnick, S., & de Vries, C. G. (2016). The impact of competition on prices with numerous firms. *Journal of Economic Theory*, *165*, 1–24.

Gentry, M., & Pesendorfer, M. (2021a). Deal-expectation effects in consumer demand. *Working Paper*.

Gentry, M., & Pesendorfer, M. (2021b). Pricing with bargain hunting consumers. *Games and Economic Behavior*, *129*, 549–569.

Gilbert, R. J. (2015). E-books: A tale of digital disruption. *Journal of Economic Perspectives*, *29*(3), 165–84.

Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, *102*(1), 469–503.

Gowrisankaran, G., Nevo, A., & Town, R. (2015). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review*, *105*(1), 172–203.

Heidhues, P., & Kőszegi, B. (2008). Competition and price variation when consumers are loss averse. *American Economic Review*, *98*(4), 1245–68.

Heidhues, P., & Kőszegi, B. (2014). Regular prices and sales. *Theoretical Economics*, *9*(1), 217–251.

Hendel, I., & Nevo, A. (2006a). Measuring the implications of sales and consumer inventory behavior. *Econometrica*, *74*(6), 1637–1673.

Hendel, I., & Nevo, A. (2006b). Sales and consumer inventory. *The RAND Journal of Economics*, *37*(3), 543–561.

Hendel, I., & Nevo, A. (2013). Intertemporal price discrimination in storable goods markets. *American Economic Review*, *103*(7), 2722–51.

Hicks, J. (1936). *Value and capital*. Oxford University Press.

Ho, K., & Lee, R. S. (2019). Equilibrium provider networks: Bargaining and exclusion in health care markets. *American Economic Review*, *109*(2), 473–522.

Hoch, S. J., & Ha, Y.-W. (1986). Consumer learning: Advertising and the ambiguity of product experience. *Journal of Consumer Research*, *13*(2), 221–233.

Horn, H., & Wolinsky, A. (1988). Bilateral monopolies and incentives for merger. *The RAND Journal of Economics*, *19*(3), 408–419.

Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2013). The allocation of talent and us economic growth. *National Bureau of Economic Research*.

Johnson, J. P. (2017). The agency model and mfn clauses. *Review of Economic Studies*, *84*(3), 1151–1185.

Johnson, J. P. (2020). The agency and wholesale models in electronic content markets. *International Journal of Industrial Organization*, *69*, 102581.

Johnson, J. P., & Myatt, D. P. (2006). On the simple economics of advertising, marketing, and product design. *The American Economic Review*, *96*(3), 756–784.

Kahnerman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under uncertainty. *Econometrica*, *47*(2), 263–291.

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, *121*(4), 1133–1165.

Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, *97*(4), 1047–1073.

Kőszegi, B., & Rabin, M. (2009). Reference-dependent consumption plans. *American*

Economic Review, *99*(3), 909–36.

Kreinovich, V., Nguyen, H. T., Sriboonchitta, S., & Kosheleva, O. (2015). Modeling extremal events is not easy: Why the extreme value theorem cannot be as general as the central limit theorem. *Departmental Technical Reports*.

Kroft, K., Leal-Vizcaíno, R., Laliberté, J.-W. P., & Notowidigdo, M. J. (2021). A welfare analysis of tax structures with love-of-variety preferences. *Working paper*.

Lamadon, T., Mogstad, M., & Setzler, B. (2020). Imperfect competition, compensating differentials and rent sharing in the u.s. *Working Paper*.

Lee, K. H., & Musolff, L. (2021). Entry into two-sided markets shaped by platform-guided search. *Job Market Paper, Princeton University*.

Li, H. (2021). Are e-books a different channel? multichannel management of digital products. *Quantitative Marketing and Economics*, *19*, 1–47.

MacKay, A., & Miller, N. (2019). Estimating models of supply and demand: Instruments and covariance restrictions. *Working Paper*.

Mankiw, N. G., & Whinston, M. (1986). Free entry and social inefficiency. *RAND Journal of Economics*, *17*, 48–58.

Marvel, H. P., & Peck, J. (1995). Demand uncertainty and returns policies. *International Economic Review*, *36*(3), 691–714.

Mazumdar, T., Raj, S. P., & Sinha, I. (2005). Reference price research: Review and propositions. *Journal of Marketing*, *69*(4), 84–102.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, *673*.

Meng, J., & Weng, X. (2018). Can prospect theory explain the disposition effect? a new perspective on reference points. *Management Science*, *64*(7), 3331–3351.

Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies*, *78*(1), 356–393.

Nevo, A. (2011). Empirical models of consumer behavior. *Annual Review of Economics*, *3*, 51–75.

Oestreicher-Singer, G., & Sundararajan, A. (2012). The visible hand? demand effects of recommendation networks in electronic markets. *Management science*, *58*(11), 1963–1981.

Online book services market size, share & trends analysis report by product (trade, education, science, technology & medicine), by region, and segment forecasts, 2020 - 2027 (Tech. Rep.). (2020). San Francisco, CA: Grand View Research.

Özer, Ö., & Zheng, Y. (2016). Markdown or everyday low price? the role of behavioral motives. *Management Science*, *62*(2), 326–346.

Pei, Z., Paswan, A., & Yan, R. (2014). E-tailer's return policy, consume's perception of return policy fairness and purchase intention. *Journal of Retailing and Consumer Services*, *21*(3), 249-257.

Pesendorfer, M. (2002). Retail sales: A study of pricing behavior in supermarkets. *The Journal of Business*, *75*(1), 33–66.

Reimers, I., & Waldfogel, J. (2017). Throwing the books at them: Amazon's puzzling long run pricing strategy. *Southern Economic Journal*, *83*(4), 869–885.

Resnick, S. I. (1987). *Extreme values, regular variation and point processes.* New York: Springer.

Romer, P. (1994). New goods, old theory, and the welfare costs of trade restrictions. *Journal of Development Economics*, *43*(1), 5–38.

Rose, N. L. (1985). The incidence of regulatory rents in the motor carrier industry. *The RAND Journal of Economics*, *16*(3), 299–318.

Schwert, G. W. (1981). Using financial data to measure effects of regulation. *The Journal of Law and Economics*, *24*(1).

Sheu, G., & Taragin, C. (2020). Simulating mergers in a vertical supply chain with bargaining. *Working Paper*.

Shiller, B. R. (2013). Digital distribution and the prohibition of resale markets for information goods. *Quantitative Marketing and Economics*, *11*(4), 403–435.

Spence, M. (1976a). Product differentiation and welfare. *American Economic Review*, *66*(2), 407–414.

Spence, M. (1976b). Product selection, fixed costs, and monopolistic competition. *American Economic Review*, *43*(2), 217–235.

Spiegler, R. (2012). Monopoly pricing when consumers are antagonized by unexpected price increases: a "cover version" of the heidhues–kőszegi–rabin model. *Economic Theory*, *51*(3), 695–711.

U.S. Census bureau, e-commerce retail sales as a percent of total sales [ECOMPCTSA]. (2003). Retrieved from `https://fred.stlouisfed.org/series/ECOMPCTSA` (FRED, Federal Reserve Bank of St.)

van Osselaer, S. M. J., & Alba, J. W. (2000). Consumer learning and brand equity. *Journal of Consumer Research*, *27*(1), 1–16.

Villas-Boas, J. M., & Winer, R. S. (1999). Endogeneity in brand choice models. *Management Science*, *45*(10), 1324–1338.

Zhang, X., Chan, T. Y., & Xie, Y. (2018). Price search and periodic price discounts. *Management Science*, *64*(2), 495–510.

Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science*, *32*(1), 153–169.

APPENDIX A

Appendix of Chapter 1

Proofs of Claims, Propositions, and Theorems

Proof of Theorem 1

Proof. Assuming symmetric prices the inverse demands when there are J_0 and J_1 varieties are parallel if and only if there exists a $d(J_0, J_1)$ such that for all p then $Q(p, J) = Q(p + d(J_0, J_1), J_1)$; that is

$$\mathbb{P}(\varepsilon_{0m} < \delta - \alpha p + \max_{1 \leq j \leq J_0} \varepsilon_j) = \mathbb{P}(\varepsilon_{0m} < \delta - \alpha(p + d(J_0, J_1)) + \max_{1 \leq j \leq J_1} \varepsilon_j).$$

Since ε_{0m} is independent of $\max_{1 \leq j \leq J_0} \varepsilon_j$ this can only be true if the distribution of the maxima for J_0 and J_1 of ε_j for $j \geq 1$ is the same, that is

$$\max_{1 \leq j \leq J_0} \varepsilon_j \quad \stackrel{d}{=} \quad -\alpha d(J_0, J_1) + \max_{1 \leq j \leq J_1} \varepsilon_j$$

Let F be the CDF of ε, then the equation above implies that for all natural number n there exists $t(n)$ such that for all x:

$$F(x) = F^n(x + t(n)).$$

Iterating on both sides implies

$$F^{nm}(x + t(nm)) = F^{nm}(x + t(n) + t(m))$$

we recognize an instance of the functional equation $t(nm) = t(n) + t(m)$ which has the unique solution $t(n) = c\log(n)$.[1] Therefore:

$$F(x) = F^y(x + c\log y),$$

letting $x = 0$, $s = c\log y$, we get $F(0) = F^{e^{s/c}}(s)$, and so:

$$F(s) = e^{\log F(0) e^{-s/c}},$$

which is a Gumbel distribution with location parameter $c\log(-\log F(0))$ and dispersion parameter c. This derivation proves that the parallel demands condition implies that the random utility shocks $(\varepsilon_{ij})_{j=1}^{\infty}$ follow the Gumbel distribution. Moreover, if the random utility shocks $(\varepsilon_{ij})_{j=1}^{\infty}$ follow the Gumbel distribution then $F(x) = e^{\log F(0) e^{-x/c}}$ and $F^n(x) = e^{\log F(0) e^{log(n) - x/c}} = F\left(x - clog(n)\right)$ and so parallel demands hold:

$$\mathbb{P}\left(\varepsilon_{0m} < \delta - \alpha p + \max_{1 \leq j \leq J_0} \varepsilon_j\right) = \mathbb{P}\left(\varepsilon_{0m} < \delta - \alpha(p + clog(J_1) - clog(J_0)) + \max_{1 \leq j \leq J_1} \varepsilon_j\right).$$

$\square$

Proof of Theorem 2

Proof. Let the random utility shocks (ε_j) be i.i.d. and distributed according to F in the domain of attraction of the Gumbel distribution. Let $G(x) = exp[-exp(-x)]$ be the

[1]It is easy to extend the formula for real numbers through rationals, note

$$F(x) = F^n(x + t(n)) = F^m(x + t(m))$$

implies

$$F(x) = F^{n/m}(x + t(n) - t(m)),$$

so we can consistently define $t(n/m) = t(n) - t(m)$.

Gumbel distribution. Then there exist sequences (a_n, b_n) such that

$$F^n(a_n x + b_n) \to G(x),$$

Furthermore, $\lim_{n\to\infty} \frac{a_n}{a_{[nt]}} = 1$ and $\lim_{n\to\infty} \frac{b_n - b_{[nt]}}{a_{[nt]}} = -c\log(t)$ for any $t > 0$ and some $c \in \mathbb{R}$ where $[nt]$ is the integer part of nt (see Resnick (1987) Chapter 1). Since the convergence $F^n(a_n x + b_n) \to G(x)$ is uniform (see Resnick (1987) Chapter 0) and F^n is uniformly continuous, then for any $\epsilon > 0$ there exists η and $N(\eta, \epsilon)$ such that for all $x \in \mathbb{R}$ and all $J_0, J_1 > N(\eta, \epsilon)$ we have $\left|\frac{a_{J_1}}{a_{J_0}} - 1\right| \leq \eta$ and

$$\begin{aligned}
\left|F^{J_0}(a_{J_0}x + b_{J_0}) - F^{J_1}(a_{J_0}x + b_{J_1})\right| \leq & \left|F^{J_0}(a_{J_0}x + b_{J_0}) - F^{J_1}(a_{J_1}x + b_{J_1})\right| \\
& + \left|F^{J_1}(a_{J_1}x + b_{J_1}) - F^{J_1}(a_{J_0}x + b_{J_1})\right| \\
< & \ \epsilon
\end{aligned}$$

Therefore, for any $p \in \mathbb{R}$

$$\begin{aligned}
& |Q(p, J_0) - Q\left(p + b_{J_1} - b_{J_0}, J_1\right)| \\
& = \left|\mathbb{P}\left(\max_{j\in\{1,\ldots,J_0\}} u_{ij}(p) > u_{i0}\right) - \mathbb{P}\left(\max_{j\in\{1,\ldots,J_1\}} u_{ij}(p + b_{J_1} - b_{J_0}) > u_{i0}\right)\right| \\
& = \left|\int_{\mathbb{R}} \left(F^{J_1}\left(\varepsilon_{i0} - \alpha(y - p) - \delta + \alpha(b_{J_1} - b_{J_0})\right) - F^{J_0}\left(\varepsilon_{i0} - \alpha(y-p) - \delta\right)\right) f_0(\varepsilon_{i0}) d\varepsilon_{i0}\right| \\
& < \epsilon
\end{aligned}$$

where f_0 is the probability density of ε_{i0}. We conclude that the inverse aggregate demands are asymptotically parallel. $\square$

Proof of Proposition 1

Proof. Redefine $\widetilde{\varepsilon}_{i0} = \varepsilon_{i0} - (1-\sigma)\nu_i$. Then the proof follows from Theorem 1. $\square$

Proof of Theorem 3

Proof. Let F be the CDF of the random utility shocks. Define Condition A as: for all $(\delta_n)_{n=1}^{J_0}$ bounded vector of real non-negative numbers there exists $f((\delta_n)_{n=1}^{J_0})$ such that $F(x) = \Pi_{n=1} F\left(x - \delta_n + f((\delta_n)_{n=1}^{J_0})\right)$. Theorem 1 applies for vectors of constants $(\delta, \dots, \delta)$ of any size, and shows that the only possible candidate CDF F that satisfies condition A must be Gumbel. Therefore if Condition A is going to hold for any $(\delta_n)_{n=1}^{J_0}$ bounded vector of real non-negative numbers, then F must be Gumbel. Condition A is a rephrasing of parallel WTP CDFs and so, Gumbel is necessary for parallel WTP CDFs.

Moreover, if $(\epsilon_{ij})_{j=1}^{\infty}$ are i.i.d. Gumbel then $\delta_j + \epsilon_{ij} \sim F_j(\mu_j, \beta)$ are also Gumbel, where μ_j is the position parameter of the Gumbel distribution and β is the scale parameter ($\{\mu_j\}$ is well defined, because $\{\delta_j\}$ is bounded.) Then

$$\begin{aligned}\mathbb{P}\left(\delta_j + \epsilon_{ij} < x\right) &= F_j(x) \\ &= exp\left(-exp\left(\frac{\mu_j - x}{\beta}\right)\right).\end{aligned}$$

Let $j^* = argmax_{j\in J_0}\{\delta_j + \epsilon_{ij}\}$, we have

$$\begin{aligned}F_{j^*}(x) &= \Pi_{j\in J_0} F_j(x) \\ &= exp(-\Sigma_{j\in J_0} exp(\frac{\mu_j - x}{\beta})) \\ &= exp(-exp(\frac{\mu - x}{\beta})),\end{aligned}$$

where $\mu = \beta log \sum_{j \in J_0} exp(\frac{\mu_j}{\beta})$.

Similarly, let $j^{**} = argmax_{j \in J_1} \{\delta_j + \epsilon_{ij}\}$ for $J_1 \neq J$. We have

$$F_{j^{**}}(x) = exp\left(-exp\left(\frac{\mu' - x}{\beta}\right)\right)$$

where $\mu' = \beta log \sum_{j \in J_1} exp(\frac{\mu_j}{\beta})$. The above derivation shows that we have parallel WTP distributions by letting

$$\begin{aligned} t_{J_1} &= \mu' - \mu \\ &= \beta log \frac{\sum_{j \in J_1} exp(\frac{\mu_j}{\beta})}{\Pi \sum_{j \in J_0} exp(\frac{\mu_j}{\beta})}. \end{aligned}$$

□

Proof of Theorem 4

Proof. Take (α_n, β_n) and the nondegenarate CDF H such that $\Pi_{j=1}^n F(\alpha_n x + \beta_n - \delta_j) \to H(x)$ for all x. Because

$$F^n(\alpha_n x + \beta_n) \leq \Pi_{j=1}^n F(\alpha_n x + \beta_n - \delta_j) \leq F^n(\alpha_n x + \beta_n)$$

and by continuity, there exists $\gamma_n \in [0, K]$ such that $\Pi_{j=1}^n F(\alpha_n x + \beta_n - \delta_j) = F^n(\alpha_n x + \gamma_n) \to H(x)$. But because F is in the domain of attratcion of the Gumbel, by Proposition 0.2 of Resnick (1987) there exists a and b such that $H(x) = G(ax + b)$ is a rescaling of the Gumbel distribution.

The rest of the proof follows exactly the same steps as the proof of Theorem 1, starting from $\Pi_{j=1}^n F(\alpha_n x + \beta_n - \delta_j) \to G(x)$. We have $\lim_{n \to \infty} \frac{a_n}{a_{[nt]}} = 1$ and $\lim_{n \to \infty} \frac{\gamma_n - \gamma_{[nt]}}{a_{[nt]}} =$

$-c\log(t)$ for any $t > 0$ and some $c \in \mathbb{R}$ where $[nt]$ is the integer part of nt (see Resnick (1987) Chapter 1).

Since the convergence $\Pi_{j=1}^{n} F(\alpha_n x + \beta_n - \delta_j) = F^n(\alpha_n x + \gamma_n) \to G(x)$ is uniform (see Resnick (1987) Chapter 0) and F^n is uniformly continuous, then for any $\epsilon > 0$ there exists η and $N(\eta, \epsilon)$ such that for all $x \in \mathbb{R}$ and all $J_0, J_1 > N(\eta, \epsilon)$ we have $\left|\frac{a_{J_1}}{a_{J_0}} - 1\right| \leq \eta$ and

$$\begin{aligned}
\left|F^{J_0}(a_{J_0}x + \gamma_{J_0}) - F^{J_1}(a_{J_0}x + \gamma_{J_1})\right| &\leq \left|F^{J_0}(a_{J_0}x + \gamma_{J_0}) - F^{J_1}(a_{J_1}x + \gamma_{J_1})\right| \\
&\quad + \left|F^{J_1}(a_{J_1}x + \gamma_{J_1}) - F^{J_1}(a_{J_0}x + \gamma_{J_1})\right| \\
&< \epsilon
\end{aligned}$$

Therefore, for any $p \in \mathbb{R}$

$$\begin{aligned}
&\left|\mathbb{P}\left(WTP_i(J_0) \leq x\right) - \mathbb{P}\left(WTP_i(J_1) \leq x + \frac{\gamma_{J_1} - \gamma_{J_0}}{\alpha}\right)\right| \\
&= \left|\mathbb{P}\left(\max_{j \in \{1,\ldots,J_0\}} \{\frac{\delta_j + \varepsilon_{ij} - \varepsilon_{i0}}{\alpha}\} \leq x\right) - \mathbb{P}\left(\max_{j \in \{1,\ldots,J_1\}} \{\frac{\delta_j + \varepsilon_{ij} - \varepsilon_{i0}}{\alpha}\} \leq x + \frac{\gamma_{J_1} - \gamma_{J_0}}{\alpha}\right)\right| \\
&= \left|\int_{\mathbb{R}} \left(F^{J_1}\left(\alpha x + \varepsilon_{i0} - \delta_j + \gamma_{J_1} - \gamma_{J_0}\right) - F^{J_0}\left(\alpha x + \varepsilon_{i0} - \delta_j\right)\right) f_0(\varepsilon_{i0}) d\varepsilon_{i0}\right| \\
&< \epsilon
\end{aligned}$$

where f_0 is the probability density of ε_{i0}. We conclude that the willingness-to-pay densities are asymptotically parallel. $\square$

Proof of Proposition 2

Proof. Assume parallel demands (Definition 1) and let $d(J_0, J_1)$ be such that $P(Q, J_0) + d(J_0, J_1) = P(Q, J_1)$. Then $\Lambda = \int_0^Q \big(P(s, J_1) - P(s, J_0)\big) ds = d(J_0, J_1) * Q$. $\square$

Proof of Proposition 3

Proof. Observe:

$$
\begin{aligned}
d(J_0, J_1) &= p_1 - P(Q_1, J_0) \\
&= \left(\frac{p_1 - p_0}{Q_1 - Q_0} - \frac{P(Q_1, J_0) - p_0}{Q_1 - Q_0} \right) (Q_1 - Q_0)
\end{aligned}
$$

Now assume $(p(J), Q(J))_{J \in \mathbb{R}}$ is a continuously differentiable interpolation of $(p(J), Q(J))_{J \in \mathbb{N}}$ which exists by the Stone-Weierstrass theorem. Then by the Taylor approximation theorem:

$$
\begin{aligned}
d(J_0, J_1) &= \left(\frac{p_1 - p_0}{Q_1 - Q_0} - \frac{P(Q_1, J_0) - p_0}{Q_1 - Q_0} \right) (Q_1 - Q_0) \\
&= \left(\frac{\frac{dp}{dJ}}{\frac{dQ}{dJ}} - \frac{\partial P(Q, J)}{\partial Q} \right) \frac{dQ}{dJ} \triangle J + O((\triangle J)^2)
\end{aligned}
$$

□

Proof of Proposition 5

Proof. Let $d = d(J_0, J_1)$. Observe by assumption $Q_{J_1}(\mathbf{p_{J_1}}) = Q_{J_0}(\mathbf{p_{J_0}} + (\rho - d)\mathbf{1_{J_0}})$, then the second part of the theorem follows directly from the first-order Taylor approximation:

$$
Q_{J_1}(\mathbf{p_{J_1}}) = Q_{J_0}(\mathbf{p_{J_0}}) + (\rho - d) \frac{dQ_{J_0}(\mathbf{p_{J_0}} + t\mathbf{1_{J_0}})}{dt} + O\left((\rho - d)^2\right)
$$

where $\frac{dQ_{J_0}(\mathbf{p_{J_0}} + t\mathbf{1_{J_0}})}{dt}$ is the directional derivative in the direction $\mathbf{1_{J_0}}$. And so

$$
d = \left(\frac{\rho}{\Delta Q} - \left(\frac{dQ_{J_0}(\mathbf{p_{J_0}} + t\mathbf{1_{J_0}})}{dt} \right)^{-1} \right) \Delta Q + O\left((\rho - d)^2\right)
$$

$\square$

APPENDIX B

Appendix of Chapter 2

B.1. Model Details

B.1.1. BNE vs. Commitment Solution

In solving for the retailer's pricing problem, there are two solution concepts: the non-commitment (BNE) solution and the commitment solution. The difference between the two solution concepts is as follows. The BNE solution solves

$$r^* = \arg\max_r \pi(r, F), \qquad \forall r^* \in F$$

while the commitment solution solves

$$F^* = \arg\max_F \int \pi(r, F) dF(r).$$

In the commitment solution, the retailer can commit to a pricing policy ex ante and chooses a price distribution to maximize the expected profit, accounting for the effect that the price distribution has on consumer reference expectations. Thus, in the commitment solution, the retailer may set prices that are not optimal in the sense that for some $r' \in F^*$, $r' \notin \arg\max \pi(r, F^*)$.

Gentry & Pesendorfer (2021b) shows that the optimal pricing policy of the commitment solution is also a price distribution instead of a single price. Figure B1 plots the retailer's optimal pricing policy and profit function under BNE and commitment solutions. Without commitment, every price in the support of the equilibrium price distribution must be individually optimal. With commitment, (almost) all prices charged by the monopolist are individually sub-optimal.

The qualitative effects of consumers' bargain-hunting behavior on retailer's profit also depend on whether or not the retailer can commit ex ante to individually sub-optimal prices. If such commitment is feasible, then the retailer will exploit bargain-hunting consumers to increase prices and profits. If not, then the retailer's temptation to exploit bargain-hunting behavior will reduce prices and profits.

The literature on monopoly pricing under reference dependence has focused mainly on the commitment solutions, e.g., Spiegler (2012) and Heidhues & Kőszegi (2014). In this paper, we instead focus on the BNE solution since in the book retail market, it is not clear that there exist formal commitment devices for the retailer, i.e., the retailer is not able to commit ex ante to individually sub-optimal prices.

B.1.2. Proof of Retailer's Pricing Equilibrium

We show that the BNE involves a mix-strategy equilibrium where all prices in the support give the retailer the same profit. The proof follows the proof of Proposition 4 in Gentry & Pesendorfer (2021b). The retailer's pricing problem is

$$\max_{r} \quad \pi(r, F) = (r - c) \cdot D(r, F).$$

As we have shown in Equation (2.5),

$$\begin{aligned}
D(r,F) &\equiv a - br + \int_{\underline{r}}^{\bar{r}} \rho(R-r)dF(R) \\
&= a - br + \delta^{+}\int_{r}^{\bar{r}} (R-r)dF(R) + \delta^{-}\int_{\underline{r}}^{r} (R-r)dF(R) \\
&= (a+\delta^{-}r_e) - (b+\delta^{-})\cdot r + (\delta^{+}-\delta^{-})\cdot \int_{r}^{\bar{r}} (R-r)dF(R) \\
&\equiv a' - b'r + \Delta \int_{r}^{\bar{r}} (R-r)dF(R)
\end{aligned}$$

where $a' \equiv a + \delta^{-}r_e, b' \equiv b + \delta^{-}$, and $\Delta \equiv \delta^{+} - \delta^{-}$.

For F to be a mixed-strategy equilibrium, all r must yield the same profit. In particular, every retail rate in the support must yield the same profit as the supremum rate $\bar{r}$: $\pi(r,F) = \pi(\bar{r},F)$, $\forall r \in [\underline{r},\bar{r}]$. Dividing both profit functions by $r-c$, we obtain the equivalent indifference condition

$$D(r,F) - \frac{(\bar{r}-c)\cdot(a'-b'\bar{r})}{r-c} = 0, \qquad \forall r \in [\underline{r},\bar{r}].$$

Since the left-hand side is a constant that equals zero over the support of r, we must have the derivative with respect to r equals 0 for all r in the support:

$$-b' - \Delta(1-F(r)) + \frac{(\bar{r}-c)(a'-b'\bar{r})}{(r-c)^2} = 0, \qquad \forall r \in [\underline{r},\bar{r}].$$

Using the fact that $F(\bar{r}) = 1$, we have

$$\bar{r} = \frac{a+\delta^{-}r_e}{2(b+\delta^{-})} + \frac{c}{2}. \tag{B.1}$$

Also, since $F(\underline{r}) = 0$, we have

$$\text{(B.2)} \qquad \underline{r} = \sqrt{\frac{(\bar{r}-c)(a'-b'\bar{r})}{b+\delta^+}} + c = (\bar{r}-c)\sqrt{\frac{b+\delta^-}{b+\delta^+}} + c.$$

Therefore,

$$F(r) = 1 - \frac{b+\delta^-}{\delta^+ - \delta^-}\left(\frac{(\bar{r}-c)^2}{(r-c)^2} - 1\right).$$

The last step is to find r_e. Recall that the equal profit condition requires that the profit function is a constant for $r \in [\underline{r}, \bar{r}]$. Thus, the derivative of profit with respect to r must equal to zero for $r \in [\underline{r}, \bar{r}]$. Since $\pi(r,F) = (r-c)\cdot D(r,F)$, we have

$$\begin{aligned}\frac{\partial \pi(r,F)}{\partial r} = a - br + \delta^+ \int_r^{\bar{r}} (R-r)dF(R) + \delta^- \int_{\underline{r}}^{r} (R-r)dF(R) \\ + (r-c)\cdot(-b - \delta^+(F(\bar{r}) - F(r)) - \delta^-(F(r) - F(\underline{r}))).\end{aligned}$$

We use the fact that $\frac{\partial \pi(r,F)}{\partial r}\big|_{r=\underline{r}} = 0$ to find

$$\begin{aligned}\frac{\partial \pi(r,F)}{\partial r}\Big|_{r=\underline{r}} &= a - b\underline{r} + \delta^+ \int_{\underline{r}}^{\bar{r}} (R-\underline{r})dF(R) + (\underline{r}-c)\cdot(-b-\delta^+) \\ &= a - b\underline{r} + \delta^+(r_e - \underline{r}) + (\underline{r}-c)\cdot(-b-\delta^+) \\ &= a + \delta^+ r_e + (b+\delta^+)c - 2(b+\delta^+)\underline{r} = 0.\end{aligned}$$

Thus,

$$\text{(B.3)} \qquad \underline{r} = \frac{a+\delta^+ r_e}{2(b+\delta^+)} + \frac{c}{2}.$$

Lastly, we use Equation (B.1), (B.2) and (B.3) to solve for

$$r_e = \frac{a(\sqrt{b+\delta^+} - \sqrt{b+\delta^-}) + c((b+\delta^+)\sqrt{b+\delta^-} - (b+\delta^-)\sqrt{b+\delta^+})}{\delta^+\sqrt{b+\delta^-} - \delta^-\sqrt{b+\delta^+}}.$$

B.1.3. Proof of Proposition 1

We first show the upstream and downstream profits in the BNE with general δ^- values. Let the list price $l = 1$ and base share $s^0 = 1$ to simplify notation. Also, we impose the technical condition that $\frac{a}{b} > c^U + c^D$ to ensure that demand is positive.

In the mix-strategy equilibrium, the retailer earns identical profits for all prices in the support. Specifically, the retailer's profit for all prices equals to the profit at the supremum price $\bar{r}$. From Equation (2.5), it equals to the profit under the simple linear demand curve $a' - b'r$. Therefore,

$$\pi^D = \frac{(a' - b'c)^2}{4b'} = \frac{(a + \delta^- r_e - (b+\delta^-)c)^2}{4(b+\delta^-)}$$

where $c = w + c^D$.

Now we can derive the publisher's profit. Since retailer's profit is the same for all prices, the total quantity is

$$q = \int_{\underline{r}}^{\bar{r}} \frac{\pi^D}{r - c} dF(r).$$

Using Equation (2.6), we have

$$\begin{aligned}
\pi^U &= (w - c^U) \cdot q \\
&= (c - c^U - c^D) \cdot \pi^D \cdot \int_{\underline{r}}^{\bar{r}} \frac{1}{r-c} dF(r) \\
&= (c - c^U - c^D) \cdot \pi^D \cdot \frac{2(b+\delta^-)(\bar{r}-c)^2}{3(\delta^+ - \delta^-)} \cdot \left(\frac{1}{(\underline{r}-c)^3} - \frac{1}{(\bar{r}-c)^3} \right).
\end{aligned}$$

The bargaining solution is found by maximizing the Nash product. The first order condition is

$$\lambda \pi^D \pi^{U'} + (1-\lambda) \pi^U \pi^{D'} = 0$$

where primes indicate derivatives with respect to w (equivalently c).

After the vertical merger, the profit of the joint entity is

$$\pi^M = \frac{(a + \delta^- \tilde{r}_e - (b+\delta^-)\tilde{c})^2}{4(b+\delta^-)}$$

where $\tilde{c} = c^U + c^D$ and $\tilde{r}_e$ is the value of r_e when $c = \tilde{c}$.

Now consider the special case when $\delta^- = 0$, which helps to simplify the math and facilitate comparison of the scenarios with and without price reference effects. When $\delta^- = 0$,

$$\pi^D = \frac{(a-bc)^2}{4b}$$

and

$$\begin{aligned}
\pi^U &= (c - c^U - c^D) \cdot \frac{(a-bc)^2}{4b} \cdot \frac{(a-bc)^2}{2b \cdot 3\delta^+} \cdot \frac{1}{(\frac{a-bc}{2b})^3} \cdot \frac{(b+\delta^+)^{\frac{3}{2}} - b^{\frac{3}{2}}}{b^{\frac{3}{2}}} \\
&= (c - c^U - c^D) \cdot (a-bc) \cdot \frac{(b+\delta^+)^{\frac{3}{2}} - b^{\frac{3}{2}}}{3b^{\frac{1}{2}}\delta^+}.
\end{aligned}$$

Plug into the first order condition of the Nash product and solve for

$$c^\star = \frac{a\lambda}{2b} + \frac{2-\lambda}{2}(c^U + c^D).$$

We notice three things. First, $c^\star$ does not depend on δ^+. Thus, the wholesale price (and as a result the retailer's profit) does not depend on δ^+ when $\delta^- = 0$. Second, when $\lambda = 0$, $c^\star = c^U + c^D$ and $w^\star = c^\star - c^D = c^U$. This is the case when the retailer has all the bargaining power and would set the wholesale price to be the marginal cost of the publisher. Third, $c^U + c^D < c^\star < \frac{1}{2} \cdot \frac{a}{b} + \frac{1}{2}(c^U + c^D) < \frac{a}{b}$ for $\lambda \in (0, 1)$.

Then the pre-merger joint profits are

$$\pi^{\star D} + \pi^{\star U} = \frac{(a - bc^\star)^2}{4b} + (c^\star - c^U - c^D) \cdot (a - bc^\star) \cdot \frac{(b + \delta^+)^{\frac{3}{2}} - b^{\frac{3}{2}}}{3b^{\frac{1}{2}}\delta^+}.$$

The pre-merger joint profits are strictly increasing in δ^+, which corresponds to the increasing curve in Figure 2.5. As δ^+ approaches 0 from above, the joint profits converge to

$$\lim_{\delta^+ \to 0} (\pi^{\star D} + \pi^{\star U}) = \frac{(a - bc^\star)^2}{4b} + (c^\star - c^U - c^D) \cdot (a - bc^\star) \cdot \frac{1}{2}.$$

Notice that the post-merger joint profits are

$$\pi^{\star M} = \frac{(a - b(c^U + c^D))^2}{4b}.$$

Thus,

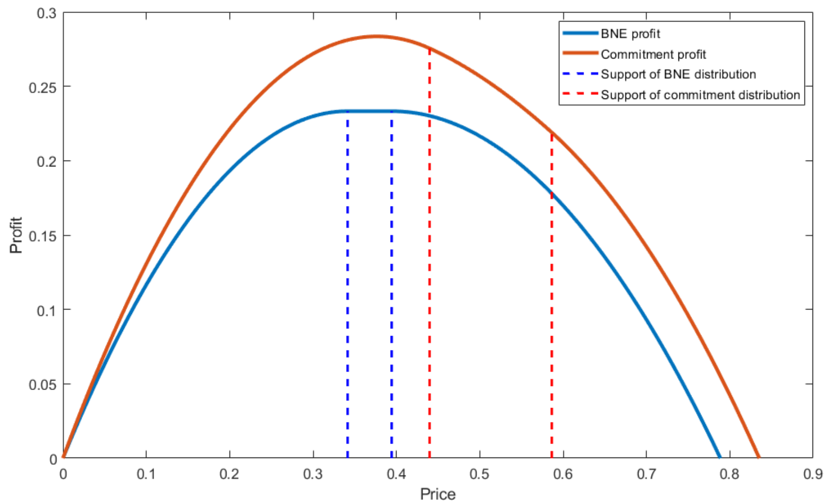
$$\pi^{\star M} - \lim_{\delta^+ \to 0} (\pi^{\star D} + \pi^{\star U}) = \frac{b(c^\star - c^U - c^D)^2}{4} > 0$$

as long as $\lambda \neq 0$. This means when $\delta^+ \to 0$, the post-merger joint profits are higher than the pre-merger joint profits.

However, we notice that $\pi^{\star D} + \pi^{\star U}$ is increasing in δ^+ and converges to $+\infty$ when $\delta^+ \to +\infty$, while $\pi^{\star M}$ stays as a constant. Thus, the post-merger joint profits are lower than the pre-merger joint profits for a significantly large δ^+. The proposition then follows from the continuity of the equilibrium joint profits with respect to δ^+.

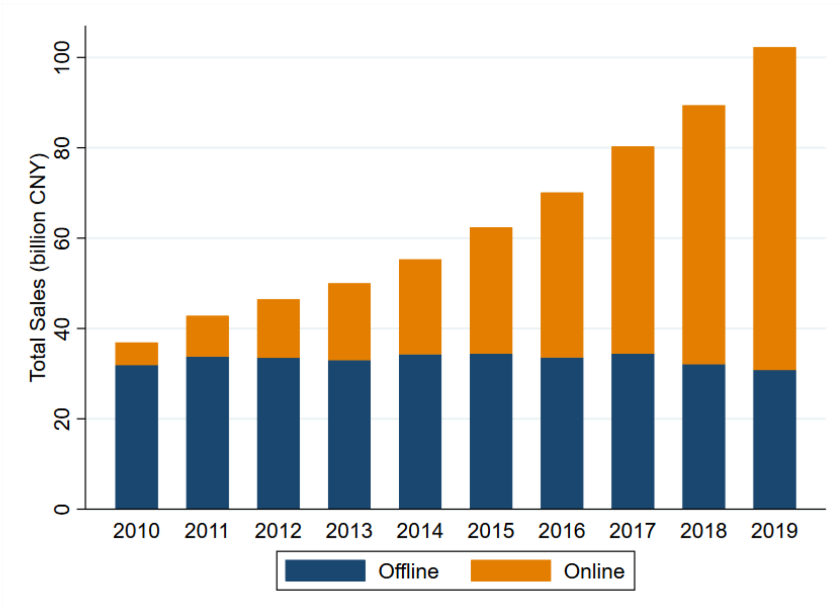
B.2. Additional Figures

Figure B1. BNE vs. Commitment Solution



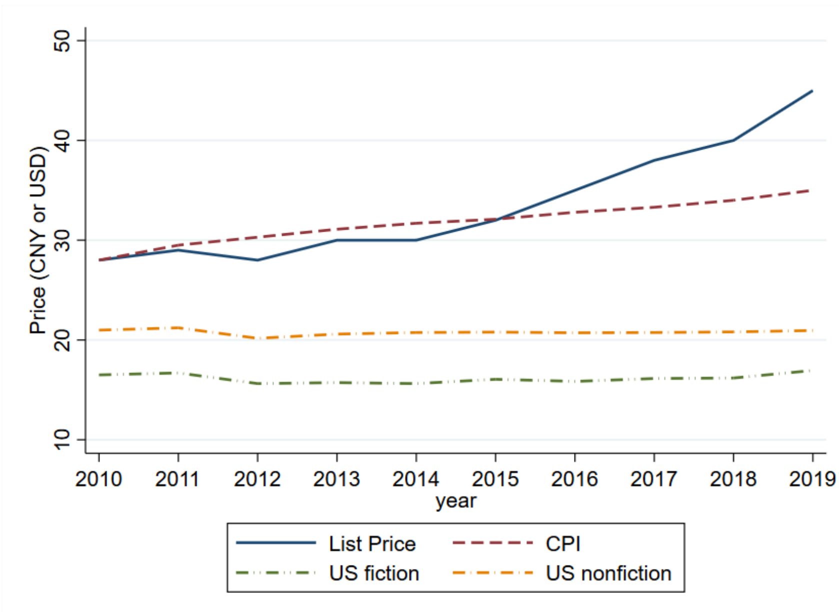
Notes: Illustration of retailer's optimal pricing policy and profit function under BNE and commitment solutions. $a = 1, b = 1, c = 0, \delta^{+} = 1$, $\delta^{-} = 0.5$. $\pi^{commitment} = 0.253 > \pi^{BNE} = 0.233$.

Figure B2. Total Sales



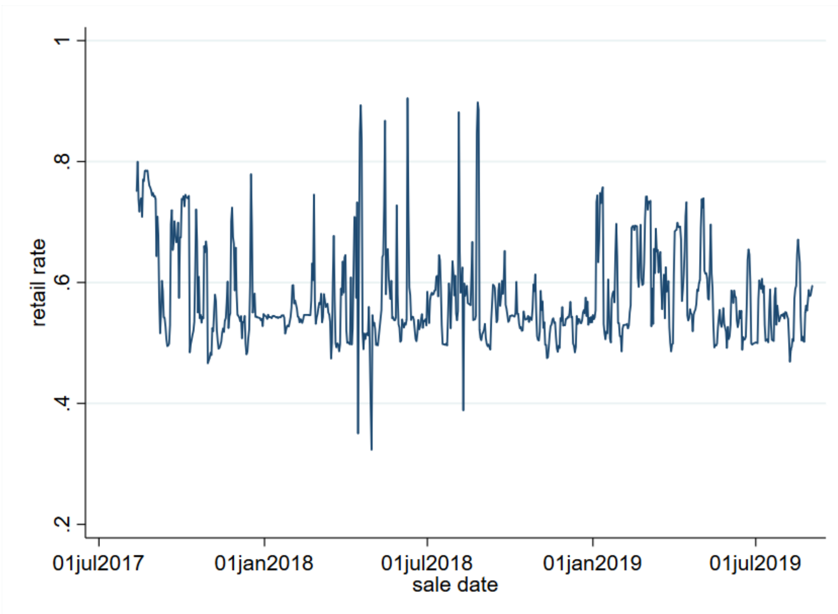
Notes: This figure plots the total sales (measured by list price times total quantity) of online and offline channels in the book retail market in China. 1 USD ≈ 6.6 CNY in 2018. Data sources: 2020 China Book Retail Market Report.

Figure B3. List Price



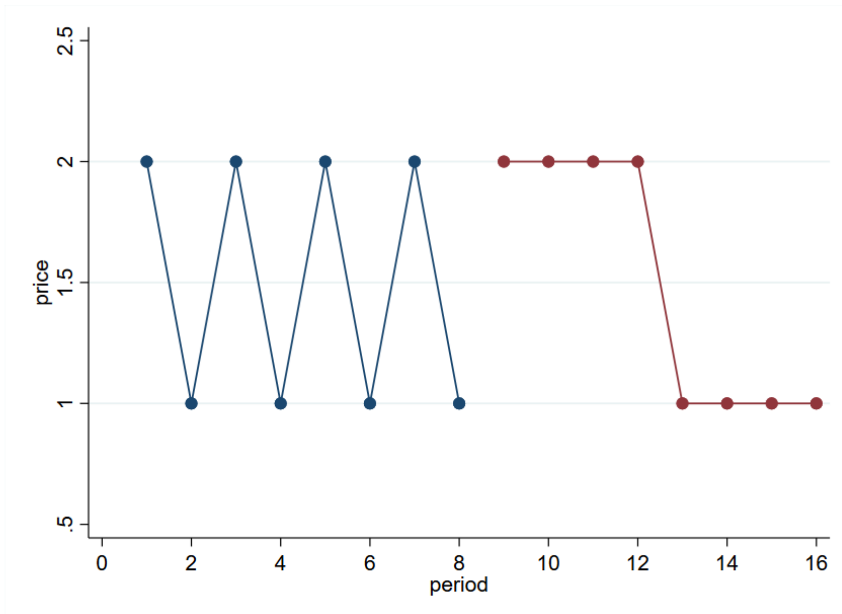
Notes: This figure plots the median list price of books published between 2010 to 2019 in China and the US (CNY and USD, respectively). 1 USD $\approx$ 6.6 CNY in 2018. The CPI in China is normalized to the 2010 level. The list prices for fiction and non-fiction books in the US are reported separately. Data sources: 2020 China Book Retail Market Report; Average Book Prices 2019 (https://tln.lib.mi.us/dept/technical-services/acq/files/AverageBookPrices2019.pdf)

Figure B4. Retail Rate of the Book *To Live*



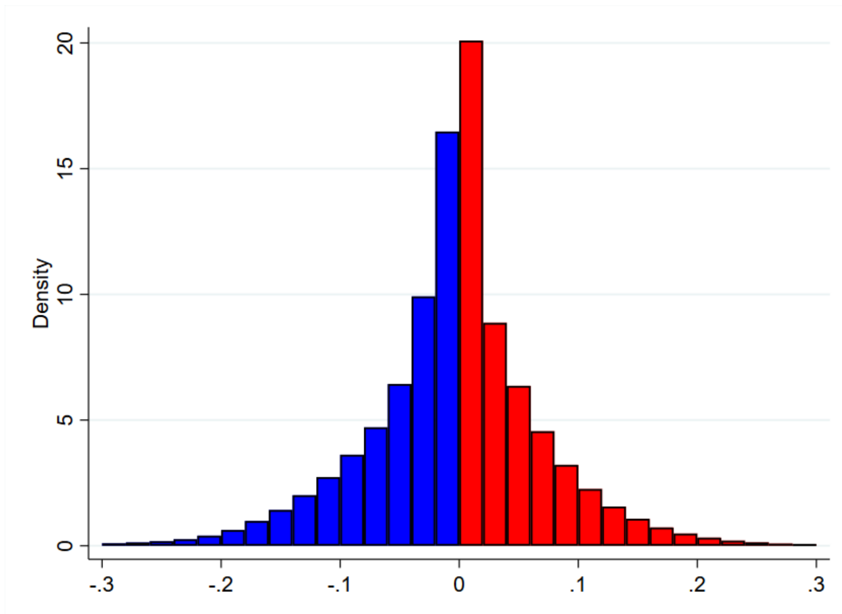
Notes: This figure plots the daily retail rate of the book *To Live* (2017 version) from Aug 12, 2017 to Sept 2, 2019.

Figure B5. Price Persistence



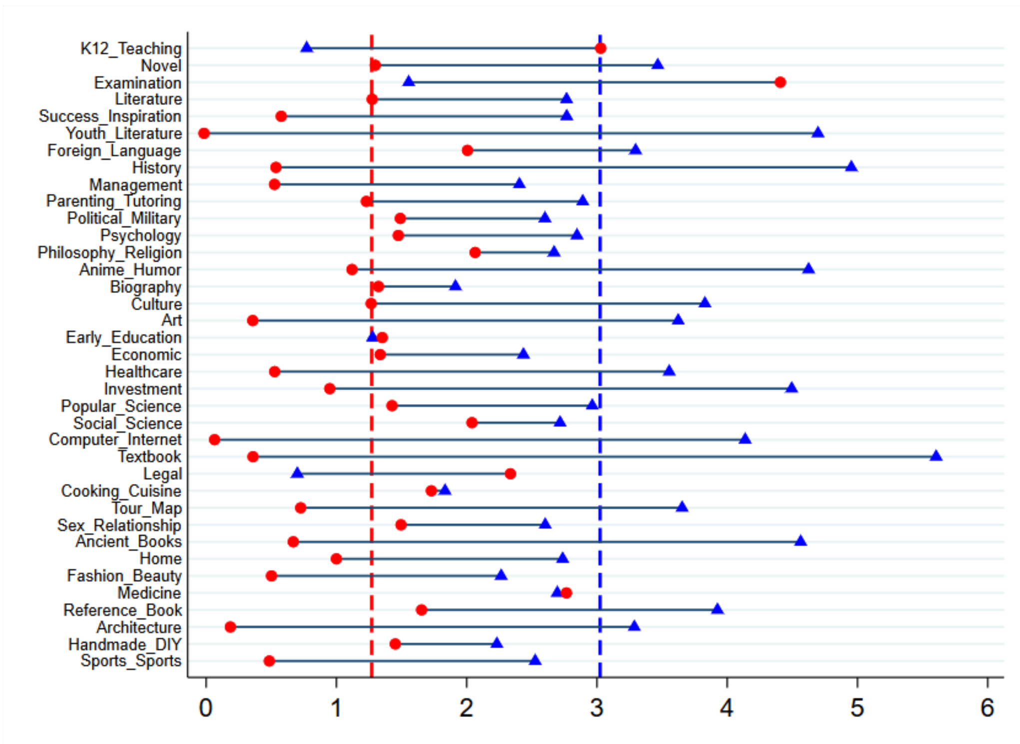
Notes: This figure illustrates the difference between a periodical high-low price pattern and a persistent price pattern.

Figure B6. Reference Effect Histogram



Notes: This figure plots the empirical reference effects pooled across all books. The positive part corresponds to the term that multiples δ^+ while the negative part corresponds to the term that multiples δ^- in Equation (2.9).

Figure B7. Heterogeneity of δ^{+} and δ^{-} by Book Category



Notes: This figure plots the estimated δ^{+} and δ^{-} for each category. The triangles (circles) represent the estimated values of δ^{+} (δ^{-}). The dashed lines are the average values of the δ^{+} and δ^{-} estimates (not weighted by the number of books in each category).

B.3. Additional Tables

Table B1. Full Sample

	Obs	Mean	S.D.	p25	p50	p75
	Title level					
Publish date	1,044,464	2014.56	3.82	2012.41	2015.24	2017.50
Quantity	1,044,464	515.0	7,175	6	25	113
List price	1,044,464	61.61	579.0	28	38	58
Retail rate	1,044,464	0.662	0.118	0.582	0.667	0.747
Wholesale rate	1,044,464	0.593	0.094	0.550	0.620	0.650
	Title-date level					
Sale date	81,445,362	2018.38	0.76	2017.72	2018.40	2019.04
Publish date	81,445,362	2015.21	3.38	2013.91	2016.00	2017.41
Quantity	81,445,362	6.604	70.41	1	1	3
List price	81,445,362	52.73	118.6	28.5	39	56
Retail rate	81,445,362	0.622	0.159	0.500	0.641	0.753
Wholesale rate	81,445,362	0.564	0.094	0.500	0.600	0.630

Notes: This table presents summary statistics of the full sample of books sold on DangDang.com from Jan 1, 2017 to Sept 2, 2019. List prices are in the unit of CNY. 1 USD $\approx$ 6.6 CNY in 2018. Retail and wholesale rates are the ratio of retail and wholesale prices over the list price.

Table B2. Demand Estimates: Robustness

	(1) OLS	(2) IV	(3) OLS	(4) IV
Price	-0.0269***	-0.0144***	-0.00574***	-0.0100***
	(0.00168)	(0.00274)	(0.000504)	(0.00108)
$\text{Price}^2/100$	0.00490***	0.00128*		
	(0.000577)	(0.000735)		
$\log(s_{j\|g})$			0.790***	0.788***
			(0.00278)	(0.00279)
δ^+	2.098***	2.573***	1.211***	0.984***
	(0.102)	(0.136)	(0.0453)	(0.0654)
δ^-	1.255***	1.711***	0.577***	0.390***
	(0.0803)	(0.107)	(0.0379)	(0.0547)
Book FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
First-stage F Stat		715.33		634.21
R-squared	0.619		0.876	
Obs	2,217,222	2,214,612	2,217,222	2,214,612

Notes: This table reports the estimated coefficients of the model in Equation (2.9). The unit of observation is a book-date. The dependent variable is log of quantity sold. All specifications include book and date level fixed effects. The IV specifications use the average lagged prices in the previous week as an instrument for price. The nested share is calculated as the fraction of quantity sold over quantity sold of all books in the same category for that date. Standard errors are clustered at the book level and presented in parentheses below the coefficients. *p<0.1, **p<0.05, ***p<0.01.

Table B3. Top 10 Publishers' Quantity, Title and Revenue Shares

Rank	Publisher	Quantity	Title	Revenue
1	Beijing United Publishing Company	3.28	0.51	2.95
2	Hunan Literature and Art Publishing House	3.03	0.33	2.75
3	CITIC Publishing House	2.47	0.56	3.73
4	The Commercial Press	2.39	0.96	2.55
5	Posts & Telecom Press	2.22	1.80	2.57
6	Nan Hai Publishing Company	2.16	0.17	2.14
7	People's Literature Publishing House	2.14	0.48	2.07
8	China Construction Industry Press	2.11	1.24	2.39
9	Writers Publishing House	2.10	0.36	1.51
10	Foreign Language Teaching and Research Press	1.93	0.65	2.04
Total		23.83	7.05	24.70

Notes: This table reports the number of quantity sold, number of unique titles, and total revenues of the top 10 publishers ranked by quantity sold as the percentage of the total values across all publishers. Numbers are in percentage.

Table B4. Bargained Wholesale Price and Publisher Size

	Dependent variable: Wholesale rate		
	(1)	(2)	(3)
Publisher Size	0.00466***	0.00478***	0.00562***
	(0.00222)	(0.00222)	(0.00109)
Publish Year FE	No	Yes	Yes
Category FE	No	No	Yes
Dep. Var. Mean	0.522	0.522	0.522
R-squared	0.003	0.007	0.206
Obs	6,222	6,222	6,222

Notes: This table reports regression results of the wholesale rate on the publisher size controlling for different book-level fixed effects. Publisher size is defined as the log of total quantity sold in the sample period for books published by the publisher.

APPENDIX C

Appendix of Chapter 3

C.1. Additional Tables

Table C1. Results of Full Sample: Prices

VARIABLES	(1) log_price	(2) log_price	(3) log_price	(4) log_price	(5) log_price	(6) log_price
post_event	-0.00276	0.00480*	-0.00308	-0.00394	0.00932	0.0247***
	(0.00261)	(0.00259)	(0.00260)	(0.00271)	(0.00775)	(0.00711)
Constant	3.442***	3.436***	3.442***	3.442***	3.296***	3.285***
	(0.00182)	(0.00180)	(0.00181)	(0.00189)	(0.00579)	(0.00531)
Observations	147,737	147,737	147,737	146,389	147,737	147,737
R-squared	0.881	0.891	0.882	0.891	0.854	0.867
Book Time Trend	N	Y	N	N	N	Y
Cate Time Trend	N	N	Y	Y	N	N
Cate Week FE	N	N	N	Y	N	N
Pre Sale Weight	N	N	N	N	Y	Y

Robust standard errors in parentheses

Notes: This table reports the estimation results from the event study analysis. The unit of observation is a book-week. The dependent variable is log price. All specifications include book and week level fixed effects, as well as week since publication fixed effects. Standard errors are clustered at the book level and presented in parentheses below the coefficients. *p<0.1, **p<0.05, ***p<0.01.

Table C2. Results of Selected Sample: Prices

VARIABLES	(1) log_price	(2) log_price	(3) log_price	(4) log_price	(5) log_price	(6) log_price
post_event	-0.000327	0.0153***	-0.000345	-0.00212	0.0221**	0.0302***
	(0.00383)	(0.00382)	(0.00380)	(0.00387)	(0.00993)	(0.0101)
Constant	3.367***	3.357***	3.367***	3.370***	3.288***	3.282***
	(0.00260)	(0.00260)	(0.00258)	(0.00263)	(0.00710)	(0.00720)
Observations	75,019	75,019	75,019	74,368	75,019	75,019
R-squared	0.883	0.894	0.884	0.892	0.884	0.897
Book Time Trend	N	Y	N	N	N	Y
Cate Time Trend	N	N	Y	Y	N	N
Cate Week FE	N	N	N	Y	N	N
Pre Sale Weight	N	N	N	N	Y	Y

Notes: This table reports the estimation results from the event study analysis. The unit of observation is a book-week. The dependent variable is log price. All specifications include book and week level fixed effects, as well as week since publication fixed effects. Standard errors are clustered at the book level and presented in parentheses below the coefficients. *p<0.1, **p<0.05, ***p<0.01.

Table C3. Results of Full Sample: Sales

VARIABLES	(1) log_quant	(2) log_quant	(3) log_quant	(4) log_quant	(5) log_quant	(6) log_quant
post_event	0.118***	0.0244	0.102***	0.103***	-0.0478	-0.323***
	(0.0188)	(0.0154)	(0.0187)	(0.0194)	(0.0729)	(0.0689)
Constant	2.581***	2.646***	2.592***	2.591***	4.588***	4.794***
	(0.0131)	(0.0107)	(0.0130)	(0.0135)	(0.0544)	(0.0515)
Observations	147,747	147,747	147,747	146,399	147,747	147,747
R-squared	0.756	0.808	0.761	0.779	0.791	0.859
Book Time Trend	N	Y	N	N	N	Y
Cate Time Trend	N	N	Y	Y	N	N
Cate Week FE	N	N	N	Y	N	N
Pre Sale Weight	N	N	N	N	Y	Y

Notes: This table reports the estimation results from the event study analysis. The unit of observation is a book-week. The dependent variable is log quantity. All specifications include book and week level fixed effects, as well as week since publication fixed effects. Standard errors are clustered at the book level and presented in parentheses below the coefficients. *p<0.1, **p<0.05, ***p<0.01.

Table C4. Results of Selected Sample: Sales

VARIABLES	(1) log_quant	(2) log_quant	(3) log_quant	(4) log_quant	(5) log_quant	(6) log_quant
post_event	0.114***	0.00382	0.0913***	0.0907***	-0.182**	-0.432***
	(0.0267)	(0.0222)	(0.0260)	(0.0269)	(0.0878)	(0.0883)
Constant	2.786***	2.861***	2.802***	2.800***	4.763***	4.943***
	(0.0181)	(0.0151)	(0.0177)	(0.0183)	(0.0628)	(0.0632)
Observations	75,026	75,026	75,026	74,375	75,026	75,026
R-squared	0.770	0.823	0.776	0.790	0.782	0.858
Book Time Trend	N	Y	N	N	N	Y
Cate Time Trend	N	N	Y	Y	N	N
Cate Week FE	N	N	N	Y	N	N
Pre Sale Weight	N	N	N	N	Y	Y

Notes: This table reports the estimation results from the event study analysis. The unit of observation is a book-week. The dependent variable is log quantity. All specifications include book and week level fixed effects, as well as week since publication fixed effects. Standard errors are clustered at the book level and presented in parentheses below the coefficients. *p<0.1, **p<0.05, ***p<0.01.