Higher Prices for Larger Quantities? Non-Monotonic Price-Quantity Relations in B2B Markets

Wei Zhang

Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China; wzhang15@hku.hk

Sriram Dasu

Marshall School of Business, University of Southern California, Los Angeles, CA 90089, USA; dasu@marshall.usc.edu

Reza Ahmadi

Anderson School of Management, University of California, Los Angeles, 110 Westwood Plaza, Los Angeles, CA 90095, USA; reza.ahmadi@anderson.ucla.edu

Abstract

We study a microprocessor company selling short-life-cycle products to a set of buyers that includes large consumer electronic goods manufacturers. The seller has a limited capacity for each product and negotiates with each buyer for the price. Our analysis of their sales data reveals that larger purchases do not always result in bigger discounts. While existing theories cannot explain this non-monotonic pattern, we develop an analytical model and show that the non-monotonicity is rooted in how sellers value capacity when negotiating with a buyer. Large buyers accelerate the selling process and small buyers are helpful in consuming the residual capacity. However, satisfying mid-sized buyers may be costly because supplying these buyers can make it difficult to utilize the remaining capacity, which is too much for small buyers but not enough for large buyers. We briefly discuss the implications for capacity rationing and posted pricing and potential applications to other industries.

[*Keywords*: data-driven; revenue management; pattern analysis; bargaining; semiconductor]

1 Introduction

Price and quantity are the most common business concepts. While conventional wisdom suggests that larger purchase quantities are generally associated with lower prices (e.g., Spence 1977, Oren et al. 1982, Jeuland and Shugan 1983, and Weng 1995), our empirical observations raise doubts. In this research, we interacted with managers of a large semiconductor company and obtained a sales data set that spans a three-year period. As we examine the data, we observe some compelling instances wherein larger-quantity buyers pay higher prices. Five such examples are illustrated in Table 1. For each of the five products, we sort buyers about equally into three groups—small, medium, and large—according to their total purchase quantities. We then calculate the average price received by each group by summing the total purchase value and dividing it by the total quantity. Interestingly, the average price received by the medium-quantity group is less than those obtained by the other two groups. Further, the large-quantity group received the highest average price in four of the five examples. In fact, many other "anomalies" are observed beyond the five examples. In particular, if we rank buyers of a product according to their total purchase quantities, we find that in about 26% of the cases, a buyer pays a higher average price than a neighboring, smaller-quantity buyer does. These pricing "anomalies" may lead to overall non-monotonic price patterns for some products. In order to check whether a non-monotonic price-quantity relation generally exists, we conduct more rigorous analyses in this paper, both empirically and analytically.

In the first step, we use a set of linear and nonlinear regressions to control for other possible influences on price. Although total payment increases with total quantity in almost all cases, the

Table 1: Quantity-Weighted Average Price for Three Customer Segments

Product	Category	Number of Customers	Lifespan (year)	Small Amt. Avg. Price	Medium Amt. Avg. Price	Large Amt. Avg. Price
1	Desktop CPU	40	2.77	\$ 58.38	\$ 55.32	\$ 58.78
2	Desktop CPU	5	0.98	\$ 29.13	\$ 27.45	\$ 45.01
3	Desktop CPU	6	0.90	\$ 27.54	\$ 25.33	\$ 28.46
4	Desktop CPU	11	0.86	\$ 92.50	\$ 91.06	\$ 93.43
5	Memory	9	0.87	\$ 2.59	\$ 2.47	\$ 2.51

discount received by a buyer is statistically a non-monotonic function of the buyer's demand share (or relative size) for a product. Specifically, the discount increases with demand share for small quantities. However, as demand share increases, the discount decreases and then increases again. In brief, we observe an N-shaped discount curve, which cannot be explained by the existing literature.

To gain a deeper understanding of the observed phenomenon as well as to provide a theoretical justification, we then develope an analytical model that is largely based on the practices of the company we are studying. We use the model to investigate the price-quantity relation in a business-to-business (B2B) market where the product life cycle is short, capacity is inflexible, and prices are set through *one-shot* negotiations. Our model suggests that the non-monotonic price-quantity relation is rooted in how the seller values the capacity when negotiating with a buyer. Large buyers accelerate the selling process and small buyers are helpful in consuming the residual capacity. However, satisfying mid-sized buyers may be costly because supplying these buyers can make it difficult to utilize the remaining capacity, which is too much for small buyers but not enough for large buyers. Technically, a value function for the remaining capacity that is first convex and then concave (or S-shaped) is sufficient to lead to a non-monotonic price curve. Such a value function can arise quite naturally in practice because demand is expected to be finite and thus capacity is valuable neither too higher nor too lower than the expected demand. Finally, we show that our model fits the data better and it can yield the price-quantity curves found in the data.

Knowledge gleaned on the price-quantity relation will be useful in B2B markets that have similar economics to the semiconductor industry, where capacity is finite and inflexible and prices are negotiated. Due to the impact of one transaction on subsequent transactions, it is important for the seller to control the capacity allocated to each buyer, if possible, prior to price negotiations. To optimize the trade-off between the profit from the current buyer and that from future buyers, a good understanding of the price-quantity relation is necessary. This knowledge will also help the seller optimize the posted price, which balances the profit between buyers who choose to take the price and those who choose to bargain.

The rest of this paper is organized in the following way. We present a brief literature review in Section 2. In Section 3, we introduce the industry and firm practices. In Section 4, we show our empirical observation through linear and nonlinear regressions. We then build a model in Section 5 and analyze the problem in Section 6. We discuss the managerial implications of our finding in

Section 7 and conclude in Section 8. All the proofs are in the electronic companion.

2 Related Literature

Our work is related to four areas of research. The first area is about the quantity-discount pricing policy. In the operations manangement and marketing literature, quantity-discount pricing policy has been widely studied as a tool for price discrimination (e.g., Spence 1977 and Oren et al. 1982) and channel coordination (e.g., Jeuland and Shugan 1983 and Weng 1995). In these papers, it is assumed that one party will offer the contract in a take-it-or-leave-it fashion and buyers with greater demand receive lower prices. Assuming that one party has full bargaining power simplifies the analysis, but it also ignores prevailing practices in which buyers negotiate. Our study assumes that the seller does not have the power to dictate the price for every buyer.

The second related area is B2B price bargaining. In the marketing literature, Kahli and Park (1989) analyzed a two-party bargaining problem wherein the inventory policies of the seller and the buyer are described by a simple economic order quantity (EOQ) model and they showed that the optimal discount increases with purchase quantity. In the economics literature, Snyder (1998) and Chipty and Snyder (1999) discussed the impact of buyer demand size on price discount. Snyder (1998) showed that when many suppliers compete to sell to one buyer at a time in a repeated game, the price offered to the seller in equilibrium initially increases with buyer size and then decreases with buyer size. However, the result requires that suppliers cooperate and buyers appear sequentially over an infinite horizon, which are both very strong assumptions in supply chains. More importantly, our data exhibits a more complicated price pattern that is not explained by their model. In a very different setting from ours, Chipty and Snyder (1999) showed that a merger enhances (worsens) buyers' bargaining position if the supplier's payoff function is concave (convex) in total transaction size.

Other studies on B2B bargaining have assumed that the size of the pie is given and explored how the pie is allocated among channel members. While Dukes et al. (2006) and Lovejoy (2010) focused on the impact of channel structure, Nagarajan and Bassok (2008) considered suppliers in an assembly chain who form multilateral bargaining coalitions and compete for a position in the bargaining sequence. We supplement this branch of literature by considering a seller that sequentially negotiates with a group of buyers, investigating the impact of a buyer's relative quantity size both empirically and analytically, and discussing the plausibility and possible implications of the non-monotonic discount curve.

Our research is also related to revenue management. Kuo et al. (2011) is the first paper to study revenue management for limited inventories when buyers negotiate. They considered a dynamic setting with fixed compositions of price-takers and bargainers and assumed that each buyer only buys one unit of the product and that the posted price is updated frequently. The authors characterized the optimal posted price and the resulting negotiation outcome as a function of inventory and time. They also showed that negotiation is an effective tool to achieve price discrimination. In contrast, our paper considers a dynamic, capacity-rationing problem in a B2B market in which buyers request different quantities and quantities influence prices. Our work is also related to research on dynamic and stochastic knapsack problems that study the optimal admission or pricing policies with limited capacity. Talluri and van Ryzin (2004) categorize this type of problems as revenue management with group arrivals but they offer no solutions. While early studies such as Gallego and van Ryzin (1994) and Kleywegt and Papastavrou (1998) showed that the optimal expected revenue is concave in capacity if all demands require the same amount of resources, Kleywegt and Papastavrou (2001) showed that concavity does not hold in general when demands are heterogeneous, which provides support for our analysis. However, Kleywegt and Papastavrou (2001) focused on characterizing the conditions under which concavity holds, and we focus on characterizing the property of the value function under which the price-quantity relation is non-monotonic.

Lastly, our research is related to capacity management in the semiconductor industry. Wu et al. (2005) provide a good review of the literature on capacity planning in the high-tech industry. In the specific setting of semiconductor industry, Karabuk and Wu (2003) studied how to cope with both demand and capacity uncertainties and coordinate marketing and manufacturing decisions in strategic capacity planning. Cohen et al. (2003) proposed a model to estimate the imputed costs of a equipment supplier in deciding the timing of production. Terwiesch et al. (2005) empirically studied the demand forecast sharing process between a buyer of customized equipment and a set of equipment suppliers. Karabuk and Wu (2005) studies the incentive issues when product managers compete for capacity allocations. Peng et al. (2012) worked with Intel and developed an equipment procurement framework that allows Intel to make a combination of base and flexible capacity reservations with suppliers. Different from all these studies, this paper helps semiconductor companies understand the value of their capacities when customers arrive sequentially and randomly.

To summarize, our paper makes the following contributions. First, we provide an empirical analysis that reveals the existence of a non-monotonic price-quantity relation in the semiconductor industry. Second, we develop a model to investigate this phenomenon and find a plausible explanation. Third, we show a simple and sufficient condition for the price-quantity relation to be non-monotonic.

3 Industry and Firm Practices

Market Structure. The microprocessor market is intensely competitive, with rapid technological advancements, short product life cycles, and regular pricing activities. Many competing sellers such as Intel, Nvidia, and Advanced Micro Devices (AMD), sell multiple product lines primarily to original equipment manufacturers (OEMs), such as Hewlett-Packard (HP), Lenovo, and Dell.

Capacity Inflexibility. For sellers to remain competitive, they need production capacity with up-to-date process technology, which requires heavy capital investments. Some sellers like Intel (which is known as an integrated device manufacturer) manufacture products in house, while others like AMD (known as a fabless company) only focus on product design and outsource production to third-party foundries. In both cases, because the manufacturing facilities are costly and construction lead times are long, capacities are inflexible during a selling season. Although capacity configuration at a manufacturing facility can be altered, doing so disrupts flows in the manufacturing facility and causes increased manufacturing cycle times (Karabuk and Wu 2003). Another important factor is that sellers allocate the capacity to product lines based on demand forecasts and start production several month before demand realizes. The forecasts represent sales commitments from product line managers and sometimes (non-binding) purchase commitments from customers. Once the capacity is allocated accordingly to a product, the production starts almost immediately. By doing so, sellers can first increase the capacity utilization. Second, due to the long production lead time, which is on average six to twelve weeks, sellers can build up inventory in advance in order to satisfy customers that normally require immediate delivery. Last but not least, semiconductor manufacturing entails significant learning and it takes time for yields to ramp up and for quality to improve. Thus, when facing a supply shortage, it is not only costly but also risky to seek an alternative source. Given these facts, it is important for sellers' product managers to provide accurate forecasts and to sell according to allocated capacities.

Price Negotiation. Although each product has a posted price, the final price for each buyer is usually set through negotiations. Major buyers are sophisticated, drive hard bargains, and often enjoy higher annual revenues than sellers (Cooper 2008). Buyers know that the marginal production cost of microprocessors is low and sellers are eager to discount prices to fully utilize their capacities. Moreover, buyers can allocate their business among competing sellers. Lacking full pricing power, sellers are unable to use pricing strategies, such as take-it-or-leave-it price schedules or a menu of contracts, and have to engage in negotiations. Once a price is settled, the duration of contract can vary for different buyers and products; price renegotiations happen frequently but not in all cases.[1] In our data set, approximately 45% of the purchases did not involve any renegotiation.

Procurement Quantity. The purchase quantity, however, is normally not a term for negotiation. To produce their products, buyers need other inputs from different suppliers, so it is costly to manipulate purchase quantities once production plans have been made. In principle, buyers can allocate their requirements among alternative semiconductor firms, but normally at a very early stage before production plans are finalized.[2] This is because products offered by different sellers differ in technical features and it normally requires a buyer to design the final product in a specific way in order to use the component. In addition, sellers' brand images in the consumer market may matter, so substitutions cannot be easily made. As a result, buyers' procurement managers prefer to stick to their internal production plans and procure the desired quantity at the best possible price. In summary, buyers determine their purchase quantities based on their production plans prior to negotiating with suppliers, and incur costs if they are unable to procure these amounts and must switch to an alternative seller. Their contracts with sellers typically do not include any commitment or requirement for minimum product purchases.

[1] Companies may use different types of contracts in terms of price flexibility, which is manifested in contract specifications about how frequently and to what extent the price can be renegotiated. Hence, price renegotiation happens when a contracted price expires.

[2] For example, Apple Inc. allocated about one-third of the A9 processor orders to Taiwan Semiconductor Manufacturing Company in April 2015, but the production of 2015 iPhones and iPads that use the A9 processor was not started until August. Sources: http://appleinsider.com/articles/15/04/15/apple-makes-last-minute-decision-to-use-tsmc-for-30-of-a9-chip-orders-for-next-iphone/; http://www.macrumors.com/2015/08/07/iphone-6s-production-late-august/.

Technology Upgrades. Another aspect of this industry with a major impact on negotiations is the risk of obsolescence. Sellers are aware that technological advancements from rivals can cause a rapid decline in demand for existing products. Although they are aware of development cycles in the industry and can anticipate when rivals will introduce products, they must constantly consider the likelihood of a demand shock and the possibility of having to salvage inventories (Karabuk and Wu 2003).

4 Empirical Observation and Analysis

The data provided by a major global semiconductor company for use in this study encompasses 3,826 products and 251 buyers over a three-year period. Each record in the data set consists of customer ID, product ID, product category, product brand (subcategory), sales territory, bill quantity, bill value in USD, unit price, and date of transaction. The products sold include central processing units (CPUs), graphics processing units (GPUs), and embedded chips, among others.

4.1 Data Preparation

4.1.1 Fixed-Price Contracts

As a buyer normally purchases a product through multiple transactions over time, the price may be renegotiated. In this paper, we focus on purchases in which prices are fixed over the entire product life cycle. We say that such transactions are made under *fixed-price* contracts. The analysis for *fixed-price* contracts or *one-shot* price bargaining is simpler than for repeated negotiations.

Let I and J be the indices of buyer and product, respectively. Let $\mathcal{T}_{ij}$ be the set of dates at which buyer i purchased product j. Let q_{ijt} and p_{ijt} denote the transaction quantity and price for customer i and product j at time $t \in \mathcal{T}_{ij}$. We define an *instance* θ_{ij} as the set of transactions related to buyer $i \in I$ and product $j \in J$; i.e., $\theta_{ij} := \{(t, q_{ijt}, p_{ijt}) : t \in \mathcal{T}_{ij}\}$. As stated earlier, in this industry it is a common practice for a buyer to enter a price negotiation with a fixed target quantity. For fixed-price contracts, we have $p_{ijt} = p_{ij}$ for all $t \in \mathcal{T}_{ij}$, and the life-cycle purchase quantity is the target for the price bargaining. Hence, we focus on the relationship between the total purchase quantity, $TQ_{ij} = \sum_{t \in \mathcal{T}_{ij}} q_{ijt}$, and the fixed price p_{ij} for such instances. Of course, transaction-level data may contain information about factors that impact negotiations, such as when a purchase starts and how long it lasts. We therefore use the transaction-level data to construct measures for these factors.

4.1.2 Normalization

Widely varying prices and market sizes for different products compel us to normalize the data to the same scale. Prices of the 425 brands (or product subcategories) range from several dollars to more than \$100 per unit. Thus, an observed price that is the lowest for one product may be higher than any observed price for another product. For that reason, in place of the price and total quantity, we use two ratio metrics: (1) effective discount (*ED*) and (2) demand share (*DS*). Define

$$ED_{ij} \quad := \quad 1 - p_{ij} / \max_{i' \in I, t \in \mathcal{T}} p_{i'jt}; \tag{1}$$

$$DS_{ij} \quad := \quad TQ_{ij} / \sum_{i' \in I} TQ_{i'j}, \tag{2}$$

where $\mathcal{T} := \cup \mathcal{T}_{ij}$. Both variables have the range of $[0, 1]$. *ED* is a measure of price level relative to the highest price ever paid for the product. Note that the posted prices of the seller who offers this data set are very stable, usually lasting for more than a year, so for a fixed-price contract the nominal discount rate is almost constant over time and *ED* is very close or equal to the nominal discount rate. *DS* is a measure of total quantity relative to the market size of the product. The advantage of *ED* and *DS* is that they simultaneously control the mean and the variation across different products.[3] However, the demand shares in a large amount of instances are concentrated around zero (as shown in Figure 1 on the left) due to the 20-80 rule: 80% of customers contribute only 20% of sales. Therefore, it is difficult to examine how quantity affects price discount in the majority of instances if we use *DS*. To avoid such a shortcoming, we use a power transformed of demand share ($PTDS$) which is defined as $PTDS = DS^\gamma$ for a $\gamma \in (0, 1)$. $PTDS$ has serveral advantages. First, it is a monotonic transformation of *DS*, so the price-quantity relationship is preserved. Second, its empirical distribution is more spread out (as shown in Figure 1 on the right). Third, the range $[0, 1]$ is preserved and maintaining the unit range makes it convenient to interpret the results. As shown in Table 2, the distribution has the highest spread when γ is

[3] Note that how the demand share of a customer evolves over time is irrelevant for our analysis. We only use the *ex-post* demand share as a normalized quantity, which is constant over time.

Figure 1: Histograms for the Selected Subset of Fixed-Price Contracts

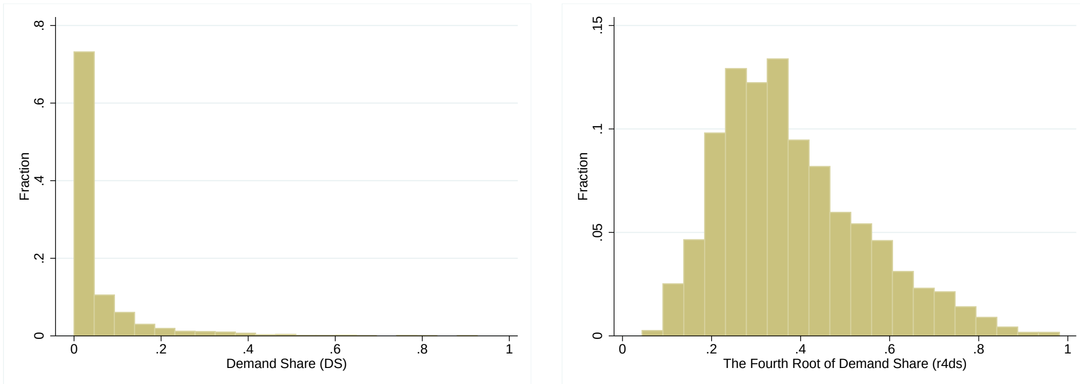


Table 2: Measuring the Distributions of Power-Transformed Demand Share

	$DS^{0.15}$	$DS^{0.20}$	$DS^{0.25}$	$DS^{0.30}$	$DS^{0.35}$	$DS^{0.40}$	$DS^{0.45}$
St.Dev.	0.1409	0.1556	**0.1624**	**0.1642**	**0.1626**	0.1589	0.1539
W	0.9873	0.9743	0.9572	0.9365	0.9131	0.8875	0.8604

Note. The Shapiro-Wilk W statistic measures the straightness of the normal probability plot of a variable; larger values of W indicate better normality.

between 0.25 and 0.35, and $\gamma = 0.25$ yields a distribution that is the closest to normal in this range.[4] In this paper, we focus on the fourth-root transformation (i.e., $\gamma = 0.25$). Later, to check the robustness, we will also show the results for γ =0.15, 0.2, 0.3, and 0.35. Thus, our objective reduces to identifying the relation between ED_{ij} and $PTDS_{ij}$ while controlling for other factors.

4.1.3 Data Filtering

The data set corresponds to a three-year time period from January 1, 2009 to March 25, 2012. Instances that started prior to January 1, 2009, and those that lasted beyond March 25, 2012, have missing data (or are truncated). Because prices in the microprocessor market are decreasing over time, instances that started early with left-truncation will appear to have smaller total quantities and higher prices than subsequent instances. Similarly, instances that started late with right-truncation will appear to have smaller total quantities and lower prices than preceding instances. Mixing these two effects may generate a non-monotonic relation between price and quantity. To mitigate this truncation effect, we focus on instances with an observed starting date at least one

[4]Note that our primary goal is to spread out the distribution as much as possible.

quarter later than January 1, 2009, and an observed ending date at least one quarter earlier than March 25, 2012. There are 6,573 instances (about 53%) that satisfy such criteria. Furthermore, to focus on regular purchases but not transactions for one-time substitutions or sample testing, which follow different selling processes, we drop another 312 instances (about 4.7%) that have only one purchase record. Finally, products that have an extremely small number of buyers are often customized and may follow a different selling process. In addition, such products tend to have extreme-demand-share buyers as well as narrow price dispersions. Hence, our proxies for the posted price of such products could be downward-biased. To avoid this bias or lack of good proxy, we drop another 1,551 instances (about 18%) and consider products that have more than three buyers. In this way, we obtain a subset of data with 2,346 fixed-price instances and 2,364 price-renegotiated instances. In the electronic companion, we use the Heckman selection model to correct for the selection bias and show that our data filtering does not change the price-quantity pattern.

4.2 Variables

Aside from the demand share, other variables may also influence a customer's discount. According to the generalized Nash bargaining model (Nash 1950; Roth and Malouf 1979), these variables fall into three broad categories: the seller's outside options, the buyer's outside options, and their respective bargaining powers. As far as we can imagine, the seller's outside options are affected by production cost, salvage value, buyer-side competition, time of purchase, capacity or inventory level, and demand uncertainty. The buyer's outside options are affected by the value of adopting a different product, seller-side competition, posted price, and time of purchase. Bargaining powers are affected by the value of the business relationship, the bargaining skills of salespersons and procurement managers, and the buyer's reputation for committing to a forecast. In the following, we explain the variables included in our regression. Table 3 provides the summary statistics for the portion of data we use, and Table 4 shows the correlation among the variables.

Power transformation of Demand Share. We first focus on the relationship between *ED* and the fourth root of demand share (*r4ds*). Later, we will consider other power transformations of demand share (e.g., $DS^{0.15}$, $DS^{0.2}$, $DS^{0.3}$, and $DS^{0.35}$) to check the robustness.

Cbase. This variable counts *the total number of buyers* for a product, and is thus a measure of a product's popularity and the buyer-side competition.

Table 3: Summary Statistics for Fixed-Price Instances (N = 2,346)

Variables	Mean	S.D.	Min	Max	Variables	Mean	S.D.	Min	Max
ED	.1126	.1781	0	.9986	*lndod*	3.9237	1.8581	0	6.9527
DS	.0516	.0942	3.8e-6	.9263	*lndrt*	4.5246	1.3531	0	6.8977
$DS^{0.25}$	.3818	.1624	.0440	.9811	*M3*	.2928	.4552	0	1
Cbase	19.66	13.43	4	48	*CapL*	.4614	.3052	.1761	1.176
TSQ	8.37e5	2.35e6	826	3.06e7	*Cshr*	.0516	.1038	2.24e-6	.7897
Herf	.2593	.1787	.0561	.9815	*Vrate*	.3112	.2506	0	.9375

TSQ. The total sales quantity of a product, which also implies the popularity of the product.

Herf. This is the *Herfindahl Index* for the demand structure of a product, which measures the degree of demand concentration. It is calculated as the square root of the sum of the square of demand shares across all the buyers (Weinstock 1982).

lndod. The discount received by a buyer for a product is related to the time when the buyer starts to purchase, because effective prices (or price-performance ratio) in the semiconductor industry are decreasing over time in general. The later a buyer arrives for a product, the greater discount (relative to the highest price) he may obtain due to better outside options. Hence, to capture such a time effect, we use the logarithm of *days of delay*, which is calculated as the difference in number of days between the starting date of an instance and the first date that the product was ever purchased. Note that, as relative measures, *lndod* and *ED* are compatible.[5] In addition, *lndod* is also a measure of demand uncertainty, because uncertainty is resolved over time.

lndrt. The discount may be also related to the rate of purchase given the same total quantity, so we also control the logarithm of *duration* of an instance. It is calculated as the number of days between the first date and the last date of an instance. We can see that the life span of an instance in the data set is fairly short, with an average duration of 232 days.

M3. Another dimension of the time of purchase is related to the seller's financial cycle. It is well known that the *end-of-quarter effect* may bring buyers an edge in the bargaining. We introduce *M3* as a binary variable with a value equal to 1 if the date of the price negotiation is in the third month of a quarter and 0 otherwise.

[5] Another candidate may be the days of delay relative to the introduction date of a product. Compare two scenarios. In scenario I, the first customer delays her purchase 0 days and receives the highest price among all the customers. In scenario II, the first customer delays her purchase 100 days and also receives the highest price. Other customers purchase one day after and all receive the same effective discount. Hence, the absolute delay is irrelevant.

CapL. The remaining capacity level at the time of price negotiation is a consideration for both the seller and the buyer. However, we do not have information about the capacity level. We approximate the total capacity level using the total sales of a product divided by the semiconductor industry capacity-utilization rate (about 85%),[6] and the available capacity level for a buyer using the difference between the total capacity level and the cumulative contracted sales prior to this buyer. We then use $CapL$ =(available capacity level)/(total sales) as a control variable.

Cshr. A buyer that contributes to a large portion of the seller's overall sales can have significant bargaining power. To capture the bargaining power of a buyer in this regard, we calculate *Cshr* (i.e., customer share) as 100 times the total quantity purchased by a buyer over the observed period divided by the seller's total sales volume across all products and the observed period. This measures the value of a buyer to the seller.[7] Note that using the total purchase value as a control variable may cause an issue of endogeneity as the value depends on the price discount.

Vrate. Since we use the highest price paid for a product as the approximation of the posted price, the larger the price variation of a product, the greater the computed *ED*. Hence, we should control the price variation of a product. However, prices depend on discounts received by buyers. To avoid the endogeneity issue, we use the *fraction of price-renegotiated instances* to measure a product's price variation, given that buyers with renegotiable-price contracts are more likely to get the posted price at the beginning and price variations are larger than with fixed-price contracts. In addition, *Vrate* measures the uncertainty of the product's value, because price renegotiations normally happen when uncertainties are resolved. For fixed-price contracts, discounts are likely to be larger when uncertainties are higher.

Product-line (or brand) fixed effect. To capture product-line-specific impacts such as the seller-side competition, production cost and salvage value, we use binary variables for the major 16 brands that have at least 100 observed instances in the original data set. It is important to note that the seller's salespeople are organized by product lines. Hence, negotiated prices of a product are all subject to the same impact from the bargaining ability of a salesperson or a group of salespeople.

[6] http://www.semiconductors.org/industry_statistics/industry_statistics/, accessed March 2015. We obtain very similar results if we use random utilization rates (e.g., a normally distributed random variable with mean 0.85 and standard deviation 0.1, capped by 0 and 1).

[7] Note that *Cshr* is a *ex-post* measure and it may not accurately capture a buyer's bargaining power at the beginning of the observed period. However, as we can see later, *Cshr* is highly collinear with the buyer fixed effects, indicating that the bargaining power dynamics do not change drastically.

Table 4: Correlation Matrix for the Selected Fixed-Price Instances

	ED	*r4ds*	*Cbase*	*TSQ*	*Herf*	*lndod*	*lndrt*	*M3*	*CapL*	*Cshr*	*Vrate*
ED	1.00										
r4ds	0.06	1.00									
Cbase	-0.22	-0.42	1.00								
TSQ	0.02	-0.15	0.10	1.00							
Herf	0.15	0.07	-0.63	-0.09	1.00						
lndod	0.30	-0.13	-0.22	0.07	0.24	1.00					
lndrt	-0.16	0.14	-0.01	0.06	0.00	-0.11	1.00				
M3	0.03	0.08	-0.13	-0.01	0.05	0.02	-0.09	1.00			
CapL	-0.16	0.38	-0.07	-0.07	-0.11	-0.65	0.12	-0.00	1.00		
Cshr	0.22	0.29	-0.32	-0.01	0.20	-0.02	0.10	0.05	0.13	1.00	
Vrate	0.36	-0.18	0.05	-0.02	-0.04	0.24	-0.13	0.01	-0.21	0.05	1.00

In other words, the impact of salesperson ability is product-line-specific and thus can be captured by the product-line fixed effect.

Buyer fixed effect. Apart from purchasing value, a buyer's bargaining power is also affected by unobservable factors, such as the experience of the procurement manager and the reputation for honoring a commitment. Hence, we use binary variables to control the buyer fixed effect for the 10 major buyers in terms of total purchase value with the seller and use others as the reference.

Quarter fixed effect. To capture the industry dynamics that are cyclical within a financial year, we use the first quarter as the reference and binary variables for the other three quarters.

Location fixed effect. The degree of market competition on both the buyer and seller sides may depend on the location. We use binary variables for nine of the ten recorded sales territories, such as greater China and North America. Additionally, location may also be an indicator of cost level and demand uncertainty.

Interaction effect. The impact of capacity level may interact with time elapsed. The likelihood of a technology shock occurring increases over time after a product is introduced; once a shock occurs, the seller may have to salvage the remaining capacity. Hence, the capacity has less and less value as time elapses and we thus include the interaction between *CapL* and *lndod*.

Though we try to control for as many variables as possible, we still confront the "omitted variable" problem due to a lack of information. Thus, consistent estimators can be obtained only when the omitted variables are uncorrelated with our regressors. The factors we do not control for here are the net cost of switching to an alternative product and contract terms other than price and quantity for a buyer. We show in the electronic companion that under certain mild assumptions, the estimated coefficients are just the scaled true marginal effects when these two factors are relevant but missing. Hence, the shape of the price-quantity relationship will be preserved. Last, note that we consider only the instances with one-shot bargaining (or fixed-price contract) for new products, and thus it is reasonable not to consider any reference effect from previous prices or discounts.

4.3 Regression Analysis

In this section, we try to identify the empirical relationship between the effective discount and demand share in two steps. The first step is to explore the underlying pattern by segmenting the demand share and computing the average effective discount received by buyers in each segment. Based on the observed pattern, if one exists, we obtain a reasonably well-fitted functional form through piecewise polynomial regression in the second step. Our analyses in the two steps are both necessary and complementary. The first step provides us with information about the shape of the function, the possible location(s) of the knot(s), and the order of polynomial functions we need. The second step allows us to test the statistical significance of the functional form.

4.3.1 Average Discounts by Segments

For robustness, we consider two different ways of segmentation, the details of which are given in Table 5. In model (i), we divide the instances into six segments according to *r4ds*. We use wider ranges for the first and last segments in order to include more instances in the "tails." In model (ii), we use nine segments. Incorporating the aforementioned variables, we run the linear regression $ED = \sum_{k>1} a_k \cdot seg_k + b'X + \epsilon$ for each model, where seg_k is a binary indicator for segment k, X is the vector of controlled covariates (including the constant), and ϵ is the error term. The estimation results for a are summarized in Table 5 and b in Table 7.[8]

We can see that in both models the marginal impact of demand share on discount displays a similar non-monotonic pattern. In model (i), the estimated coefficient increases with demand share for the first four segments, then decreases in segment 5, and increases again in segment 6. In model

[8] We use the *regress* command with the *robust* option in STATA. With this option, STATA estimates the standard errors using the Huber-White sandwich estimators. The robust standard errors can effectively deal with minor problems regarding normality, heteroscedasticity, and some observations that exhibit large residuals. The point estimates of the coefficients are exactly the same as those in ordinary OLS.

Table 5: Summary Statistics and Regression Results for Demand Share Segments

Model	Segmt.	Range of $r4ds$	Summary Stats. of $r4ds$			Regression		
			Mean	S.D.	Obs.	Coef.	Robust S.E.	P value
	1	0˜0.2	.1581	.0313	235	-	-	-
	2	0.2˜0.35	.2772	.0423	916	.0141	.0097	0.149
	3	0.35˜0.5	.4142	.0433	683	.0245	.0111	0.028
(i)	4	0.5˜0.65	.5669	.0406	335	.0428	.0153	0.005
	5	0.65˜0.8	.7187	.0439	145	.0348	.0178	0.052
	6	0.8˜1	.8625	.0459	32	.0933	.0312	0.003
	1	0˜0.15	.1226	.0199	85	-	-	-
	2	0.15˜0.25	.2087	.0267	423	.0023	.0156	0.885
	3	0.25˜0.35	.2993	.0289	643	.0031	.0159	0.843
	4	0.35˜0.45	.3949	.0293	518	.0171	.0168	0.308
(ii)	5	0.45˜0.55	.4965	.0286	291	.0145	.0183	0.429
	6	0.55˜0.65	.5922	.0285	209	.0409	.0206	0.048
	7	0.65˜0.75	.6977	.0303	106	.0194	.0230	0.398
	8	0.75˜0.85	.7905	.0269	55	.0425	.0257	0.098
	9	0.85˜1	.8987	.0374	16	.1223	.0476	0.010

(ii), the coefficient increases until segment 6, then decreases in segment 7, and increases again. These results indicate that the discount is likely to be an N-shaped function of demand share. In Figure 2, we plot the average discount received by each segment and the smooth line connecting them. In both graphs, all the other controlled variables take their mean values.

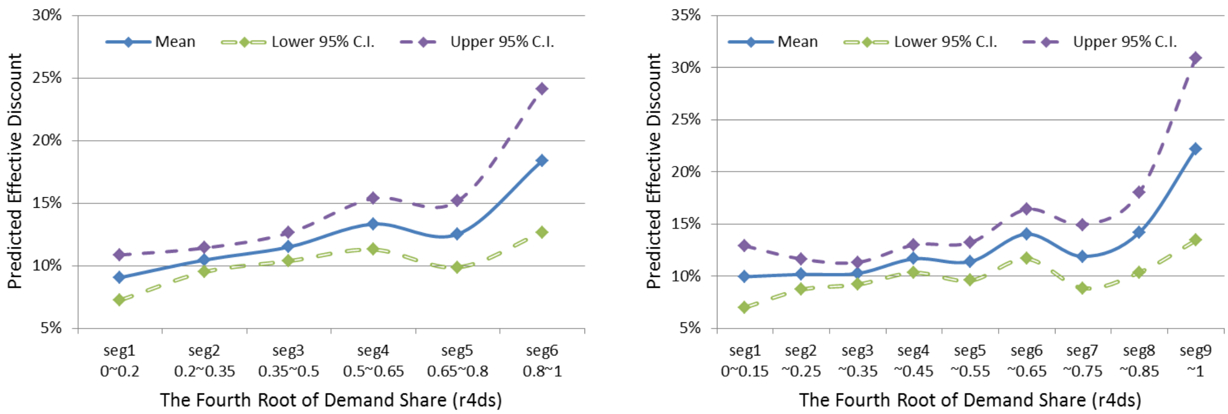
Although the smooth lines look like an "N" in both graphs of Figure 2, it is still difficult to tell how significantly the underlying shape is an N based solely on the results we have obtained so far. Rather, what we can learn is that it may be inappropriate to use a simple monotone function or a polynomial function to describe the shape given the irregular pattern; that is, the shape is more likely to be a combination of a monotonically increasing curve and a V-shaped curve, which are easier to fit by polynomial functions separately. In the next step, we propose a piecewise polynomial function to fit the data and test the significance of the non-monotonicity.

4.3.2 Piecewise Polynomial Regression

To reduce the number of parameters while maintaining adequate flexibility, we use a two-segment, piecewise-quadratic function with an unknown knot to fit the data in model (iii).[9] Hence, we let

[9]A polynomial of degree three can also generate an N-shaped curve. However, a degree-three polynomial is concave on the left of "N," meaning that discount increases rapidly with demand share for very small buyers. According to our empirical observation, the discount curve should be linear or convex first, so it is difficult to have a good fit for

Figure 2: Average Marginal Impact of Demand Share in Model (i) and (ii)



Note. In both graphs, all other covariates take their mean values.

the data decide whether the function is linear or quadratic in each segment and where the two smooth lines are connected. We let B denote the location of the knot to be estimated and run a least-square regression with the following nonlinear model:

$$ED = a_1 \cdot (r4ds - B)_- + a_2 \cdot (r4ds - B)_-^2 + a_3 \cdot (r4ds - B)_+ + a_4 \cdot (r4ds - B)_+^2 + b'X + \epsilon, \quad (3)$$

where $x_- = \min\{x, 0\}$, $x_+ = \max\{x, 0\}$, X is the vector of controlled covariates (including the constant), and ϵ is the error term. We run this nonlinear-least-square (NLS) regression in Stata with the command *nl* and we notice that the estimates are sensitive to initial values for the iteration performed by *nl*. To minimize the reliance on the initial guesses, we first run the NLS regression to estimate B and then run an ordinary-least-square (OLS) regression using the estimated B to obtain other parameters. We report the estimates for a_1 to a_4 and B in Table 6, and b in Table 7.

We can see from Table 6 that a_1 is significant (at the 5% level) but a_2 is not, meaning that a linear relationship is significant in the first (left) segment. In the second (right) segment, a_3 is significant at the 10% level (for the two-sided test) and a_4 is significant at the 5% level, meaning that a quadratic relationship is significant. Because $B = 0.5668$ is highly significant (at the 0.1% level), the relationship thus cannot be described by a single linear or quadratic function. If we assume in

a degree-three polynomial. Although polynomials of high degrees can approximate any shape of curve, they have potential problems of overfitting and multicollinearity. In contrast, piecewise polynomials of lower degrees are capable of offering adequate flexibility, while having fewer parameters.

Table 6: Results of the Piecewise Polynomial Regressions

	Model (*iii*)			Model (*iv*)		
	Estimate	Robust S.E.	P value	Estimate	Robust S.E.	P value
a_1	0.2338	0.1058	0.027	0.1030	0.0319	0.001
a_2	0.2981	0.2208	0.177	-	-	-
a_3	-0.4378	0.2306	0.058	-0.3422	0.2001	0.087
a_4	2.0449	0.8653	0.018	1.9251	0.9564	0.044
B	0.5668	0.0346	0.000	0.5794	0.0449	0.000
	Model (*v*)			Model (*vi*)		
	Estimate	Robust S.E.	P value	Estimate	Robust S.E.	P value
a_1	0.1857	0.1101	0.092	0.1355	0.0973	0.164
a_2	0.2477	0.2805	0.377	0.0861	0.1726	0.618
a_3	-0.1252	0.1759	0.477	-0.3958	0.3127	0.206
a_4	0.7575	0.5609	0.177	2.6182	1.4140	0.064
B	0.5050	-	-	0.6286	-	-

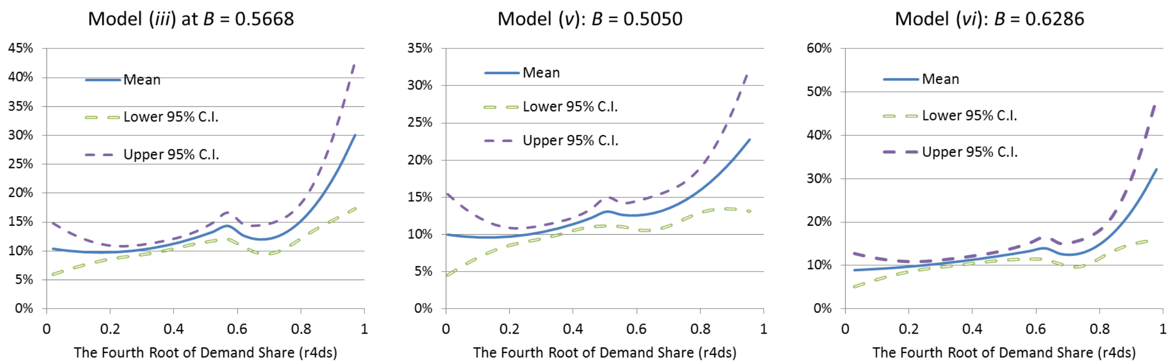
advance in model (*iv*) that the relationship is linear in the first (left) segment and quadratic in the second (right) segment, we will get similar results. Note that the minimum of the quadratic curve is achieved at $r4ds = -\frac{a_3}{2a_4} + B \approx 0.1 + B > B$, meaning that discount first decreases with demand share and then increases in this segment. Regarding *the significance of non-monotonicity*, we just need to test the null hypothsis that $a_3 \geq 0$. For this one-sided test, we have P values less than 5% for both model (*iii*) and (*iv*). Hence, we can now claim quite confidently that the empirical relationship between discount and demand share is indeed N-shaped.

To check the sensitivity of the estimated shape to the location of the knot, we run two linear regressions based on model (*iii*) with $B = 0.5668 \pm 2 \times 0.0309$. We call the two regressions model (*v*) and (*vi*), respectively, and we plot the predicted average effective discount agains $r4ds$ in Figure 3. Although setting a biased knot could smooth out the decreasing part of the curve, we still observe N-shaped curves with both model (*v*) and (*vi*).

Finally, we find that the predicted discount decreases with the number of buyers for a product, increases with a buyer's business size with the seller,[10] increases with the number of delayed days, and increases with the remaining capacity level. Additionally, impacts from time that has elapsed and capacity level influence each other in a negative way. In other words, the impact of capacity level deteriorates over time, and the impact of time delay decreases with capacity level. It is interesting to find that the effective discount is not significantly correlated with the end-of-quarter

[10] $Cshr$ and buyer fixed effect are colinear and $Cshr$ will be significant if buyer fixed effect is not included.

Figure 3: Average Marginal Impact of Demand Share: Sensitivity to Knot Location



Notes. The lines show the predicted marginal effect of *r4ds* on *ED*, with other covariates taking their mean values.

Table 7: Summary of Regression Results

Variables	(*i*) *ED*	(*ii*) *ED*	(*iii*) *ED*	(*iv*) *ED*	(*v*) *ED*	(*vi*) *ED*
f(*r4ds*)	Tab. 5	Tab. 5	Tab. 6	Tab. 6	Tab. 6	Tab. 6
Cbase	-1.13e-3***	-1.21e-3***	-1.14e-3***	-1.15e-3***	-1.11e-3***	-1.16e-3***
	(3.89e-4)	(3.88e-4)	(3.91e-4)	(3.92e-4)	(3.92e-4)	(3.92e-4)
TSQ	8.41e-10	5.90e-10	6.42e-10	9.34e-10	7.33e-10	7.87e-10
	(1.26e-9)	(1.31e-9)	(1.29e-9)	(1.27e-9)	(1.30e-9)	(1.28e-09)
Herf	-.0008	-.0059	-.0044	-.0053	.0005	-.0052
	(.0271)	(.0270)	(.0271)	(.0272)	(.0272)	(.0272)
lndod	.0382***	.0382***	.0384***	.0385***	.0382***	.0384***
	(.0047)	(.0048)	(.0047)	(.0047)	(.0047)	(.0047)
lndrt	-.0049	-.0049	-.0049	-.0051	-.0050	-.0051
	(.0039)	(.0040)	(.0039)	(.0039)	(.0040)	(.0040)
M3	-.0013	-.0007	-.0013	-.0014	-.0014	-.0014
	(.0074)	(.0074)	(.0074)	(.0074)	(.0074)	(.0074)
CapL	.0873***	.0866***	.0855***	.0865***	.0846***	.0854***
	(.0225)	(.0225)	(.0225)	(.0224)	(.0225)	(.0225)
Cshr	.2129	.2221	.1975	.2095	.2035	.2087
	(.1718)	(.1734)	(.1726)	(.1716)	(.1720)	(.1719)
Vrate	.1931***	.1919***	.1936***	.1938***	.1942***	.1932***
	(.0127)	(.0127)	(.0127)	(.0127)	(.0127)	(.0127)
*CapL**lndod*	-.0289***	-.0293***	-.0296***	-.0297***	-.0293***	-.0298***
	(.0057)	(.0058)	(.0057)	(.0057)	(.0057)	(.0057)
Constant	-.1136***	-.0994***	-.0805***	-.0902***	-.0706**	-.0588*
	(.0329)	(.0353)	(.0346)	(.0328)	(.0323)	(.0343)
F.E.s	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.3817	0.3824	0.3835	0.3830	0.3820	0.3829

Notes. Robust standard errors are in parentheses. *$p<0.1$; **$p<0.05$; ***$p<0.01$. F.E.: Fixed Effect.

effect and the demand concentration rate for the fixed-price contracts. The reasons may be that fixed-price contracts entail long-term considerations and that demand concentration is not a good measure for product popularity for this type of contract.

4.4 Robustness Checks

First, to check whether discount is influencing demand share simultaneously, we regress $r4ds$ against ED and other control variables using dummies for three quarters as instrumental variables. We find no support for simultaneity. Second, we run four additional nonlinear-piecewise-polynomial regressions using $DS^{0.15}$, $DS^{0.2}$, $DS^{0.3}$, and $DS^{0.35}$, respectively, in place of $r4ds$ in (3), and we get similar non-monotonic curves. Third, we find that omitting the buyer's net cost of switching and the value of other contract terms will not affect the non-monotonic pattern if certain mild conditions hold. Lastly, we can show that an "N" shape is still observed and significant even after correcting for sample selection biases. The details of the robustness checks are presented in the electronic companion.

4.5 Further Discussions

Our observations from the regressions are particularly interesting in that they are somewhat inconsistent with the conventional wisdom—and our intuition—which says that buyers with larger quantities should receive lower prices. While our intuition is correct for small- and large-quantity buyers, we are not aware of a discount valley for medium-sized buyers.

Why do we see such a discount curve? No existing theories in the literature can explain our observation. As we have discussed in Section 2, although existing theories can predict increasing or V-shaped discount curves, they work under very different premises and cannot be combined to generate other discount curves. Hence, we do not yet fully understand the mechanism of price negotiation in B2B markets like the semiconductor industry.

Our empirical finding has important implications for both buyers and sellers. Because larger quantities may not lead to lower prices, it may not be wise for a buyer to increase the purchase size. *However, there should not be an arbitrage opportunity for a buyer, because the total purchase cost still increases with quantity.*[11] For the seller, it may be imperative to rethink posted pricing

[11] We find that in only 7% of instances across the entire data set, a buyer pays less in total than another buyer of

and capacity rationing, given this non-monotonic relationship between price and quantity. We will continue discussion in the following sections, in which we try to use an analytical model to understand and explore the rationale behind our observation.

5 Modelling and Verification

In order to study the rationale behind the empirical observation, we build an analytical model in this section. We focus on selling a particular product and make the following key assumptions. (i) The capacity is fixed and does not expire over time. (ii) Buyers arrive sequentially and randomly. (iii) A technology shock will end the buyer arrival process. (iv) The seller can decide the capacity allocation. (v) Prices are determined by Nash bargaining. We then verify with the data that the non-monotonicity is rooted in how the seller values capacity by comparing different models in terms of goodness-of-fit as well as the statistical significance of non-monotonicity in each model.

5.1 The Model

Consider a seller $\mathcal{A}$ (she) that sells a new product to a group of OEMs. $\mathcal{A}$ has a fixed capacity κ due to a long production lead time and a short product life cycle. To be specific, we denote κ as the maximum amount of the product that can be produced during the product life cycle. In light of current industry practice, major microprocessor companies produce according to their forecasts and normally start production immediately after the capacity is set up. Thus, the capacity does not expire over time until the end of the product life cycle. The selling starts at time 0 and ends when there are no more buyers or the capacity is sold out or salvaged. There are M potential buyers who arrive stochastically for supply contracts. We consider a general *non-homogeneous Poisson arrival process* wherein the arrival rate of the i-th buyer after the arrival of the (i-1)-th buyer is $\lambda_i < \infty$ for $i = 1, 2, \cdots, M$. For simplicity, we assume that the arrival process is only determined by market characteristics and is independent of buyer identities or the history of the arrival process.

As stated earlier, in the semiconductor industry, technological advancements of competing products will lead to obsolescence of the focal product. When such a technological shock happens, potential buyers will change their adoption decisions. However, for buyers that have already adopted

the same product who buys less. This number is only 4% among fixed-price instances.

the product and integrated it into their product designs, the switching cost will be high; thus, existing buyers will continue their purchase until they phase out products that use the focal product. We assume that the arrival time of the technological shock is exponentially distributed with rate λ_0 and that seller $\mathcal{A}$ salvages the remaining capacity at marginal value s when the shock arrives. Let $\delta_i \in (0, 1)$ represents the probability of the shock arriving after the $(i-1)$-th and prior to the i-th buyer. Using the memoryless property of exponential distribution, we can check that $\delta_i = \frac{\lambda_0}{\lambda_0+\lambda_i}$.

Let D_i denote the total purchase requirement (or demand) of the i-th buyer and Q_i the capacity allocated to buyer i. We assume that buyers accept partial fulfillment as long as $Q_i \in [\eta D_i, D_i]$, where $\eta \in (0, 1]$ is plausibly an industry standard that is exogenous and identical for all the buyers. Hence, if $\eta < 1$, seller $\mathcal{A}$ faces a dynamic capacity management problem wherein she must decide the degree of fulfillment ρ_i for buyer i in order to maximize the total expected revenue. Let K_i be the total available capacity when buyer i arrives, so K_i/D_i is the maximum level of fulfillment and $\rho_i \in [\eta, 1] \cap [0, K_i/D_i]$. Note that ρ_i is not relevant if $K_i/D_i < \eta$.

Demand is unknown to the seller *ex ante* but is exogenously given because each buyer's production plan is determined in advance and the production needs inputs from different suppliers so that it is costly for buyer i to manipulate D_i. Although demand may not be exogenous from a buyer's point of view, that is not a concern of this paper. From the seller's standpoint, demand can be correlated and thus the distribution of each oncoming demand is history-dependent because buyers may be subject to the same demand shock and (or) competition in the same market. We define "history" as the set of information that is revealed to the seller. Let $\psi(t)$ denote the history up to time t, and ψ_i the history up to the arrival of buyer i. We assume that D_i follows distribution function (cdf) $F(\cdot|\psi_{i-1})$, where $\psi_0 = \emptyset$. For the purpose of analysis, we make the following technical assumptions: (1) that the expectation of the demand from a buyer is always finite, and (2) that there exists a lower envelope for the possible forms of F.

Technical Assumption 1. $\int_0^{+\infty} D dF(D|\psi) < \infty$ *for any* $\psi \in \mathcal{H}$*, where* $\mathcal{H}$ *stands for the set of all possible histories.*

Technical Assumption 2. *There exists an increasing and continuous function* $F_0(\cdot)$ *such that (i)* $F_0(0) = 0$*, (ii)* $F_0(+\infty) = 1$*, and (iii)* $F_0(x) \leq F(x|\psi)$ *for any* $x \in [0, +\infty)$ *and* $\psi \in \mathcal{H}$*.*

The sequence of events with the i-th buyer is modeled as follows. (1) Buyer i arrives at t_i and proposes an acceptable range $[\eta D_i, D_i]$ for quantity. (2) Seller $\mathcal{A}$ decides ρ_i. (3) Buyer i stays if $\rho_i \geq \eta$ and leaves permenently if otherwise. (4) If buyer i stays, they settle the transaction price w_i for quantity $Q_i = \rho_i D_i$ through *Nash bargaining*, in which information is assumed to be symmetric for simplicity.

Let β_i denote the exogenous, relative bargaining power of buyer i against seller $\mathcal{A}$. It captures exogenous factors such as bargaining skills and net cost of keeping a long-term relationship. We assume that β_i is known given the identity of buyer i. Conditional on history ψ, the bargaining power β of a potential buyer follows distribution $B(\cdot|\psi)$. The *generalized Nash bargaining model* predicts that if player j's payoff and outside option for the focal transaction are $\Pi_j(w)$ and d_j given the transaction price w, where $j \in \{\mathcal{A}\} \cup \{1, 2, \cdots, M\}$, then the bargaining results in price $w^* = \arg\max_w (\Pi_i(w) - d_i)^{\beta_i} \cdot (\Pi_A(w) - d_A)^{1-\beta_i}$. In particular, if $\Pi_i(w) - d_i + \Pi_A(w) - d_A$ is independent of w, then w^* splits the pie between the buyer and the seller in proportion to their respective bargaining powers.

For buyer i, let r_i and r_i' denote the profit margins before subtracting the cost of the product purchased from seller $\mathcal{A}$ and an alternative supplier, respectively, p the posted price for $\mathcal{A}$'s product, $\tilde{c}_i$ the marginal cost of buying from the alternative, Q_i' the quantity available from the alternative, and ρ_i' the corresponding fill rate. In addition, let $l_i = \frac{Q_i'}{Q_i} = \frac{\rho_i' \cdot D_i}{\rho_i \cdot D_i} = \frac{\rho_i'}{\rho_i}$. Accordingly, the total payoff for buyer i is $\Pi_i(w_i) = (r_i - w_i) \cdot Q_i$ and the outside option is

$$\begin{aligned} d_i =& \max\left\{Q_i \cdot (r_i - p), Q_i' \cdot (r_i' - \tilde{c}_i)\right\} \\ =& Q_i \cdot \max\left\{r_i - p, l_i \cdot (r_i' - \tilde{c}_i)\right\} \\ =& Q_i \cdot \left[r_i - \min\left\{p, r_i - l_i \cdot r_i' + l_i \cdot \tilde{c}_i\right\}\right]. \end{aligned} \tag{4}$$

Let $\bar{c}_i = r_i - l_i \cdot r_i' + l_i \cdot \tilde{c}_i$ represent the net marginal cost of buying from the alternative supplier in order to keep the same margin r_i. If $\bar{c}_i > p$, it is not credible for buyer i to switch, so the outside option is to buy from seller $\mathcal{A}$ at the posted price. This is possible because products are not perfectly substitutable, and $\bar{c}_i$ includes switching costs such as searching, redesigning, damage to the brand image, and so on. We assume that $\bar{c}_i$ is unknown to the seller *ex ante* but will be revealed during the negotiation. For a potential customer who has not arrived, $\bar{c}$ follows distribution $G(\cdot|\psi)$ given history ψ. Regarding the link between l_i and D_i or Q_i, we can verify with simulation

Table 8: Summary of Notations

λ_i	Buyer arrival rate after the $(i-1)$-th buyer	ρ_i	Fill rate for buyer i	β	Bargaining power
λ_0	Arrival rate of the tech shock	η	Lower bound of fill rate	B	Distribution of β
δ_n	Probability of tech shock when n buyers have arrived	t_i	Arrival time of buyer i	Π_A	Seller's payoff
s	Marginal salvage value	ψ_i	History up to time t_i	Π_i	Buyer i's payoff
κ	The total capacity	F	Distribution of demand	d	Outside option
K_i	Capacity available to buyer i	$\bar{c}$	Net marginal cost of buying from an alternative supplier	p	The posted price
D_i	Demand of buyer i	G	Distribution of $\bar{c}$	w	Transaction price
Q_i	Capacity allocated to buyer i	c_L	Lower bound of $\bar{c}$	V	The value function

that when $\eta > 0.8$, the correlations between l_i and D_i and between l_i and Q_i, respectively, are highly insignificant and thus $\partial l_i/\partial D_i \approx 0$ and $\partial l_i/\partial Q_i \approx 0$.[12] Hence, we make the Assumption 3 to simplify our analysis. Admittedly, if η is small this assumption may not hold and the price-quantity relationship will become more complicated.

Technical Assumption 3. $\partial l_i/\partial D_i = \partial l_i/\partial Q_i = 0$ *for all* $i \in \{1, 2, \cdots, M\}$.

For seller $\mathcal{A}$, let $V(K, p, \psi(t))$ represent the expected revenue obtained *after* time t given remaining capacity K, posted price p, and history $\psi(t)$. Therefore, when bargaining with buyer i, seller $\mathcal{A}$ has expected payoff $\Pi_A(w_i) = w_i Q_i + V(K_i - Q_i, p, \psi_i)$ and outside option $d_A = V(K_i, p, \psi_i) \cdot \mathbb{I}\{\bar{c}_i \le p\} + [pQ_i + V(K_i - Q_i, p, \psi_i)] \cdot \mathbb{I}\{\bar{c}_i > p\}$, where $\mathbb{I}\{\cdot\}$ is an indicator function. In addition, we assume that $\bar{c}_i \ge c_L > s$ for every i so that the bargaining always has a solution (i.e., c_L, the highest possible price a customer would like to pay, is higher than the marginal value for the seller). Hence, we have the following lemma, which determines whether buyer i pays the posted price or engages in the price bargaining. Notice that β_i and $\bar{c}_i$ are known when the buyer arrives and are thus taken as certain in the bargaining.

[12] First, it is the industry standard to fill at least 90% by any seller according to our interaction with practitioners. Note that ρ_i may not be a monotonic function of D_i given the complexity of the value function. Thus, the intuition is that, when η is close to 1, ρ_i could stick to a boundary or bounce back and forth between η and 1 as D_i changes.

Lemma 1. *If $\bar{c}_i > p$, buyer i pays the posted price; If $\bar{c}_i \leq p$, Nash bargaining results in*

$$w_i = \beta_i \cdot \frac{V(K_i, p, \psi_i) - V(K_i - Q_i, p, \psi_i)}{Q_i} + (1 - \beta_i) \cdot \bar{c}_i. \tag{5}$$

Lemma 1 simply says that *the chance of a buyer engaging in a price negotiation increases with the posted price p*. Hence, the higher the posted price, the more bargainers. It also says that the negotiated price is a function of the available capacity and transaction quantity. Based on Lemma 1, we know that as long as $\partial V/\partial K \geq 0$, our model satisfies the property that larger quantities entail larger total payments.[13] In order to understand how w_i is affected by K_i and Q_i, we need to know more about value function V as well as other factors.

5.2 Source of Non-Monotonicity

In our model, we propose that buyer bargaining power β and net switching cost $\bar{c}$ are not the source of price-quantity non-monotonicity and thus assume for simplicity that they are independent of purchase quantity Q. To verify our conjecture, we compare three different models by running nonlinear regressions. To proceed, first note that from Lemma 1 we have

$$w_{ij} = \mathbb{I}\{\bar{c}_{ij} \geq p_j\} \cdot p_j + \mathbb{I}\{\bar{c}_{ij} < p_j\} \cdot [\beta_{ij} \cdot \triangle\, \hat{v}_{ij} + (1 - \beta_{ij}) \cdot \hat{c}_{ij}] \cdot p_j, \tag{6}$$

where $\triangle\, \hat{v}_{ij} = [V(K_{ij}, p_j, \psi_{ij}) - V(K_{ij} - Q_{ij}, p_j, \psi_{ij})] / (p_j Q_{ij})$ and $\hat{c}_{ij} = \bar{c}_{ij}/p_j$. Accordingly, we can derive the discount received by buyer i for product j:

$$1 - \frac{w_{ij}}{p_j} = \mathbb{I}\{\hat{c}_{ij} < 1\} \cdot [1 - \beta_{ij} \cdot \triangle\, \hat{v}_{ij} - (1 - \beta_{ij}) \cdot \hat{c}_{ij}]. \tag{7}$$

Now we can see that three factors can possibly contribute to the non-monotonicity we are after: $\hat{c}_{ij}$, β_{ij}, and $\triangle\, \hat{v}_{ij}$. Hence, we consider three different models. In preparation, we define $\phi(x) = a_1 \cdot (x - B)_- + a_2 \cdot (x - B)_-^2 + a_3 \cdot (x - B)_+ + a_4 \cdot (x - B)_+^2$, which is the piece-wise polynomial function we used to capture the non-monotonicity in the empirical analysis. In model (I), we assumes that the non-monotonicity is rooted in how the seller values capacity. In particular, $\hat{c} = b_c' \cdot X_c$, $\beta = b_b' \cdot X_b$, and $\triangle\, \hat{v} = \phi(r4ds) + b_v' \cdot X_v$. In model (II), we assume that the non-monotonicity is originated from the net switching cost. In particular, $\hat{c} = \phi(r4ds) + b_c' \cdot X_c$, $\beta = b_b' \cdot X_b$, and $\triangle\, \hat{v} = b_v' \cdot X_v$. In model (III), we assume that the non-monotonicity is due to quantity-dependent bargaining power.

[13]It is easy to check that $\partial(w_i Q_i)/\partial Q_i = \beta_i V_K'(K_i - Q_i, p, \psi_i) + (1 - \beta_i) \cdot \bar{c}_i$.

In particular, $\hat{c} = b_c' \cdot X_c$, $\beta = \phi(r4ds) + b_b' \cdot X_b$, and $\triangle\, \hat{v} = b_v' \cdot X_v$. Note that although the three models have the same components, they are different in structures. Regarding other explanatory variables, we use $lndod$ and 10 major brand names for X_c, $Cshr$ and $Vrate$ for X_b, and $Cbase$, $lndod$, $CapL$, and three quarters for X_v. At last, we run three nonlinear regressions based on the following equation:

$$ED_{ij} = \mathbb{I}\{\hat{c}_{ij} < 1\} \cdot [1 - \beta_{ij} \cdot \triangle\, \hat{v}_{ij} - (1 - \beta_{ij}) \cdot \hat{c}_{ij}] + \hat{\epsilon}_{ij}. \tag{8}$$

The results are summarized in Table 9. Notice that the non-monotonicity is statistically significant only in model (I). In addition, given the same number of parameters or degree of freedom (d.f.), model (I) has the highest R^2 and the lowest sum of squared residuals (SS). If we can assume that $\hat{\epsilon}$ is normally distributed, we can use Akaike's Information Criterion (AIC) (Akaike 1981) to compute the evidence ratio (i.e., how much more likely) of one model against another. We first compute the corrected AIC value defined by

$$AIC_C = N \cdot \ln\left(\frac{SS}{N}\right) + \frac{2 \cdot J \cdot N}{N - J - 1}, \tag{9}$$

where N is the number of observations and J is the number of parameters in the model plus one. Next, we can obtain the evidence ratio defined by

$$\text{Evidence Ratio} = \frac{\Pr\{\text{model (I) is correct}\}}{\Pr\{\text{model (II) is correct}\}} = \exp\left(\frac{AIC_C^{(II)} - AIC_C^{(I)}}{2}\right). \tag{10}$$

Accordingly, we know that model (I) is 2.26×10^{15} times more likely against model (II) and 239 times more likely against model (III) to be the correct one. In other words, the evidence is overwhelmingly in favor of model (I). Therefore, combining all the results, we conclude that model (I) is the correct model among the three.

Regarding other possible models, the most plausible is the combination of model (II) and (III). However, there will be a serious collinearity problem when we let β depend on $r4ds$ in (II) or let $\hat{c}$ depend on $r4ds$ in (III). Hence, more complicated models cannot provide better explanations. Lastly, if we add $r4ds$ as a linear part of $\hat{c}$ in model (I), the estimated coefficient is not significant, so it is reasonable to assume that $\hat{c}$ is not correlated with $r4ds$.

Table 9: Selected Regression Results for the Three Alternative Models

	(I)	(II)	(III)
a_1	-0.6508*** (0.2148)	-0.2475* (0.1354)	1.0113** (0.4718)
a_2	-1.0167** (0.4417)	-0.3875 (0.3095)	1.1682 (0.9095)
a_3	1.8758*** (0.5942)	0.4899 (0.3113)	-1.5467 (1.5779)
a_4	-8.7269*** (2.7016)	-2.0765 (1.2606)	5.0896 (6.7346)
B	0.5865*** (0.0196)	0.5795*** (0.0397)	0.6297*** (0.0474)
R^2	0.5140	0.4991	0.5117
d.f.	27	27	27
SS	50.6148	52.1636	50.8517
AIC_C	-8945	-8874	-8934

Notes. Standard errors are in parentheses. SS: sum of squared errors.
*p<0.1; **p<0.05; ***p<0.01.

6 Theoretical Analysis

In this section, we first derive a sufficient condition on the value function for the price-quantity curve to be non-monotonic. We then analyze the seller's problem, formulate the value function, and investigate its property. Finally, we try to simulate the price curve given a certain form of the value function.

6.1 A Sufficient Condition for Non-Monotonicity

We assume V is a non-decreasing and twice-differentiable function of capacity K. To simplify the notation, we write $V(K, p, \psi_i) = V_i(K)$. We know that how the price w_i changes with quantity Q_i depends on the sign of the first-order derivative of w_i in (5) with respect to Q_i:

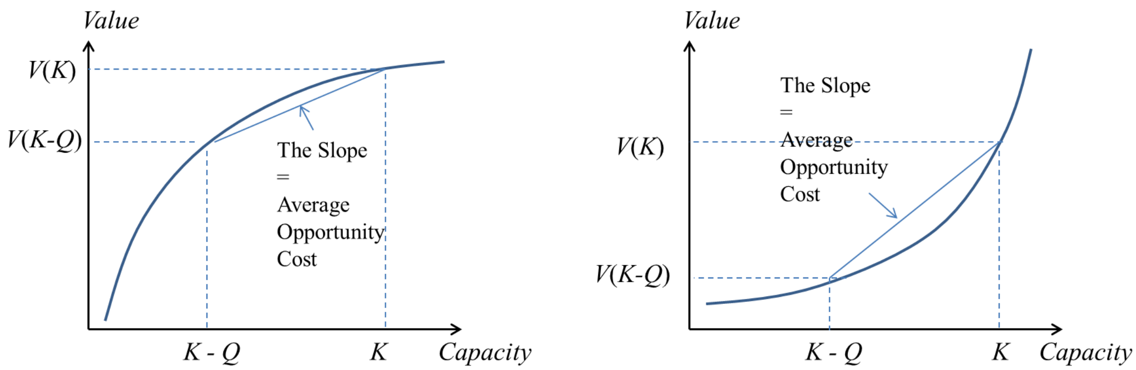
$$\frac{\partial w_i}{\partial Q_i} = \frac{\beta_i}{Q_i}\left[V_i'(K - Q_i) - \frac{V_i(K) - V_i(K - Q_i)}{Q_i}\right]. \tag{11}$$

Given that $\frac{\beta_i}{Q_i} > 0$, the sign of $\frac{\partial w_i}{\partial Q_i}$ depends on that of $V_i'(K - Q_i) - \frac{V_i(K) - V_i(K - Q_i)}{Q_i}$. A simple examination leads us to the following proposition.

Proposition 1. *If $V_i(x)$ is concave for any $x \in [0, K]$, then w_i increases with Q_i. If $V_i(x)$ is convex for any $x \in [0, K]$, then w_i decreases with Q_i.*

Although the above results are simple, they are surprising. Our initial intuition is that the value function should be concave and the price should decrease with quantity. However, in order to have the quantity discount, our simple model requires the value function to be convex. Figure

Figure 4: The Intuition Behind Proposition 1



4 illustrates the intuition behind Proposition 1. Note that given buyer i's outside option and the bargaining power, w_i depends on the seller's average opportunity cost of selling Q_i units. We can see that as Q_i increases, the average opportunity cost increases if the value function is concave and decreases if the value function is convex. Based on this observation, we suspect that the value function may not be simply convex or concave, which may be the reason for a non-monotonic price-quantity relation. In fact, we can show that a simple combination of convexity and concavity for the value function will generate a non-monotonic price-quantity curve.

Proposition 2. *If there exists $x' \in (0, K)$ such that $V_i(x)$ is strictly convex for $x \in [0, x']$, strictly concave for $x \in [x', K]$, and $V_i'(0) < V_i(K)/K$, then there exists $x'' \in (0, K)$ such that w_i increases with Q_i for $Q_i \in [0, K - x'']$ and decreases with Q_i for $Q_i \in [K - x'', K]$.*

Proposition 2 provides us with a sufficient condition for the price-quantity relation to be non-monotonic. We call such a property *convex-concave*. Actually, it is reasonable to expect the value function to be convex-concave or *S-shaped*. When the capacity is very low, the seller is unlikely to fulfill any buyer's need and will have to salvage the capacity. As the capacity increases, it becomes more and more likely that the capacity is sufficient to satisfy more buyers' needs. When the capacity is very high, it may exceed demand and the seller may have to salvage a portion. We can infer from Figure 4 that when the value function is convex-concave and the capacity level is high enough, the average cost of selling Q units for the seller first increases and then decreases with Q, which leads to a non-monotonic price-quantity relation.

However, this result alone is not satisfactory, because it cannot explain the pricing pattern we observe in the data. The discussion in the previous paragraph is based on perturbing the purchase quantity of a single buyer with a fixed-value function. Notice that the seller may update the estimation of future demand based on the demand of the current buyer. Thus, the shape of the value function may be different for buyers with different purchase quantities, which may explain the empirical pattern. In the following, we try to verify our conjecture by formulating and analyzing the value function.

6.2 Formulating the Value Function

Assume that K units of capacity is available after the $(i-1)$-th buyer leaves. Let us consider the value of this remaining capacity in four cases that constitute the sample space. First, the leftover capacity will be salvaged with probability δ_i. Second, if buyer i arrives, the buyer walks away immediately if the capacity is insufficient. Hence, if $K < \eta D_i$, the seller's expected revenue at t_i is $V_i(K)$. Third, if $K \geq \eta D_i$ and $\bar{c}_i > p$, the expected revenue is $pQ_i + V_i(K - Q_i)$. Fourth, if $K \geq \eta D_i$ and $\bar{c}_i \leq p$, the expected revenue is $(1-\beta_i)\,\bar{c}_i Q_i + \beta_i\left[V_i(K) - V_i(K - Q_i)\right] + V_i(K - Q_i)$. Note that $Q_i = \rho^*(K, D_i, \beta_i, \psi_i) \cdot D_i$ can differ for different parameter values. As a result,

$$\begin{aligned} V_{i-1}(K) = \delta_i \cdot s \cdot K + (1-\delta_i) \cdot \Bigg\{ & \int_{K/\eta}^{+\infty} V_i(K)\, dF(D|\psi_{i-1}) \\ & + \int_0^1 \int_0^{K/\eta} \int_p^{+\infty} \left[p\rho^* D + V_i(K - \rho^* D)\right] dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \\ & + \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} \left[\beta V_i(K) + (1-\beta)\, V_i(K - \rho^* D)\right] dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \\ & + \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} (1-\beta)\, \bar{c}\rho^* D dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \Bigg\}. \end{aligned} \tag{12}$$

Apparently, this is a complicated function and it is not obvious how $V_{i-1}(K)$ is affected by various parameters. Hence, we try to derive approximations for the value function in two cases: one with a finite buyer group and the other with a very large buyer group (i.e., $M \to \infty$).

6.3 An Approximation with Finite M

In this section, we derive an upper and a lower bound for $V_{i-1}(K)$ and we present the results in Theorem 1. In preparation, we define

$$\nu(p, \psi_{M-1}) \quad = \quad p - s - \mathbf{E}_{\beta,\bar{c}}\left[(p-(1-\beta)\bar{c}-\beta s)\cdot \mathbb{I}\left\{\bar{c}\le p\right\}|\psi_{M-1}\right]. \tag{13}$$

The key idea of the proof for the lower bound is the following. First, setting $\rho = 1$ in (12) leads to a lower bound of $V_{i-1}(K)$. We then utilize the fact that the lower bound is a separable function for D and we iteratively plug the lower bound into (12). We complete the proof by induction. The proof for the upper bound is similar, except that we use $\rho^* \le 1$. An essential tool we use is the law of iterative expectation, based on which we have $\mathbf{E}_{\beta,\bar{c}}\left[\nu(p,\psi_{i-1})|\psi_{i-2}\right] = \nu(p,\psi_{i-2})$ and $\int_0^{+\infty}\int_0^{\lambda K} DdF\left(D|\psi_{i-1}\right)dF\left(D'|\psi_{i-2}\right) = \mathbf{E}\left[\mathbf{E}\left[D\cdot\mathbb{I}\left\{D\le\lambda K\right\}|\psi_{i-1}\right]|\psi_{i-2}\right] = \int_0^{\lambda K} DdF\left(D|\psi_{i-2}\right)$.

Theorem 1. *For any* $1 \le i \le M$ *and* $K \ge 0$,

$$V_{i-1}(K) \ge sK + \left(1-\delta_i^l\right)\cdot\nu(p,\psi_{i-1})\cdot\int_0^{K\cdot\lambda^{M+1-i}} DdF\left(D|\psi_{i-1}\right), \tag{14}$$

$$V_{i-1}(K) \le sK + \left(1-\delta_i^u\right)\cdot\nu(p,\psi_{i-1})\cdot\int_0^{K/\eta} DdF\left(D|\psi_{i-1}\right), \tag{15}$$

where $\delta_i^l = \delta_i - (1-\delta_i)\left(1-\delta_{i+1}^l\right)F_0\left(K-\lambda K\right)$, $\delta_i^u = \delta_i - (1-\delta_i)\left(1-\delta_{i+1}^u\right)$, *and* $\delta_M^u = \delta_M^l = \delta_M$.

Because $\lim_{K\to+\infty} F_0\left(K-\lambda K\right) = 1$, we have $\lim_{K\to+\infty}\delta_i^l = \delta_i^u$. Therefore, the upper and lower bounds converge as K goes to infinity. Both the upper and lower bounds take a functional form similar to $\hat{V}_{M-1}(K,\lambda)$, and it is reasonable to expect that $V_{i-1}(K)$ is similar to the bounds as long as they are close enough. We may also conclude that the shape of the value function is largely dependent on the demand distribution of the next buyer. However, the gap between the upper and lower bounds increases with M. Hence, the bounds will perform well when M is not extremely large. Otherwise, it may be useful to get bounds that are independent of M.

6.4 An Approximation with Infinite M

In this case, we assume that the buyer arrival rate is constantly λ_b. Hence, the probability of ending the selling process is constantly $\delta = \frac{\lambda_0}{\lambda_0+\lambda_b}$. Before we analyze $V_{i-1}(K)$, note that we can write it as $\mathbf{E}\left[\mathcal{R}\left(K,\{D_n,\beta_n,\bar{c}_n,t_n-t_{n-1}\}_{n=i}^{\infty}\right)|\psi_{i-1}\right]$, where $\mathcal{R}$ is the total revenue, which is a function of the future demand, buyer bargaining power, net marginal cost of buying outside, and arrival times. Similarly, $V_i(K) = \mathbf{E}\left[\mathcal{R}\left(K,\{D_n,\beta_n,\bar{c}_n,t_n-t_{n-1}\}_{n=i+1}^{\infty}\right)|\psi_i\right]$. Based on our assumptions, $\{D_n, \beta_n, \bar{c}_n, t_n - t_{n-1}\}_{n=i}^{\infty}$ and $\{D_n, \beta_n, \bar{c}_n, t_n - t_{n-1}\}_{n=i+1}^{\infty}$ are statistically equivalent given information ψ_{i-1}. Thus, using this condition and the law of iterative expectation, we get $\mathbf{E}\left[V_i\left(K\right)|\psi_{i-1}\right] = \mathbf{E}\left[\mathcal{R}\left(K, \{D_n, \beta_n, \bar{c}_n, t_n - t_{n-1}\}_{n=i+1}^{\infty}\right)|\psi_{i-1}\right] = V_{i-1}\left(K\right)$. Leveraging this property of the value function, we obtain an upper bound and a lower bound for $V_{i-1}\left(K\right)$. In preparation, let

$$H_i\left(K\right) = \mathbf{E}\left[p - (p - (1-\beta)\bar{c}) \cdot \mathbb{I}\left\{\bar{c} \le p\right\}|\psi_{M-1}\right] \cdot \int_0^{K/\eta} DdF\left(D|\psi_{i-1}\right), \tag{16}$$

which is an *approximate measure* for the expected revenue obtained from the i-th buyer. Let

$$h_i(K) = \mathbf{E}\left[V_i\left(\left[K - D_i\right]^+\right)|\psi_{i-1}\right]/V_{i-1}\left(K\right). \tag{17}$$

It is easy to see that $h_i(K) \in [0,1]$. Now we can introduce the following theorem.

Theorem 2. *For any $i \ge 1$ and $K \ge 0$, we have*

$$\frac{s \cdot K + \frac{1-\delta}{\delta} \cdot H_i\left(K\right)}{1 + \frac{1-\delta}{\delta} \cdot \left[1 - h_i(K)\right] \cdot \left(1 - \mathbf{E}\left[\beta|\psi_{i-1}\right] \cdot G(p|\psi_{i-1})\right)} \le V_{i-1}\left(K\right) \le s \cdot K + \frac{1-\delta}{\delta} \cdot H_i\left(K\right). \tag{18}$$

If $\frac{\partial}{\partial K} V_i\left(K\right) \ge s$ for any $i \ge 1$, $K \ge 0$, and ψ_i, then $\lim_{K\to\infty} h_i(K) = 1$.

Let U_{i-1} and L_{i-1} be the upper and lower bounds in (18), respectively. We have that $L_{i-1} = U_{i-1}/\left(1 + Z_i\right)$, where $Z_i = \frac{1-\delta}{\delta} \cdot \left[1 - h_i(K)\right] \cdot \left(1 - \mathbf{E}\left[\beta|\psi_{i-1}\right] \cdot G(p|\psi_{i-1})\right)$. We can see that the percentage gap, $\frac{U_{i-1} - L_{i-1}}{U_{i-1}} = \frac{Z_i}{1+Z_i}$, goes to zero as $K \to \infty$. The absolute gap, $U_{i-1} - L_{i-1} = \frac{Z_i}{1+Z_i} \cdot U_{i-1}$, also goes to zero if $s = 0$. When $s > 0$, the size of the absolute gap depends on $Z_i \cdot K$.

The condition $\frac{\partial}{\partial K} V_i\left(K\right) \ge s$ should be satisfied by definition, because we assume that the seller can always salvage the capacity at marginal value s, and thus s should be the lowest marginal value for $V_i\left(K\right)$. This means that *the bounds will perform particularly well at the beginning of the selling season*, when the capacity is relatively large compared with the average buying quantity. The gap is also decreasing in δ, $\mathbf{E}\left[\beta|\psi_{i-1}\right]$, and $G(p|\psi_{i-1})$. In other words, the bounds are closer to the true value function when the leftover capacity is more likely to be salvaged, buyers are more powerful on average, and buyers are more likely to engage in price bargaining.

6.5 Discussion

Note that $H_i\left(K\right)$ can be written in the form of $a' \cdot \int_0^{K \cdot a''} DdF\left(D|\psi_{i-1}\right)$, where a' and a'' are parameters independent of K. Hence, the upper and lower bounds given by Theorem 1 and 2 can all be written in the form of $a \cdot s \cdot K + a' \cdot \int_0^{K \cdot a''} DdF\left(D|\psi_{i-1}\right)$, where $a, a',$ and a'' are parameters independent of K. Moreover, we learn from Eq. (20) that the value function of a single-period problem takes a similar form. Therefore, we are basically approximating the value function by a single-period problem in which the seller treats the next buyer as the last one. This is very likely to be the mental heuristic used by a salesperson. More importantly, the approximations in both cases suggest that the shape of the value function depends much on the demand distribution of the next buyer, which supports our conjecture in Section 6.1. We find that if the demand is normally distributed, the value function is very likely to be *convex-concave* as described in Proposition 2.

Proposition 3. *If the demand is normally distributed, then the bounds are all convex-concave.*

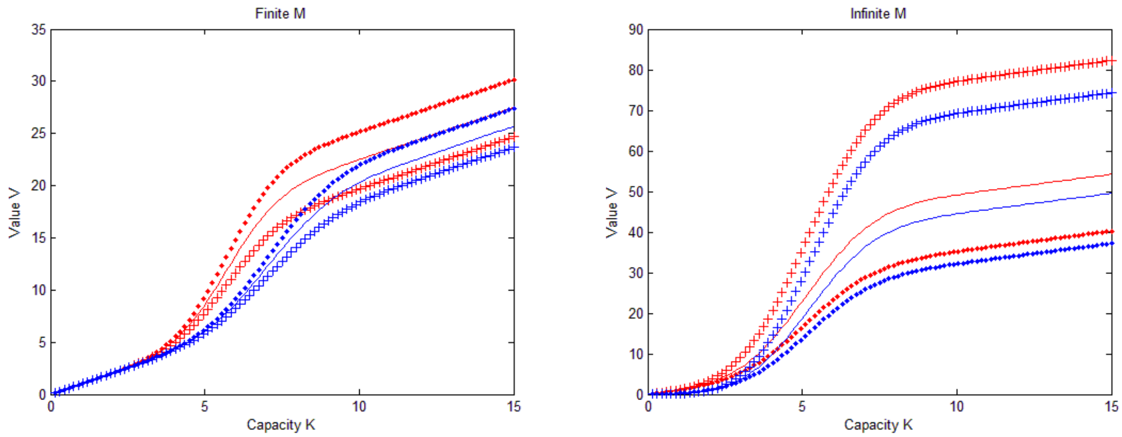
Other distributions—for example, any unimodal distribution—may also generate S-shaped bounds, and it is quite natural to expect a unimodal demand distribution.

6.6 Plotting the Price Curve

From (5) we know that the negotiated transaction price w_i is a linear combination of both parties' outside options. The seller's outside option is the average opportunity cost of selling Q_i, which depends on Q_i, capacity K_i, and the value function. In this section, we first investigate the performance and the shape of the bounds given a normal demand distribution and then try to plot the price curve using different parameter settings.

Without loss of generality, we consider negotiating with buyer i for quantity Q_i at time t_i. As in the regression models, we control for capacity level, bargaining power, posted price, demand uncertainty, as well as all the other buyer-, product-, and market-related factors. In the base case, we set $K = 10$, $\beta_i = 0.8$, $\mathbf{E}\left[\beta|\psi_i\right] = \beta_i$, $p = 8$, $s = 1$, $c_L = 6$, $\bar{c}_i = 7$, $G\left(\bar{c}|\psi_i\right) = 1 - \exp\left(c_L - \bar{c}\right)$, and $\eta = 0.9$. We assume that arrival rates satisfy $\frac{\lambda_i}{\lambda_0} = \frac{M-i}{i^2}$, which means that buyer arrival rate is linear in the number of potential buyers and the technology-shock arrival rate increases quadratically in the number of buyers that have arrived. In addition, we assume that the demand of the next buyer is normally distributed with mean μ and standard deviation $\sigma = \mu \cdot C_V$, where $C_V = 0.25$ is a constant. When observing Q_i, the seller updates belief and set $\mu = \min\left\{12, 16 - s_m \times Q_i\right\}$ where s_m captures the market structure—larger s_m means more concentrated demand. This way of updating means that if buyer i is very large, the rest of the buyers are likely to be small, especially when the seller

Figure 5: Illustrations of Bounds for the Value Function



Note. The three sets of lines illustrate the upper and lower bounds of V_i for different M and δ. In the case of finite M: "+" for $M - i = 3$; "–" for $M - i = 4$; "•" for $M - i = 5$. In the case of infinite M: "+" for $\delta = 0.2$; "–" for $\delta = 0.3$; "•" for $\delta = 0.4$.

knows in advance the market structure and the identities of the buyers. F_0 is normal distribution with mean 12 and standard deviation $12 \cdot C_V$. Finally, we set $h_i(K) \approx 1 - s \cdot \mathbf{E}\left[D_i|\psi_{i-1}\right]/V_{i-1}(K)$. See the proof of Theorem 2 for justifications.

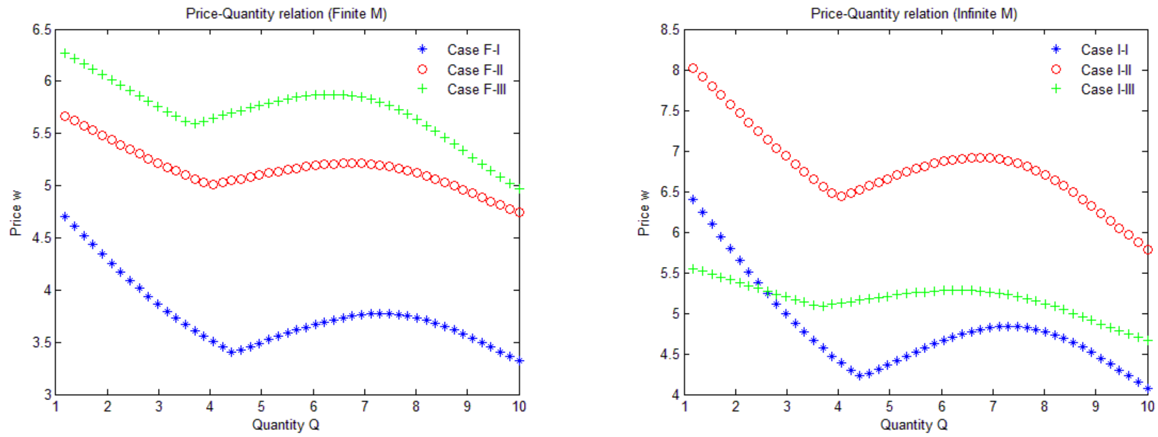
In Figure 5, we present three numerical examples of the bounds for both finite and infinite M. We can see that the bounds for finite M perform better with smaller M; the performance of the bounds for infinite M depends on the assumption of δ, and they work better with larger δ. In both cases, the bounds are S-shaped given the normal demand distribution.

In Figure 6, we use the upper bounds in each case (of finite vs. infinite M) as the approximation of the value function and generate the negotiated price for three different scenarios. With an S-shaped value function, we obtain a price-quantity curve in all scenarios that is reversed-N-shaped, which is consistent with our empirical observations.

7 Managerial Implications

According to our model, the reason some buyers are receiving lower discounts than who buy less is as follows: large buyers accelerate the selling process and small buyers are helpful in consuming the residual capacity. However, satisfying mid-sized buyers may be costly because supplying these buyers can make it difficult to utilize the remaining capacity, which is too much for small buyers

Figure 6: Model-Generated Price-Quantity Relation



Note. Case F-I: $\beta_i = 0.8$; $M = 3$; $s_m = 0.9$; $C_V = 0.2$. Case F-II: $\beta_i = 0.5$; $M = 3$; $s_m = 1$; $C_V = 0.25$. Case F-III: $\beta_i = 0.5$; $M = 5$; $s_m = 1.1$; $C_V = 0.3$.

Case I-I: $\beta_i = 0.8$; $\delta = 0.4$; $s_m = 0.9$; $C_V = 0.2$. Case I-II: $\beta_i = 0.5$; $\delta = 0.4$; $s_m = 1$; $C_V = 0.25$. Case I-III: $\beta_i = 0.5$; $\delta = 0.6$; $s_m = 1.1$; $C_V = 0.3$.

but not enough for large buyers. In this section, we begin by discussing the capacity allocation decision for the seller given the non-monotonic price-quantity relation. Basically, the seller need to avoid mid-sized transactions by controlling the capacity allocation. Next, we discuss how the posted price should be set given its influence in the selling process.

7.1 Dynamic Capacity Rationing

In this setting the seller should control the capacity that is allocated to each buyer in the acceptable range. Given the complexity of the value function and the price-quantity relation, it is not immediately clear whether the seller should increase or decrease the allocated capacity. Based on our model, we derive a simple rule for deciding the quantity.

Proposition 4. *The seller should increase ρ_i if $r_i > V_i'(K - Q_i)$ and decrease if $r_i < V_i'(K - Q_i)$.*

The above result suggests that the rationing decision depends on the remaining capacity level, purchase quantity, demand distribution, and the buyer's profit margin before subtracting the cost of this component product. If we hold r_i constant, the allocated capacity should be reduced when $K - Q_i$ is close to the mean of the demand from the next buyer; otherwise, the seller should sell as much as possible. The logic is straightforward: the seller need to avoid losing the next major buyer due to insufficient capacity. Mathematically, due to the shape of the value function, it is more likely to have high $V_i'(K-Q_i)$ when $K-Q_i$ is neither too high nor too low. On the other hand, if we hold $V_i'(K-Q_i)$ constant, then capacity reduction is more likely to benefit the seller when the buyer has a lower profit margin. The logic is clear: reserve capacity for buyers who are willing to pay more. Overall, the lesson is that allocation decisions cannot be based solely on the capacity levels and purchase quantities. Further, incorrect assumptions on the value function lead to suboptimal decisions.

7.2 Posted-Price Optimization

When the posted price is determined by marketing people, two important factors related to the selling process should be considered. First, we learn from Lemma 1 that the posted price determines not only the price a buyer pays but also the number of price-takers. A price that is too low undercuts the seller's profitability; a price that is too high encourages more buyers to engage in bargaining. Second, the seller's revenue is a function of both the posted price and the average discount received by bargainers. Incorrect anticipations of the average discount will lead to suboptimal posted prices. Our model can be used by sellers to optimize the posted price while considering the price-taker-bargainer trade-off and a non-monotonic price-quantity relationship with bargainers. The optimal posted price is $p^* = \arg\max_p V_0(\kappa, p)$.

7.3 Implications for Other Industries

There are other industries that resemble the semiconductor industry. For example, in the travel industry, airline companies and hotels have limited capacities and these capacities have to be sold within a limited period. Customers in these industries include bulk buyers such as travel agencies and resellers who purchase different quantities and negotiated prices. Findings from our study may carry over to such businesses. Other examples may include movie theaters, concerts, and sports events. There are a few important differences between the semiconductor industry and the others. In the semiconductor industry the obsolescence date is stochastic and deterministic for the others. Also, purchase quantities may be subject to negotiations in other industries. However, other industries may behave as if they have stochastic obsolescence dates due to the uncertainty in arrivals of buyers (i.e., the seller is not sure if another buyer will come along before the capacity is salvaged). In addition, even if the quantity is subject to negotiation, we may still observe a non-monotonic price-quantity relationship in other industries, because selling mid-sized quantities is still costly for sellers given fixed capacities. Theoretically, Lemma 1 and Eq. (6) still hold even if both price and quantity are determined in Nash bargaining. Therefore, the formulation of the value function as well as the subsequent analysis is unchanged.

8 Concluding Remarks

In this data-driven research, we study the price-quantity relation in B2B markets where the product life cycle is short and prices are set through one-shot negotiations. Using data from the microprocessor market, we found that, statistically, the transaction price can be a non-monotonic function of the transaction quantity. Contrary to our intuition, larger quantities—in a certain range—can actually lead to higher prices. We showed the robustness of this statistical result with multiple linear regression models. While existing theories cannot explain our observation, we built a model that allows us to delve into this phenomenon and understand the rationale behind it.

Our analysis reveals that it is fairly plausible for the price-quantity relation to be non-monotonic. A sufficient condition for a non-monotonic price-quantity curve is the value function being convex-concave in capacity. Although we normally assume that the value function is increasing and concave in capacity, our model shows that this need not be true in B2B markets. Instead, if the demand is normally distributed, the value function is likely to be convex-concave. More importantly, we found that a convex-concave value function is enough to explain our empirical observation: an N-shaped discount curve. We confirmed this finding by generating a price-quantity curve that is reversed N-shaped, using our model and the assumption of normally distributed demand.

To show that our theory is more likely to be the correct one against other explanations, we either control for other factors in the empirical model or fit different structured models to the data. We also show conditions under which omitted variables do not affect the non-monotonic pattern. Admittedly, certain behavior biases could also possibly cause this pricing pattern. Verifying such possibilities, however, entails completely different modelling frameworks. It is almost impossible to include these different models in the same paper. Therefore, we hope our work can stimulate future research on behavioral explanations of the empirical observation. Other limitations of our work include the assumptions of static Nash bargaining and fixed capacity, which might be restrictive for certain semiconductor companies. Future research may obtain different price-quantity relations by considering dynamic strategic bargaining and expiration of capacity over time.

Acknowledgments

We thank the editors and three anonymous reviewers for their valuable comments. We also thank Prof. Yingying Fan and Prof. Wenguang Sun at the Marshall School of Business at the University of Southern California and Prof. Hsiao-Hui Lee at the Faculty of Business and Economics at the University of Hong Kong for their comments on our empirical analysis.

References

[1] Akaike, H. 1981. Likelihood of a Model and Information Criteria. *Journal of Econometrics* 16(1) pp.3–14.

[2] Chipty, T., C.M. Snyder. 1999. The Role of Firm Size in Bilateral Bargaining: A Study of the Cable Television Industry. *The Review of Economics and Statistics* 81(2) pp.326–340.

[3] Cohen, M.A., T.H. Ho, Z.J. Ren, C. Terwiesch. 2003. Measuring Imputed Cost in the Semiconductor Equipment Supply Chain. *Management Science* 49(12) pp.1653–1670.

[4] Cooper, R.E. 2008. AMD v. Intel: An Assault on Price Competition. *Antitrust Chronicle* 3.

[5] Dukes, A.J., E. Gal-Or, K. Srinivasan. 2006. Channel Bargaining with Retailer Asymmetry. *Journal of Marketing Research* 43(1) pp.84–97.

[6] Gallego, G., G. van Ryzin. 1994. Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons. *Management Science* 40(8) pp.999–1020.

[7] Jeuland, A.P., S.M. Shugan. 1983. Managing Channel Profits. *Marketing Science* 2(2) pp.239–272.

[8] Karabuk, S., S.D. Wu. 2003. Coordinating Strategic Capacity Planning in the Semiconductor Industry. *Operations Research* 51(6) pp.839–849.

[9] Karabuk, S., S.D. Wu. 2005. Incentive Schemes for Semiconductor Capacity Allocation: A Game Theoretic Analysis. *Production & Operations Management* 14(2) pp.175–188.

[10] Kleywegt, A.J., J.D. Papastavrou. 1998. The Dynamic and Stochastic Knapsack Problem. *Operations Research* 46(1) pp.17–35.

[11] Kleywegt, A.J., J.D. Papastavrou. 2001. The Dynamic and Stochastic Knapsack Problem with Random Sized Items. *Operations Research* 49(1) pp.26–41.

[12] Kohli, R., H. Park. 1989. A Cooperative Game Theory Model of Quantity Discounts. *Management Science* 35(6) pp.693–707.

[13] Kuo, C.-W., H.-S. Ahn, G. Aydın. 2011. Dynamic Pricing of Limited Inventories When Customers Negotiate. *Operations Research* 59(4) pp.882–897.

[14] Lovejoy, W.S. 2010. Bargaining Chains. *Management Science* 56(12) pp.2282–2301.

[15] Nagarajan, M., Y. Bassok. 2008. A Bargaining Framework in Supply Chains: The Assembly Problem. *Management Science* 54(8) pp.1482–1496.

[16] Nash, J. 1950. The Bargaining Problem. *Econometrica* 18(2) pp.155–162.

[17] Oren, S.S., S.A. Smith, R.B. Wilson. 1982. Nonlinear Pricing in Markets with Interdependent Demand. *Marketing Science* 1(3) pp.287–313.

[18] Peng, C., F. Erhun, E.F. Hertzler, K.G. Kempf. 2012. Capacity Planning in the Semiconductor Industry: Dual-Mode Procurement with Options. *Manufacturing & Service Operations Management* 14(2) pp.170–185.

[19] Roth, A.E., M.K. Malouf. 1979. Game-Theoretic Models and the Role of Information in Bargaining. *Psychological Review* 86 pp.574–594.

[20] Snyder, C.M. 1998. Why Do Larger Buyers Pay Lower Prices? Intense Supplier Competition. *Economics Letters* 58 pp.205–209.

[21] Spence, M. 1977. Nonlinear Prices and Welfare. *Journal of Public Economics* 8 pp.1–18.

[22] Terwiesch, C., Z.J. Ren, T.H. Ho, M.A. Cohen. 2005. An Empirical Analysis of Forecast Sharing in the Semiconductor Equipment Supply Chain. *Management Science* 51(2) pp.208–220.

[23] Tulluri, K.T., G.J. van Ryzin. 2004. *The Theory and Practice of Revenue Management.* Kluwer Academic Publishers, Boston.

[24] Weinstock, D.S. 1982. Using the Herfindahl Index to Measure Concentration. *Antitrust Bulletin* 27 pp.285–301.

[25] Weng, Z.K. 1995. Channel Coordination and Quantity Discounts. *Management Science* 41(9) pp.1509–1522.

[26] Wu, S.D., M. Erkoc, S. Karabuk. 2005. Managing Capacity in the High-Tech Industry: A Review of Literature. *Engineering Economist* 50(2), pp.125–158.

Electronic Companion for "Higher Prices for Larger Quantities? Non-Monotonic Price-Quantity Relations in B2B Markets"

1 Robustness Checks

1.1 Simultaneity

One possible concern regarding our regression models is that they suffer simultaneity between price and quantity; i.e., not only is discount affected by demand share but demand share may also be affected by discount. This is not valid in our problem, because each buyer's production plan is determined in advance and the production needs inputs other than microprocessors from different suppliers so that it can be costly for buyers to manipulate their demand. To verify this practice with our data, we pick the *Quarter* from which an instance started in a year as an instrumental variable (IV). *Quarter* is a categorical variable used to control for time fixed effects in our previous regressions. Note that *Quarter* can directly influence ED given the cyclical nature of the semiconductor industry; However, *Quarter* can hardly relate to demand share directly because it is very unlikely that buyers of a particular demand share prefer to start a purchase from a particular quarter or are required to do so. Using Stata command *ivregress* and the *gmm* estimator, we regress $r4ds$ against ED and other control variables with dummies of three quarters as the IVs for ED. The details about the Stata command can be found at http://www.stata.com/manuals13/rivregress.pdf. The estimated coefficient for ED is -0.0227 with robust standard error 0.1719 and P value 0.895, which indicates that ED does not affect $r4ds$. We test the over-identifying restriction using the Hansen's J statistic (Hansen 1982) and obtain $\chi^2(2) = 0.1217$ and P value = 0.9410, which indicate that the IVs are valid. Therefore, the data suggests that quantity is not a term for negotiation.

1.2 Alternative Transformations of Demand Share

To check the robustness of power transformations of demand share, we run four additional nonlinear-piecewise-polynomial regressions using $DS^{0.15}$, $DS^{0.2}$, $DS^{0.3}$, and $DS^{0.35}$, respectively, in place of

Table 10: Regressions with Alternative Transformations of Demand Share

	$DS^{0.15}$	$DS^{0.2}$	$DS^{0.3}$	$DS^{0.35}$
a_1	.2543** (.1047)	.2390** (.1027)	.2242** (.1117)	.2335* (.1210)
a_2	.3705* (.2220)	.3263 (.2108)	.2843 (.2371)	.3072 (.2760)
a_3	-.6653** (.3374)	-.5221** (.2664)	-.3978* (.2125)	-.3560* (.1928)
a_4	4.6090** (1.9210)	2.8293** (1.1836)	1.6585** (.7133)	1.3402** (.5815)
B	.7159*** (.0265)	.6352*** (.0317)	.5060*** (.0386)	.4517*** (.0412)

Notes. B is obtained by NLS regression, and others are obtained by OLS regression given B. Robust standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

$r4ds$ in (3). The results are summarized in Table 10. In all four models, we can get similar non-monotonic curves composed of an increasing piece on the left and a quadratic piece on the right. Although a_3 is only significant at 10% level for $DS^{0.3}$ and $DS^{0.35}$, the results are meaningful as discussed earlier. First, we use robust standard errors that are generally greater than the ordinary standard errors. Second, we require a continuous discount curve, but the results will be more significant if we allow the curve to be discontinuous at the knot. Third, the P value should be halved for one-sided test, so the non-monotonicity is actually significant at 5% level. Hence, power transformations of demand share not only preserve the unit range, but also offer a robust way to study the price-quantity relationship.

1.3 Omitted Variables

Here we focus on the discussion of linear regressions, given that we only run nonlinear regressions to estimate knot B. Let $f(DS)$ denote the vector of demand-share-related variables, X the vector of other covariates included in our model, and $Z = (Z_1, Z_2)'$ the vector of those not included. As discussed earlier, Z_1 is the buyer's net cost of switching to an alternative product, and Z_2 is the value of other contract terms for the buyer. We show in Section 5.2 that the demand-share-dependent switching cost is not supported by the data, so it seems plausible to assume that Z_1 is uncorrelated with $f(DS)$. Hence, $Z_1 = b'_{z1} \cdot X + \varepsilon_{z1}$, wherein b_{z1} is a constant vector and ε_{z1} is a random variable that is independent of $f(DS)$. Next, price discount and other contract terms, if any, are determined through the same bargaining process, so $Z_2 = b_{z2}\,(\gamma'_{ds} \cdot f(DS) + \gamma'_x \cdot X + \gamma_{z1} \cdot Z_1) + \varepsilon_{z2}$. Therefore, if ED is determined by a linear model that $ED = \gamma'_{ds} \cdot f(DS) + \gamma'_x \cdot X + \gamma'_z \cdot Z + \varepsilon_1$, wherein $\gamma_z = (\gamma_{z1}, \gamma_{z2})'$. Defining $\gamma_0 = (1 + \gamma_{z2} b_{z2}) \cdot (\gamma_x' + \gamma_{z1} b_{z1}')$ and plugging Z into ED, we get

$$ED = (1 + \gamma_{z2} b_{z2}) \cdot \gamma_{ds}' \cdot f(DS) + \gamma_0' X + \varepsilon_0. \tag{19}$$

We can see that the true marginal effect of $f(DS)$ is just scaled if we regress ED against $f(DS)$ and X, and thus the underlying non-monotonic pattern will not be affected.

1.4 Sample Selection Bias

Recall that we dropped products with three or fewer customers because we do not have good proxies of the posted price for these products. Such a selection may cause biases because different products may have different price-quantity patterns. To correct for potential biases, we assume that the effective discount is censored for products with three or fewer customers and adopt the Heckman selection model or the Type-II Tobit model (Greene 2012). The full maximum likelihood method is used with the STATA command, *Heckman*. We run five selection models, $(S1)$ - $(S5)$, each with a different power-transformed demand share. For the ordinary regression, we use the same set of variables as in (3) except that we fix the knots according to the results in Table 6 and 10. In addition, we define $Lds = (ptds - B)_-$ and $Rds = (ptds - B)_+$. For the selection equation, we include 1) $Herf$, which measures the downstream market structure, 2) TSQ, which captures the size of the market, 3) $AvgP$, the average price of a product that captures the product quality and market segment, 4) $Lifecycle$, the life cycle of a product, 5) quarterly fixed effect, and 6) brand fixed effect. As shown in Table 11, an "N" shape is still observed and significant.

2 Technical Proofs

Proof of Lemma 1. If $\bar{c}_i > p$, then $\Pi_i(w_i) - d_i = (p - w_i) Q_i$ and $\Pi_A(w_i) - d_A = (w_i - p) Q_i$. Hence, it must be that $w_i = p$. If $\bar{c}_i \leq p$, $\Pi_i(w_i) - d_i = (\bar{c}_i - w_i) Q_i$ and $\Pi_A(w_i) - d_A = w_i Q_i + V_i (K_i - Q_i, p) - V_i (K_i, p)$. Hence, $\Pi_i(w_i) - d_i = (\bar{c}_i - w_i) Q_i = \beta_i (\Pi_i(w_i) - d_i + \Pi_A(w_i) - d_A) = \beta_i (\bar{c}_i Q_i + V_i (K_i - Q_i, p) - V_i (K_i, p))$, which results in (5). ■

Table 11: Test of Selection Bias on the Price-Quantity Relationship

		$DS^{0.15}$	$DS^{0.2}$	$DS^{0.25}$	$DS^{0.3}$	$DS^{0.35}$
DS-related Variables in the Ordinary Regression	Lds	.2561**	.2408**	.2337**	.2267**	.2363**
		(.1035)	(.1015)	(.1046)	(.1105)	(.1197)
	Lds^2	.3699*	.3262	.3059	.2857	.3093
		(.2223)	(.2106)	(.2183)	(.2359)	(.2743)
	Rds	-.6331*	-.4971*	-.4249*	-.3803*	-.3405*
		(.3340)	(.2637)	(.2283)	(.2101)	(.1906)
	Rds^2	4.4135**	2.7097**	1.9726**	1.5916**	1.2862**
		(1.8983)	(1.1696)	(.8580)	(.7044)	(.5741)
Dependent Variables in the Selection Equation	$Herf$	-4.4533***	-4.4533***	-4.4533***	-4.4533***	-4.4533***
		(.1936)	(.1936)	(.1936)	(.1936)	(.1936)
	TSQ	7.2E-7***	7.2E-7***	7.2E-7***	7.2E-7***	7.2E-7***
		(1.3E-7)	(1.3E-7)	(1.3E-7)	(1.3E-7)	(1.3E-7)
	$AvgP$	-6.8E-4***	-6.8E-4***	-6.8E-4***	-6.8E-4***	-6.8E-4***
		(2.3E-4)	(2.3E-4)	(2.3E-4)	(2.3E-4)	(2.3E-4)
	$Lifecycle$	2.1E-3***	2.1E-3***	2.1E-3***	2.1E-3***	2.1E-3***
		(2.1E-4)	(2.1E-4)	(2.1E-4)	(2.1E-4)	(2.1E-4)
	Quarter F.E.	Yes	Yes	Yes	Yes	Yes
	Brand F.E.	Yes	Yes	Yes	Yes	Yes

Notes. Robust standard errors are in parentheses. *$p<0.1$; **$p<0.05$; ***$p<0.01$.

Proof of Proposition 1. Suppose $V_i(x)$ is concave for any $x \in [0, K]$. Thus, we have $V_i'(x) < V_i'(K - Q_i)$ for any $x \in [K - Q_i, K]$. As a result, we have

$$\frac{V_i(K) - V_i(K - Q_i)}{Q_i} = \frac{1}{Q_i} \int_{K-Q_i}^{K} V_i'(x)dx < \frac{1}{Q_i} \int_{K-Q_i}^{K} V_i'(K - Q_i)\, dx = V_i'(K - Q_i)\,.$$

Similarly, we can get the result for $V_i(x)$ being convex for any $x \in [0, K]$. ■

Proof of Proposition 2. Let $L(x) = \frac{V_i(K)-V_i(x)}{K-x}$, and we have $L(0) > V_i'(0)$ and $\lim_{x\to K} L(x) = V_i'(K)$. By continuity, there exist $x'' \in (0, K)$ such that $L(x) > V_i'(x)$ for $\forall x \in [0, x'')$. We claim that $x'' < K$ and $L(x) \leq V_i'(x)$ for some $x \in [x'', K]$. Suppose this claim is not true. We have $L'(x) = \frac{1}{K-x}[L(x) - V_i'(x)] > 0$ for $\forall x \in [0, K]$ and thus $L(x)$ is strictly increasing on $[0, K]$. In addition, because $V_i(x)$ is concave on $[x', K]$, we have $V_i'(x) > V_i'(K) > L(x)$ on $[x', K)$, which is a contradiction.

Now let $x'' = \min\{x \in [0, K) : L(x) \leq V_i'(x)\}$. By continuity, we must have $L(x'') = V_i'(x'')$, which indicates $L'(x'') = 0$. Suppose $x' < x''$. Concavity requires that $V_i'(x) > V_i'(x'') = L(x'') >$

$L(x)$ on (x', x''), which is a contradiction. Hence, $x' \geq x''$. Now, suppose $\exists x_0 \in (x'', K)$ such that $L(x_0) > V_i'(x_0)$. We consider two cases: (I) $x' \leq x_0$ and (II) $x' > x_0$. In case (I), $V_i'(x)$ decreases for all $x > x_0$ by concavity. However, $L(x)$ increases as long as $L(x) > V_i'(x)$. In order to have $\lim_{x \to K} L(x) = V_i'(K)$, we need that $L(x)$ decreases while $L(x) > V_i'(x)$, which is a contradiction. In case (II), we have that $x'' < x_0 < x'$ and $V_i'(x)$ increases for all $x \leq x_0$ by convexity. However, $L'(x'') = 0 < V_i''(x'')$, so by continuity there exist $\epsilon > 0$ such that

$$\frac{L(x'' + \epsilon) - L(x'')}{\epsilon} < \frac{V_i'(x'' + \epsilon) - V_i'(x'')}{\epsilon}.$$

Therefore, in order to have $L(x_0) > V_i'(x_0)$, we need $L(x)$ to increase while $L(x) < V_i'(x)$, which is a contradiction. As a result, $L(x) \leq V_i'(x)$ for all $x \in [x', K)$. Suppose $L(x) = V_i'(x)$ for all $x \in [x', K)$. We then have $L(x) = \lim_{x \to K} L(x) = V_i'(K)$ for all $x \in [x', K)$, but $V_i'(x) > V_i'(K)$ for some $x \in [x', K)$ by concavity, which is a contradiction. The result follows. ■

Proof of Theorem 1. Part I. We begin with the last (i.e., the M-th) buyer. Given that no more selling opportunities will exist after the last buyer, we have $V_M(K) = sK$. From Lemma 1, we know that the price for the M-th buyer is $w_M = p$ or $w_M = \beta_M \cdot s + (1 - \beta_M) \cdot \bar{c}_M$. Thus, we have $w_M > s$ and the seller should sell as much to the last buyer as possible; i.e., $\rho_M^* = \min\{1, K/D_M\}$. Plugging $V_M(K) = sK$ and $Q_M = K \wedge D_M$ into (12), we can obtain

$$V_{M-1}(K) = sK + (1 - \delta_M) \cdot \nu(p, \psi_{M-1}) \cdot \left[\int_0^K D dF(D|\psi_{M-1}) + \int_K^{K/\eta} K dF(D|\psi_{M-1}) \right]. \quad (20)$$

Although in general $V_{i-1}(K)$ in (12) is not a separable function for the three random variables, β, $\bar{c}$, and D, we find in (20) that D can be multiplicatively separated from β and $\bar{c}$. Basically, $\nu(p, \psi_{M-1})$ is only related to β and $\bar{c}$, and it measures the expected margin obtained from a buyer above the salvage value. Note that $\nu(p, \psi_{M-1})$ is finite because $\beta \in [0, 1]$ and $\bar{c} \in [c_L, p]$. Furthermore, under the assumption of finite demand expectation, we have that $\int_K^{K/\eta} K dF(D|\psi_{M-1})$ approaches zero as K increases, so this term can be ignored when $K \to +\infty$.

Lemma 2. $\lim_{K \to +\infty} \int_K^{K/\eta} K dF(D|\psi) = 0$ *given that* $\int_0^{+\infty} D dF(D|\psi) < +\infty$.

Proof. First, we have

$$\int_0^{+\infty} DdF\left(D|\psi\right) = \int_0^{K} DdF\left(D|\psi\right) + \int_K^{K/\eta} DdF\left(D|\psi\right) + \int_{K/\eta}^{+\infty} DdF\left(D|\psi\right) < \infty$$

for any $K > 0$. Second, $\int_0^{+\infty} DdF\left(D|\psi\right) = \lim_{K\to\infty}\int_0^{K} DdF\left(D|\psi\right)$. Thus,

$$\lim_{K\to\infty}\int_K^{K/\eta} DdF\left(D|\psi\right) = \lim_{K\to\infty}\int_{K/\eta}^{+\infty} DdF\left(D|\psi\right) = 0.$$

Furthermore, $\int_K^{K/\eta} KdF\left(D|\psi\right) \leq \int_K^{K/\eta} DdF\left(D|\psi\right)$, so $\lim_{K\to\infty}\int_K^{K/\eta} KdF\left(D|\psi\right) = 0$. □

Therefore, based on (20), for any scalar $\lambda \in (0,1)$, we can easily have a lower bound $\hat{V}_{M-1}\left(K,\lambda\right)$ for $V_{M-1}\left(K\right)$:

$$V_{M-1}\left(K\right) \geq sK + \left(1-\delta_M\right)\cdot \nu(p,\psi_{M-1})\cdot \int_0^{\lambda K} DdF\left(D|\psi_{M-1}\right) = \hat{V}_{M-1}\left(K,\lambda\right). \tag{21}$$

We can see that the shape of $\hat{V}_{M-1}\left(K,\lambda\right)$ (i.e., how $\hat{V}_{M-1}\left(K\right)$ changes with K) depends on the shape of distribution $F\left(\cdot|\psi_{M-1}\right)$. If $F\left(\cdot|\psi_{M-1}\right)$ is exponential, then $\hat{V}_{M-1}\left(K,\lambda\right)$ is increasing and concave in K. If $F\left(\cdot|\psi_{M-1}\right)$ is normal, then $\hat{V}_{M-1}\left(K,\lambda\right)$ is S-shaped. Using the result for the last buyer, we move on and consider $V_{i-1}\left(K\right)$ in general.

Part II. We now show the proof for the lower bound. To prove by induction, we suppose for $i < M$ that

$$V_i\left(K\right) \geq sK + \left(1-\delta_{i+1}^l\right)\cdot \nu(p,\psi_i)\cdot \int_0^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_{i-1}\right).$$

Now, using (12), we have

$$\begin{aligned} V_{i-1}\left(K\right) \geq & sK + \left(1-\delta_i\right)\cdot \\ & \left\{\left(1-\delta_{i+1}^l\right)\cdot \mathbf{E}_{\beta,\bar{c}}\left[\nu(p,\psi_i)|\psi_{i-1}\right]\cdot \int_{K/\eta}^{+\infty}\int_0^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right)\right. \\ & + \left(1-\delta_{i+1}^l\right)\cdot \mathbf{E}_{\beta,\bar{c}}\left[\nu(p,\psi_i)|\psi_{i-1}\right]\cdot \int_0^{K/\eta}\int_0^{(K-D')^+\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right) \\ & \left. + \nu(p,\psi_{i-1})\cdot\int_0^{K/\eta} D'dF\left(D'|\psi_{i-1}\right)\right\} \end{aligned}$$

$$
\begin{aligned}
=&sK + (1-\delta_i)\cdot\nu(p,\psi_{i-1})\cdot\Bigg\{\left(1-\delta^l_{i+1}\right)\cdot\int_0^{+\infty}\int_0^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right)\\
&-\left(1-\delta^l_{i+1}\right)\cdot\int_0^{K/\eta}\int_0^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right)\\
&+\left(1-\delta^l_{i+1}\right)\cdot\int_0^{K/\eta}\int_0^{(K-D')^+\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right)\\
&+\int_0^{K/\eta} D'dF\left(D'|\psi_{i-1}\right)\Bigg\}\\
=&sK + (1-\delta_i)\cdot\nu(p,\psi_{i-1})\cdot\Bigg\{\left(1-\delta^l_{i+1}\right)\cdot\int_0^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_{i-1}\right)\\
&-\left(1-\delta^l_{i+1}\right)\cdot\int_0^{K/\eta}\int_{(K-D')^+\cdot\lambda^{M-i}}^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right)\\
&+\int_0^{K/\eta} D'dF\left(D'|\psi_{i-1}\right)\Bigg\}.
\end{aligned}
$$

Here for the first equality, we add and subtract $\left(1-\delta^l_{i+1}\right)\cdot\int_0^{K/\eta}\int_0^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right)$ in the curly braces. Further, we have

$$
\begin{aligned}
&\int_0^{K/\eta}\int_{(K-D')^+\cdot\lambda^{M-i}}^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right) \le \int_0^{+\infty}\int_{(K-D')^+\cdot\lambda^{M-i}}^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_i\right)dF\left(D'|\psi_{i-1}\right)\\
&= \mathbf{E}\left[\mathbf{E}\left[D\cdot\mathbb{I}\left\{(K-D')^+\cdot\lambda^{M-i}\le D\le K\cdot\lambda^{M-i}\right\}|\psi_i\right]|\psi_{i-1}\right]\\
&= \mathbf{E}\left[D\cdot\mathbb{I}\left\{(K-D')^+\cdot\lambda^{M-i}\le D\le K\cdot\lambda^{M-i}\right\}|\psi_{i-1}\right]\\
&= \int_0^{+\infty}\int_{(K-D')^+\cdot\lambda^{M-i}}^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_{i-1}\right)dF\left(D'|\psi_{i-1}\right)\\
&= \int_0^{K\cdot\lambda^{M-i}}\int_{K-D/\lambda^{M-i}}^{+\infty} DdF\left(D'|\psi_{i-1}\right)dF\left(D|\psi_{i-1}\right)\\
&= \int_0^{K\cdot\lambda^{M-i}} D\cdot\left[1-F\left(K-D/\lambda^{M-i}|\psi_{i-1}\right)\right]dF\left(D|\psi_{i-1}\right)\\
&\le \int_0^{K\cdot\lambda^{M-i+1}} D\cdot\left[1-F\left(K-D/\lambda^{M-i}|\psi_{i-1}\right)\right]dF\left(D|\psi_{i-1}\right) + \int_{K\cdot\lambda^{M-i+1}}^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_{i-1}\right)\\
&\le \left[1-F\left(K-\lambda K|\psi_{i-1}\right)\right]\cdot\int_0^{K\cdot\lambda^{M-i+1}} DdF\left(D|\psi_{i-1}\right) + \int_{K\cdot\lambda^{M-i+1}}^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_{i-1}\right)\\
&\le \left[1-F_0\left(K-\lambda K\right)\right]\cdot\int_0^{K\cdot\lambda^{M-i+1}} DdF\left(D|\psi_{i-1}\right) + \int_{K\cdot\lambda^{M-i+1}}^{K\cdot\lambda^{M-i}} DdF\left(D|\psi_{i-1}\right).
\end{aligned}
$$

Here, we first extend the range of integral for D' to $[0,+\infty)$ given that D and F are both positive.

Next, we rewrite the double integral as iterated expectations. Third, we apply the law of iterated expectations. Fourth, we write the expectation as a double integral again. Fifth, we change the sequence of integral and then simplify the expression in the next step. Sixth, we split the integral into two parts and apply $F\left(K - D/\lambda^{M-i}|\psi_{i-1}\right) \le 1$ for the part from $K \cdot \lambda^{M-i+1}$ to $K \cdot \lambda^{M-i}$. Seventh, given $0 \le D \le K \cdot \lambda^{M-i+1}$, we have $F\left(K - \lambda K|\psi_{i-1}\right) \le F\left(K - D/\lambda^{M-i}|\psi_{i-1}\right)$. Finally, by assumption, we have $F_0\left(K - \lambda K\right) \le F\left(K - \lambda K|\psi_{i-1}\right)$. Now we can write

$$\begin{aligned}
V_{i-1}(K) \ge & sK + (1-\delta_i) \cdot \nu(p, \psi_{i-1}) \cdot \left\{ \left(1 - \delta_{i+1}^l\right) \cdot \int_0^{K \cdot \lambda^{M-i}} DdF\left(D|\psi_{i-1}\right) \right. \\
& - \left(1 - \delta_{i+1}^l\right) \cdot \int_{K \cdot \lambda^{M-i+1}}^{K \cdot \lambda^{M-i}} DdF\left(D|\psi_{i-1}\right) \\
& - \left(1 - \delta_{i+1}^l\right) \cdot \left[1 - F_0\left(K - \lambda K\right)\right] \cdot \int_0^{K \cdot \lambda^{M-i+1}} DdF\left(D|\psi_{i-1}\right) \\
& \left. + \int_0^{K/\eta} DdF\left(D|\psi_{i-1}\right) \right\} \\
= & sK + (1-\delta_i) \cdot \nu(p, \psi_{i-1}) \cdot \left\{ \left(1 - \delta_{i+1}^l\right) \cdot F_0\left(K - \lambda K\right) \cdot \int_0^{K \cdot \lambda^{M-i+1}} DdF\left(D|\psi_{i-1}\right) \right. \\
& \left. + \int_0^{K/\eta} DdF\left(D|\psi_{i-1}\right) \right\} \\
\ge & sK + \left[1 - \delta_i + (1-\delta_i) \cdot \left(1 - \delta_{i+1}^l\right) \cdot F_0\left(K - \lambda K\right)\right] \cdot \nu(p, \psi_{i-1}) \cdot \int_0^{K \cdot \lambda^{M-i+1}} DdF\left(D|\psi_{i-1}\right) \\
= & sK + \left(1 - \delta_i^l\right) \cdot \nu(p, \psi_{i-1}) \cdot \int_0^{K \cdot \lambda^{M+1-i}} DdF\left(D|\psi_{i-1}\right).
\end{aligned}$$

Part III. We now show the proof for the upper bound. First, for $i = M$, we easily have

$$V_{M-1}(K) \le sK + (1-\delta_M) \cdot \nu(p, \psi_{M-1}) \cdot \int_0^{K/\eta} DdF\left(D|\psi_{M-1}\right)$$

according to (20). Then suppose we have the upper bound for $V_i(K)$: $sK + \bar{V}_i(K)$. Accordingly, we can use this in (12) to obtain

$$
\begin{aligned}
V_{i-1}(K) \leq \delta_i sK + (1-\delta_i)\cdot & \\
\Bigg\{ \int_0^1 \int_{K/\eta}^{+\infty} \int_{c_L}^{+\infty} & \left[sK + \bar{V}_i(K)\right] dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \\
+ \int_0^1 \int_0^{K/\eta} \int_p^{+\infty} & \left[p\rho^* D + s(K-\rho^* D) + \bar{V}_i(K)\right] dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \\
+ \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} & \left[sK + \bar{V}_i(K) - (1-\beta)\, s\rho^* D\right] dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \\
+ \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} & (1-\beta)\, \bar{c}\rho^* D dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \Bigg\} \\
&= sK + (1-\delta_i)\cdot \left\{ \mathbf{E}_{\bar{c},D,\beta}\left[\bar{V}_i(K)\,|\psi_{i-1}\right] \right. \\
+ \int_0^1 \int_0^{K/\eta} \int_p^{+\infty} & (p-s)\rho^* D dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \\
+ \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} & \left. (1-\beta)(\bar{c}-s)\rho^* D dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \right\} \\
&\leq sK + (1-\delta_i)\cdot \left\{ \mathbf{E}_{\bar{c},D,\beta}\left[\bar{V}_i(K)\,|\psi_{i-1}\right] + \nu(p,\psi_{i-1})\cdot \int_0^{K/\eta} D dF(D|\psi_{i-1}) \right\} \\
&= sK + \left(1-\delta_i + (1-\delta_i)\left(1-\delta_{i+1}^u\right)\right)\cdot \nu(p,\psi_{i-1})\cdot \int_0^{K/\eta} D dF(D|\psi_{i-1}).
\end{aligned}
$$

Note that we use the fact that $\int_0^{(K-\rho^* D)/\eta} D dF(D|\psi_i) \leq \int_0^{K/\eta} D dF(D|\psi_i)$ for the first inequality. We use that $\rho^* \leq 1$ for the second inequality given that $p > s$ and $\bar{c} > s$.■

Proof of Theorem 2. Let $G_{i-1}(\bar{c}) = G(\bar{c}|\psi_{i-1})$, $F_{i-1}(D) = F(D|\psi_{i-1})$, and $B_{i-1}(\beta) = B(\beta|\psi_{i-1})$. We first divide both sides of Eq. (12) by $1-\delta$ and add to the right side

$$
\begin{aligned}
0 = \int_0^{K/\eta} V_i(K)\, dF_{i-1}(D) & \\
& - \int_0^1 \int_0^{K/\eta} \int_p^{+\infty} V_i(K)\, dG_{i-1}(\bar{c})\, dF_{i-1}(D) dB_{i-1}(\beta) \\
& \qquad - \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} V_i(K)\, dG_{i-1}(\bar{c})\, dF_{i-1}(D) dB_{i-1}(\beta).
\end{aligned}
$$

Accordingly, we obtain

$$\begin{aligned}
\frac{V_{i-1}(K)}{1-\delta} &= \frac{\delta}{1-\delta} \cdot s \cdot K + \mathbf{E}\left[V_i(K)|\psi_{i-1}\right] \\
&+ \int_0^1 \int_0^{K/\eta} \int_p^{+\infty} \left[V_i(K-\rho^* D) - V_i(K)\right] dG_{i-1}(\bar{c})\, dF_{i-1}(D) dB_{i-1}(\beta) \\
&+ \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} (1-\beta)\left[V_i(K-\rho^* D) - V_i(K)\right] dG_{i-1}(\bar{c})\, dF_{i-1}(D) dB_{i-1}(\beta) \\
&+ p \int_0^1 \int_0^{K/\eta} \int_p^{+\infty} \rho^* D dG_{i-1}(\bar{c})\, dF_{i-1}(D) dB_{i-1}(\beta) \\
&+ \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} (1-\beta)\, \bar{c} \rho^* D dG_{i-1}(\bar{c})\, dF_{i-1}(D) dB_{i-1}(\beta). \qquad (22)
\end{aligned}$$

Because $\mathbf{E}\left[V_i(K)|\psi_{i-1}\right] = V_{i-1}(K)$, $V_i(K-\rho^* D) \leq V_i(K)$, and $\rho^* \leq 1$, we have

$$\frac{\delta}{1-\delta} \cdot \left[V_{i-1}(K) - s \cdot K\right] \leq 0 + 0 + H_i(K),$$

which results in $V_{i-1}(K) \leq s \cdot K + \frac{1-\delta}{\delta} \cdot H_i(K)$.

To derive the lower bound, we start from (22) and use the optimality of ρ^*. We have

$$\begin{aligned}
&\int_0^1 \int_0^{K/\eta} \int_p^{+\infty} \left[p\rho^* D + V_i(K-\rho^* D)\right] dG_{i-1}(\bar{c})\, dF_{i-1}(D) dB_{i-1}(\beta) \\
\geq &\int_0^1 \int_0^{K/\eta} \int_p^{+\infty} \left[pD + V_i\left([K-D]^+\right)\right] dG_{i-1}(\bar{c})\, dF_{i-1}(D) dB_{i-1}(\beta).
\end{aligned}$$

Similarly, we apply this logic to the case of $\bar{c}_i \leq p$. As a result, we get

$$\begin{aligned}
\frac{\delta}{1-\delta} \cdot \left[V_{i-1}(K) - s \cdot K\right] &\geq (1 - \mathbf{E}\left[\beta|\psi_{i-1}\right] \cdot G_{i-1}(p)) \cdot \int_0^{K/\eta} \left[V_i\left([K-D]^+\right) - V_i(K)\right] dF_{i-1}(D) \\
&\quad + H_i(K) \\
&\geq H_i(K) - (1 - \mathbf{E}\left[\beta|\psi_{i-1}\right] \cdot G_{i-1}(p)) \cdot \left[1 - h_i(K)\right] V_{i-1}(K).
\end{aligned}$$

For the last inequality above, we use the fact that $V_i\left([K-D_i]^+\right) \leq V_i(K)$ and

$$\begin{aligned}
&\int_0^{+\infty} \left[V_{i-1}(K) - V_i\left([K-D]^+\right)\right] dF_{i-1}(D) \\
\geq &\int_0^{K/\eta} \left[V_{i-1}(K) - V_i\left([K-D]^+\right)\right] dF_{i-1}(D).
\end{aligned}$$

Hence, we have $V_{i-1}(K) \geq \left(sK + \frac{1-\delta}{\delta} \cdot H_i(K)\right) / \left(1 + \frac{1-\delta}{\delta} \cdot (1 - \mathbf{E}[\beta|\psi_{i-1}] \cdot G_{i-1}(p)) \cdot [1 - h_i(K)]\right)$.

To show $\lim_{K\to\infty} h_i(K) = 1$, we need to check two cases: $s = 0$ and $s \neq 0$. If $s = 0$, then we know from the upper bound that $V_i(K)$ is bounded by $H_{i+1}(K)$, which is bounded as $K \to \infty$. In this case, we can show that from the ψ_{i-1} point of view, $V_i\left([K - D_i]^+\right)$ converges in probability to $V_i(K)$ as $K \to \infty$. To this end, note that given any ψ_i both $V_i\left([K - D_i]^+\right)$ and $V_i(K)$ are increasing in K but are bounded. Hence, they converge to the same limit $\bar{C}(\psi_i)$, and for any $\epsilon > 0$, there exist $K(\psi_i) < \infty$ such that $|V_i(K) - V_i\left([K - D_i]^+\right)| < \epsilon$. Because $K(\psi_i)$ is finite, there exist $K^*_{\epsilon,\xi} < \infty$ for any $\xi > 0$ such that $\Pr\left\{K^*_{\epsilon,\xi} < K(\psi_i)\,|\psi_{i-1}\right\} < \xi$. In other words, for any $\epsilon > 0$ and $\xi > 0$, there exist $K^*_{\epsilon,\xi} < \infty$ such that for $K \geq K^*_{\epsilon,\xi}$ we have

$$\Pr\left\{V_i(K) - V_i\left([K - D_i]^+\right) > \epsilon|\psi_{i-1}\right\} < \xi.$$

Therefore, $V_i\left([K - D_i]^+\right)$ converges in probability to $V_i(K)$ and thus

$$\mathbf{E}\left[V_i\left([K - D_i]^+\right)|\psi_{i-1}\right] \to \mathbf{E}\left[V_i(K)\,|\psi_{i-1}\right].$$

If $s \neq 0$, then $V_i(K)$ is unbounded. However, we know from the upper bound that $V_i(K) - sK$ is bounded by $H_{i+1}(K)$. Because $\frac{\partial}{\partial K} V_i(K) \geq s$, we have that $V_i(K) - sK$ is increasing in K. As a result, $\tilde{V}_i(K) = V_i(K) - sK$ converges to a limit. Applying the same logic as for the case of $s = 0$, we know that $\mathbf{E}\left[\tilde{V}_i\left([K - D_i]^+\right)|\psi_{i-1}\right] \to \mathbf{E}\left[\tilde{V}_i(K)\,|\psi_{i-1}\right]$. Accordingly, we have

$$\mathbf{E}\left[V_i(K) - sK - V_i\left([K - D_i]^+\right) + s\,[K - D_i]^+\,|\psi_{i-1}\right] \to 0.$$

Since $\mathbf{E}\left[sK - s\,[K - D_i]^+\,|\psi_{i-1}\right] = \mathbf{E}\left[s \cdot \min\{K, D_i\}\,|\psi_{i-1}\right] \to \mathbf{E}\left[sD_i|\psi_{i-1}\right] < \infty$, we know that $\mathbf{E}\left[V_i(K) - V_i\left([K - D_i]^+\right)|\psi_{i-1}\right] \to \mathbf{E}\left[sD_i|\psi_{i-1}\right]$. Therefore,

$$\begin{aligned} h_i(K) &= \frac{\mathbf{E}\left[V_i\left([K - D_i]^+\right)|\psi_{i-1}\right]}{\mathbf{E}\left[V_i(K)\,|\psi_{i-1}\right]} \\ &= 1 - \frac{\mathbf{E}\left[V_i(K) - V_i\left([K - D_i]^+\right)|\psi_{i-1}\right]}{\mathbf{E}\left[V_i(K)\,|\psi_{i-1}\right]} \\ &\to 1. \blacksquare \end{aligned}$$

Proof of Proposition 3. Let the probability density function be $f(x) = a \cdot \exp\left(-\frac{(x-b)^2}{2c}\right)$. Note that the critical component in all the bounds is $\int_0^{K/\eta} D_i dF\left(D_i|\psi_{i-1}\right)$ and $\int_0^{K} D_i dF\left(D_i|\psi_{i-1}\right)$. Without a loss of generality, we focus on $A(K) = \int_0^K x dF\left(x\right)$. The second-order condition gives $\frac{\partial^2 A}{\partial K^2} = f(K) + K \cdot f'(K)$. It is easy to check that $f'(x) = -\frac{x-b}{c} \cdot f(x)$. Hence, we have $\frac{\partial^2 A}{\partial K^2} = f(K) \cdot \left[1 - \frac{K(K-b)}{c}\right]$, which has zero points $K^*_{1,2} = \frac{b \pm \sqrt{b^2+4c}}{2}$. It is clear that only one non-negative zero point exists because $c > 0$. Therefore, $A(K)$ is convex-concave. Given sK is linear, we know that the bounds are all convex-concave. ■

Proof of Proposition 4. Note that $\Pi_A = w_i\left(Q_i\right) \cdot Q_i + V_i\left(K - Q_i\right)$, where $w_i\left(Q_i\right)$ is given by (5). Thus, Π_A can be written as

$$\beta_i\left[V_i\left(K\right) - V_i\left(K - \rho_i D_i\right)\right] + \left(1-\beta_i\right) D_i \left[r_i \rho_i - \rho_i'\left(r_i' - \tilde{c}_i\right)\right] + V_i\left(K - \rho_i D_i\right).$$

Taking the first-order derivative of Π_A with respect to ρ_i, we get

$$\partial \Pi_A / \partial \rho_i = \left(1-\beta_i\right) D_i \left[r_i - V_i'\left(K - \rho_i D_i\right)\right].$$

The result follows. ■

References

[1] Greene, W. H. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

[2] Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50(4) pp.1029–1054.