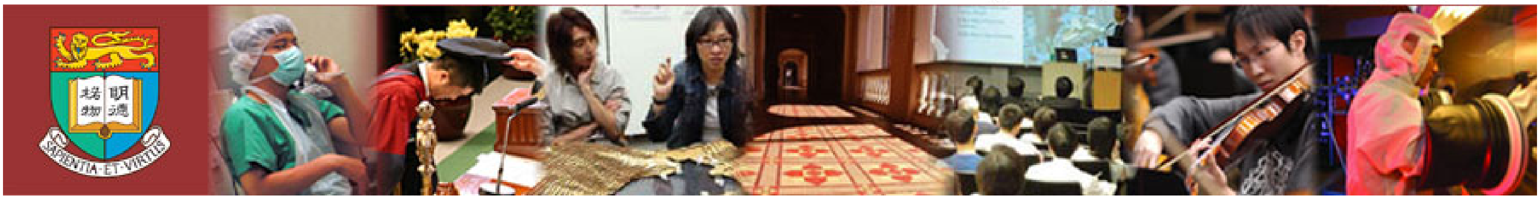


Title	**Higher Prices for Larger Quantities? Non-Monotonic Price-Quantity Relations in B2B Markets**
Author(s)	**Zhang, W; Dasu, S; Ahmadi, R**
Citation	**Management Science, 2016, v. 63 n. 7, p. 2108-2126**
Issued Date	**2016**
URL	**http://hdl.handle.net/10722/227474**
Rights	**This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.**

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. Instead, the discount curve is like an "N". 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 often too much for small buyers but not enough for large buyers. We briefly discuss the implications for capacity rationing and posted pricing as well as potential applications to other industries.

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

1 Introduction

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. We interacted with managers of a large semiconductor company and obtained a sales data set that spans a three-year period. In this data, we found several instances wherein larger-quantity buyers pay higher prices. Five such examples are illustrated in Table 1. Buyers were sorted into three groups—small, medium, and large—according to their total purchase quantities. Interestingly, the average price received by the medium-quantity group is less than the prices obtained by the other two groups; moreover, the large-quantity group received the highest average price in four of the five examples. The data set had many such instances in which larger buyers paid higher prices. 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 paid a higher average price than a neighboring, smaller-quantity buyer. These observations point to non-monotonic price patterns.

Our task was threefold in this research. First, we empirically tested whether a non-monotonic price-quantity relation generally exists. Second, we explored theoretical explanations for this phenomenon and finally, we determined the managerial implications that flow from such a pricing pattern. In the first step, we used a set of linear and nonlinear regressions to control for other possible influences on product price and found that the discount received by a buyer is statistically a non-monotonic function of the buyer's demand share (or relative size) for the product. Specifically, the discount increases with demand share for small quantities; however, as demand share increases,

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

the discount decreases and then increases again. In brief, we observed an N-shaped discount curve, which cannot be explained by the existing literature.

In the second step, we developed an analytical model based on the practices of the focal company for which the product life cycle is short, capacity is inflexible, and prices are set through *one-shot* negotiations. We showed that our model fits the data and can yield the price-quantity curves found in the data. Our model suggests that the non-monotonic price-quantity relation is rooted in how the seller values her 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 may be too much for small buyers but not enough for large buyers. Therefore, mid-sized buyers are charged a "premium."

Several managerial insights flow from our analytical model. Basically, given that each transaction has an impact on subsequent transactions, a good model of the price-quantity relation is necessary for the optimization of the trade-off between the profit from the current buyer and that of future buyers. Our analysis shows that it is important for the seller to control the capacity allocated to each buyer prior to price negotiations, if possible. Our model can also help the seller optimize the posted price, which balances the profit between buyers who take the price and buyers who choose to bargain.

The rest of this paper is organized as follows. 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 findings in Section 7 and conclude in Section 8. All of the proofs are in the electronic companion.[1]

2 Related Literature

Our work is related to four areas of research, the first of which concerns quantity-discount policies. In the operations manangement and marketing literature, quantity-discount models have been widely studied (e.g., Spence 1977 and Oren et al. 1982, Jeuland and Shugan 1983, and Weng

[1]The e-companion is available at https://drive.google.com/file/d/0ByHSOpfhKdCQdTdkNUZ1NVV2Mlk/.

1995). In this research stream, it is assumed that one party will offer a 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.

The second related area is B2B price bargaining. Kahli and Park (1989) analyzed a two-party bargaining problem that is based on a economic order quantity (EOQ) model and showed that the optimal discount increases with purchase quantity. 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, 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. Chipty and Snyder (1999) showed that a merger enhances (worsens) a buyer's bargaining position if the supplier's payoff function is concave (convex) in total transaction size. Our data exhibits a more complicated price pattern than what are predicted by these models.

Other studies on B2B bargaining have assumed a given size of the pie 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 who sequentially negotiates with a group of buyers and we investigate the impact of a buyer's purchase quantity on the discount.

Our research is also related to revenue management. Kuo et al. (2011) first studied revenue management for limited inventories when buyers negotiate and considered a dynamic setting with fixed compositions of price-takers and bargainers with the assumption that each buyer only buys one unit of the product and 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, and 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 closely related to research on dynamic and stochastic knapsack problems that study optimal admission or pricing policies when capacity is limited. Talluri and van Ryzin (2004) categorize these type of problems as revenue management with group arrivals. 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 each buyer requires the same amount, Kleywegt and Papastavrou (2001) showed that concavity does not hold in general when demands are heterogeneous. 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) provided a good review of this literature. Karabuk and Wu (2003) studied strategic capacity planning in the presence of demand and capacity uncertainties. Cohen et al. (2003) proposed a model to estimate the imputed costs of an equipment supplier. Terwiesch et al. (2005) empirically studied the demand forecast sharing process between a buyer and a set of equipment suppliers. Karabuk and Wu (2005) studied incentive issues when product managers compete for capacity. Peng et al. (2012) worked with Intel and developed an equipment procurement framework that allowed Intel to make a combination of base and flexible capacity reservations with suppliers. Different from all of these studies, the current paper helps semiconductor companies understand the value of their capacities when customers arrive sequentially and differ in order quantities.

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 frequent 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 such as Intel manufacture products in-house, while others such as AMD only focus on product design and outsource production to third-party foundries. In both cases, because 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). A common practice in this industry is for sellers to allocate capacity to product lines based on demand forecasts and start production several month before demand is realized. The forecasts represent sales commitments from product line managers and sometimes (non-binding) purchase commitments from customers. Once the capacity is allocated to a product, production starts almost immediately. In this way, sellers can first increase 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 yield to ramp up and for quality to improve. Thus, when facing a supply shortage, it is not only costly but also risky for sellers to seek an alternative source. Given these facts, it is important for sellers' product managers to provide accurate forecasts and 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 that 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 contract duration can vary for different buyers and products; price renegotiations happen frequently but not in all cases.[2] In our data set, approximately 45% of the purchases did not involve any renegotiation.

[2]Companies may use different types of contracts in terms of price flexibility that differ in terms of how frequently and to what extent prices are renegotiated. Price renegotiation happens when a contracted price expires.

Procurement Quantity. The purchase quantity is not normally a term for negotiation. In principle, although buyers can allocate their requirements among alternative semiconductor firms, they do so at a very early stage in the planning process because products offered by different sellers differ in technical features and influence the design of the buyer's products.[3] In addition, sellers' brand images may matter in buyers' markets. To produce their products, buyers need many different inputs, and the composition of the inputs are interdependent. 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 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. Contracts, however, do not stipulate any purchase commitments from the buyers.

Technology Upgrades. In this industry, the risk of obsolescence is a constant. Sellers are aware that technological advancements 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, sellers must 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 sales of 3,826 products to 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. Products sold include central processing units (CPUs), graphics processing units (GPUs), and embedded chips, among others.

[3]For example, Apple Inc. allocated about one-third of the A9 processor orders to the 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/.

4.1 Data Preparation

4.1.1 Fixed-Price Contracts

As buyer normally purchase a product through multiple transactions over time, the prices may be renegotiated. In this paper, however, 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., $\theta_{ij} := \{(t, q_{ijt}, p_{ijt}) : t \in \mathcal{T}_{ij}\}$. For fixed-price contracts, we have $p_{ijt} = p_{ij}$ for all $t \in \mathcal{T}_{ij}$. As previously stated, in this industry it is a common practice for a buyer to enter a price negotiation with a pre-determined target quantity. The life-cycle purchase quantity is the target. 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}.

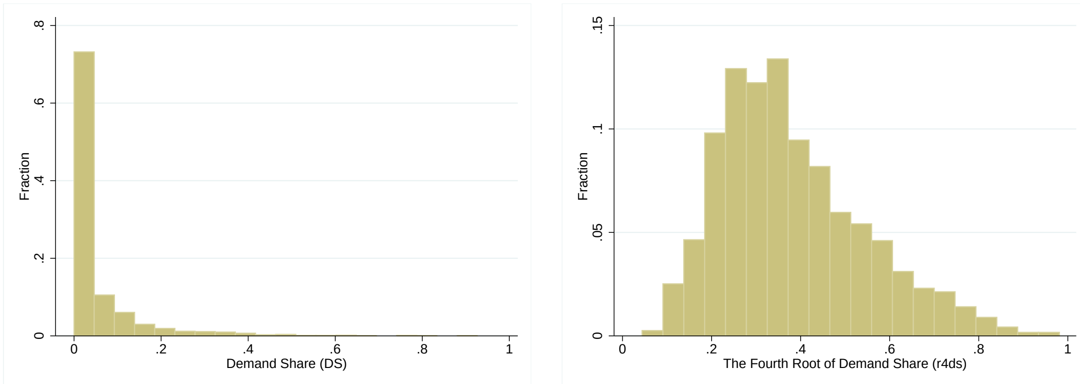
4.1.2 Normalization

Widely varying prices and market sizes for different products compel us to normalize the data. Prices of the 425 brands (or product subcategories) range from several dollars to more than \$100 per unit. In place of the price and total quantity, we use two ratio metrics: (1) effective discount (ED) and (2) demand share (DS). Define

$$\begin{aligned} ED_{ij} &:= 1 - p_{ij} / \max_{i' \in I, t \in \mathcal{T}} p_{i'jt}; && (1) \\ DS_{ij} &:= TQ_{ij} / \sum_{i' \in I} TQ_{i'j}, && (2) \end{aligned}$$

where $\mathcal{T} := \cup \mathcal{T}_{ij}$. Both variables have a range of $[0, 1]$. ED is a measure of price level relative to the highest price ever paid for the product. It is important to note that, in the case of the seller who provided us this data set, posted prices are very stable. As the vast majority of the posted prices remain unchanged for more than a year, for a fixed-price contract the nominal discount rate is almost constant over time and the ED is very close or equal to the nominal discount rate.

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



DS is the ratio of the total quantity purchased by a buyer relative to the market size for that product. The advantage of ED and DS is that they simultaneously control for the mean and variation across different products.[4] However, in the vast majority of instances, demand shares are concentrated around zero (as shown in Figure 1 on the left) due to the 20–80 rule; i.e., 80% of customers contribute only 20% of sales. Therefore, it is difficult to examine how quantity affects price discount if we use DS. To circumvent this limitation, we use a power transformation of demand share ($PTDS$); i.e., $PTDS = DS^{\gamma}$ 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 has a larger spread relative to DS (as shown in Figure 1 on the right) and third, the range of $[0, 1]$ is preserved. As shown in Table 2, the distribution has the highest spread when γ is between 0.25 and 0.35, and $\gamma = 0.25$ yields a distribution that is the closest to normal in this range.[5] We focus on the fourth-root transformation (i.e., $\gamma = 0.25$). Later, to check robustness, we consider the following values for γ: 0.15, 0.2, 0.3, and 0.35. Our objective is now reduced to identifying the relationship between ED_{ij} and $PTDS_{ij}$ while controlling for other factors.

4.1.3 Data Filtering

The data set corresponds to a three-year 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

[4]Note that the way a customer's demand share 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.

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

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.

data (or are truncated). Because prices in the microprocessor market decrease over time, instances with left-truncation appear to have smaller total quantities and higher prices than subsequent instances. Similarly, instances with right-truncation appear to have smaller total quantities and lower prices than preceding instances. Mixing these two effects may generate a non-monotonic relationship between price and quantity. To mitigate the effects of truncation, we only consider instances with an observed starting date that is at least one quarter after January 1, 2009, and an observed ending date that is at least one quarter earlier than March 25, 2012. There are 6,573 instances (about 53%) that satisfy these criteria. Furthermore, to focus on regular purchases and not transactions for one-time substitutions or sample testing—that follow a different selling process—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. Such products also tend to have buyers with very high demand-shares and narrow price dispersions and as a result, our proxies for these products' posted prices could be downward-biased. To avoid this bias, we drop another 1,551 instances (about 18%) and consider only products that have more than three buyers. In this way, we obtain a subset of the 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 the respective bargaining powers of both buyer and seller. As far as we can imagine, the seller's outside

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

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, 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.

TSQ. The total sales quantity of a product. This variable also captures the popularity of a product.

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

lndod. Because, in general, effective prices (or price-performance ratios) in the semiconductor industry decrease over time, the later a buyer arrives the greater the discount (relative to the posted price) he may obtain. To capture this effect, we use the logarithm of *days of delay*, which is calculated as the difference in the number of days between the starting date of an instance and the first date that the product was ever purchased. In addition, *lndod* is also a measure of demand uncertainty because uncertainty is resolved over time. Note that the introduction date of a product is unobserved, but the starting date of the first purchase is normally a good proxy.

lndrt. The discount may also be related to the rate of purchase. We control for this effect by the logarithm of the *duration* of an instance. Duration of an instance is given by the number of days between the first and the last date of that instance. We note parenthetically that the duration of an instance in our data set is fairly short, with an average of 232 days.

M3. It is well known that the *end-of-quarter effect* has a bearing on bargaining. We introduce *M3* as a binary variable for an instance, with a value equal to 1 if the date of the first order is in the third month of a quarter and 0 otherwise.

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

Cshr. A buyer that accounts for a large portion of the seller's business can have significant bargaining power. To capture this aspect of the bargaining power we use the measure *Cshr* (customer share). *Cshr* is 100 times the total quantity purchased by a buyer divided by the seller's total sales volume across all products during the time period being studied. It measures the value of a buyer to the seller.[7] Note that using the total purchase value as a control variable may induce endogeneity as the value depends on the price discount.

Vrate. Since we use the highest price paid for a product as an approximation of the posted price, small variations in the price of a product will result in low *ED* and thus we have to also control for price variations. However, prices depend on discounts. To avoid this endogeneity, we use the *fraction of instance in which the price was renegotiated* to measure a product's price variation. Buyers with renegotiable-price contracts are more likely to get the posted price at the beginning and experience price variations that are larger than those with fixed-price contracts.

[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 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 seller-side competition, production costs, and salvage values, we use binary variables for the 16 major brands that have at least 100 observed instances in the original data set. It is also important to note that the seller's salespeople are organized by product lines. Thus, the impact of a salesperson's bargaining ability is product-line-specific and can be captured by the product-line fixed effect.

Buyer fixed effect. 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 buyer fixed effects for the ten major buyers ranked on the basis of total purchase value. The remaining buyers serve as a 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 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 for cost and demand uncertainty.

Interaction effect. The impact of capacity level may depend on elapsed time. The likelihood of a technology shock increases over time and once a shock occurs, the seller may have to salvage the remaining capacity. Hence, the value of capacity may diminish as time elapses and we include the interaction between *CapL* and *lndod*.

Although we include many variables, we still have the "omitted variable" problem. Consistent estimators can be obtained only when the omitted variables are uncorrelated with our regressors. The factors we do not control for are the net cost of switching to an alternative product and contract terms other than price and quantity for a buyer. In the electronic companion, we show 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. Finally, note that we only consider instances with one-shot bargaining (or fixed-price contract) for new products and as such, it is reasonable not to consider any reference effects from previous discounts.

4.3 Regression Analysis

In this section, we explore the empirical relationship between the effective discount and demand share in two steps. In the first step we partition the buyers into groups based on their demand share and compute the average effective discount received by buyers in each segment. Based on the observed pattern, in the second step we use piecewise polynomial regressions to fit a functional form. Our analyses in these 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. 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 segmentation approaches that are detailed are 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 all our control 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 estimates

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
(i)	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
	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
(ii)	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
	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

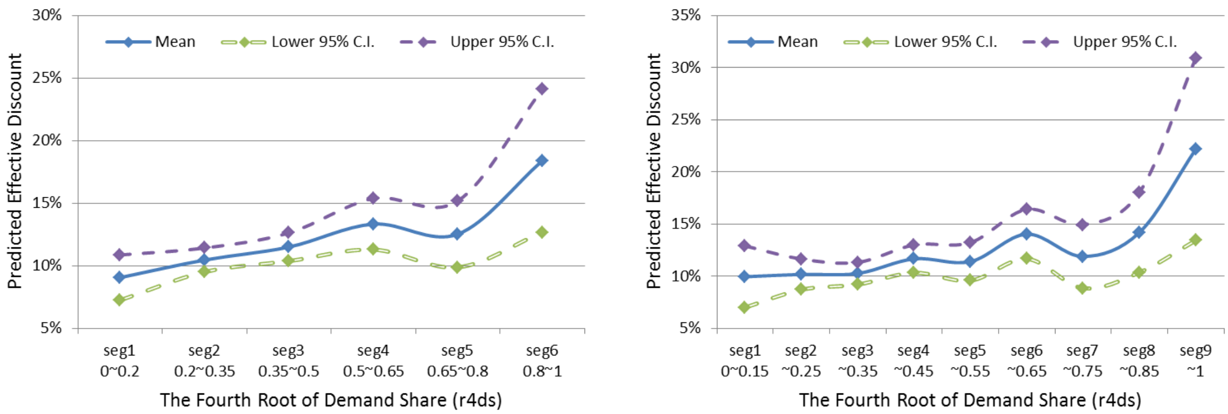
for a are summarized in Table 5 and for b in Table 7.[8]

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 in the first four segments, then decreases in segment 5, and increases again in segment 6. In model (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 a 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 determine the statistical significance of this finding. What we can infer is that it may be inappropriate to use a simple monotone function or a polynomial function. The discount curve is more likely to be a combination of a monotonically increasing curve and a V-shaped curve. Accordingly, in the next step, we fit a piecewise polynomial function and test the significance of the non-monotonicity.

[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 an ordinary OLS.

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



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

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] This allows us to infer 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 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*. The estimates are sensitive to initial values for the iteration performed by *nl*. To minimize the reliance on 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.

[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 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

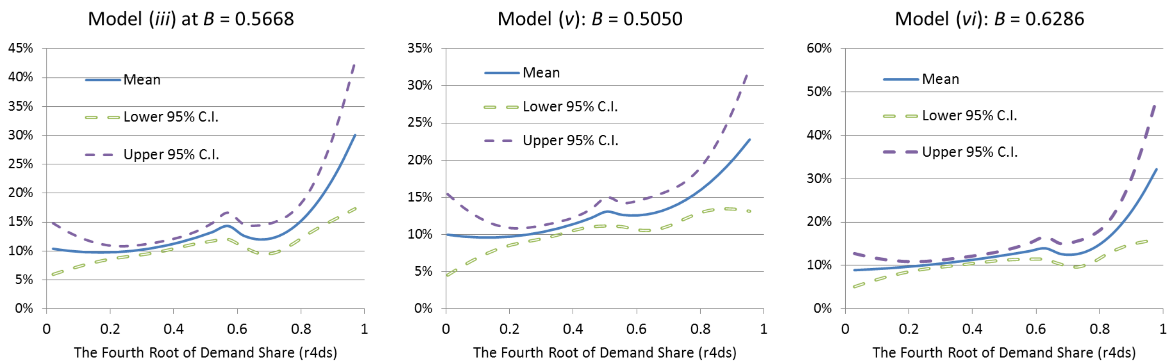
	Model (*iii*)			Model (*iv*)		
	Estimate	Robust S.E.	P value	Estimate	Robust S.E.	P value
a_1	0.2338	0.1058	0.027	0.1030	0.0319	0.001
a_2	0.2981	0.2208	0.177	-	-	-
a_3	-0.4378	0.2306	0.058	-0.3422	0.2001	0.087
a_4	2.0449	0.8653	0.018	1.9251	0.9564	0.044
B	0.5668	0.0346	0.000	0.5794	0.0449	0.000
	Model (*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	-	-

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 here. Because $B = 0.5668$ is highly significant (at the 0.1% level), the functional form cannot be described by a single linear or quadratic function. If we assume in 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 in the second segment first decreases with demand share and then increases. To establish *the significance of non-monotonicity*, we 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 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*) but with $B = 0.5668 \pm 2 \times 0.0309$. We call these two regressions model (*v*) and (*vi*), respectively, and we plot the predicted average effective discount against $r4ds$ in Figure 3. Although setting a biased knot could smooth out the decreasing part of the curve, we continue to observe N-shaped curves under both model (*v*) and (*vi*).

From this analysis we also find that the predicted discount decreases with the number of buyers

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.

for a product, increases with a buyer's size,[10] increases with the delays, and increases with the remaining capacity level. Additionally, the interaction effect between elapsed time and capacity level is negative. In other words, the impact of capacity level decreases 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 effect and the demand concentration rate, which may occur because fixed-price contracts entail long-term considerations and demand concentration is not a good measure for product popularity.

4.4 Robustness Checks

To determine 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. Next, 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 obtain similar non-monotonic curves. Third, we find that omitting the buyer's net cost of switching and the value of other contract terms do not affect the non-monotonic pattern if certain mild conditions hold. Lastly, we find that an "N" shape persists and is significant even after correcting for sample selection biases. The details of the robustness checks are in the electronic companion.

4.5 Further Discussions

Our observations from the regressions analysis are particularly interesting because they are inconsistent with conventional wisdom as well as our intuition: we anticipated that buyers with larger quantities would receive lower prices. While our intuition is correct for small- and large-quantity buyers, we observe a discount valley for medium-sized buyers that cannot be fully explained by existing theories. Why do we see such a discount curve? Although existing theories can predict increasing or V-shaped discount curves, they work under very different premises and cannot account for an "N" curve. Hence, we need models that shed light on price negotiations in B2B markets such as the semiconductor industry and identify factors that give rise to the observed patterns.

Our empirical findings also have important implications for both buyers and sellers. Because larger quantities may not always lead to lower prices, a buyer has to consider this possibility

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

while determining the purchase quantity. Despite this non-monotonic pricing pattern, *arbitrage opportunities do not necessarily exist for buyers, because empirically the total purchase cost still increases with quantity.*[11] For the seller, this non-monotonic relationship may require paying careful attention to posted prices and capacity allocated to different buyers. In the following sections we develop an analytical model to explore the rationale behind our observations.

5 Modelling and Verification

To deepen our understanding, we develop a model for a single 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.

5.1 The Model

Consider a seller $\mathcal{A}$ (she) that sells a new product to a group of OEMs. $\mathcal{A}$ has κ units of a fixed capacity. To be specific, we denote κ as the maximum amount of the product that can be produced during the product life cycle. The current industry practice is that major microprocessor companies produce according to their forecasts and start production as soon as the forecast is confirmed. Thus, the capacity does not expire during the planned product life cycle. The selling starts at time 0 and ends when no more buyers exist or the capacity is sold out. There are M potential buyers who arrive according to a general, *non-homogeneous Poisson arrival process.* 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 previously stated, in the semiconductor industry, technological advancements of competing products will lead to obsolescence. When such a technological shock happens, buyers that have already adopted the product and integrated it into their product designs, will continue their planned purchase. However, new buyers will cease to arrive and any unsold capacity will have to be salvaged. We assume that the arrival time of the technological shock is exponentially distributed with rate

[11] We find that in only 7% of instances across the entire data set, a buyer pays less in total than another buyer of the same product who buys less. This number is only 4% among fixed-price instances.

λ_0 and that seller $\mathcal{A}$ salvages the remaining capacity at a 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. 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. Each buyer's production plan is determined in advance as it is costly for buyer i to manipulate D_i. Although demand may not be exogenous from the buyer's point of view, that is not a concern of this paper. From the seller's standpoint, demands can be correlated and the distribution of each new buyer's demand can be history-dependent. 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) the expectation of the demand from a buyer is always finite, and (2) 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 his purchase quantity. (2) Seller $\mathcal{A}$ decides ρ_i. (3) Buyer i stays if $\rho_i \geq \eta$ and leaves permenently 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. It captures exogenous factors such as bargaining skills and the net cost of maintaining 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, respectively, 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

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

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 through simulations 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 Assumption 3 to simplify

[12]Note that ρ_i may not be a monotonic function of D_i given the complexity of the value function. According to our interactions with practitioners, it is the industry standard to fill at least 90% of the requirement. 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.

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

our analysis. Admittedly, when η is small this assumption may not hold and the price-quantity relationship may 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 $d_A = V(K_i, p, \psi_i) \cdot \mathbb{I}\{\bar{c}_i \leq p\} + [pQ_i + V(K_i - Q_i, p, \psi_i)] \cdot \mathbb{I}\{\bar{c}_i > p\}$, where $\mathbb{I}\{\cdot\}$ is an indicator function. In addition, we assume that $\bar{c}_i \geq c_L > s$ 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 price bargaining. Notice that β_i and $\bar{c}_i$ are known when the buyer arrives.

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 states 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 states 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.

5.2 Source of Non-Monotonicity

In our model, we conjectured that a buyer's bargaining power β and net switching cost $\bar{c}$ are not the source of price-quantity non-monotonicity and thus assumed for simplicity that they are independent of purchase quantity Q. To verify our conjecture, we compare three different models by running nonlinear regressions.

From Lemma 1 we can derive the discount received by buyer i for product j:

$$1 - \frac{w_{ij}}{p_j} = \mathbb{I}\left\{\hat{c}_{ij} < 1\right\} \cdot \left[1 - \beta_{ij} \cdot \bigtriangleup \hat{v}_{ij} - (1 - \beta_{ij}) \cdot \hat{c}_{ij}\right], \tag{6}$$

where $\bigtriangleup \hat{v}_{ij} = \left[V\left(K_{ij}, p_j, \psi_{ij}\right) - V\left(K_{ij} - Q_{ij}, p_j, \psi_{ij}\right)\right] / \left(p_j Q_{ij}\right)$, and $\hat{c}_{ij} = \bar{c}_{ij}/p_j$. Now we see that three factors can contribute to the non-monotonicity: $\hat{c}_{ij}$, β_{ij}, and $\bigtriangleup \hat{v}_{ij}$. 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 assume 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 $\bigtriangleup \hat{v} = \phi(r4ds) + b_v' \cdot X_v$. In model (II), we assume that the non-monotonicity is due to the net switching cost. In particular, $\hat{c} = \phi(r4ds) + b_c' \cdot X_c$, $\beta = b_b' \cdot X_b$, and $\bigtriangleup \hat{v} = b_v' \cdot X_v$. In model (III), we assume that the non-monotonicity is driven by the quantity-dependent bargaining power. In particular, $\hat{c} = b_c' \cdot X_c$, $\beta = \phi(r4ds) + b_b' \cdot X_b$, and $\bigtriangleup \hat{v} = b_v' \cdot X_v$. Note that although the three models have the same components, they are different in structure. The other explanatory variables in the analysis are as follows. 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.

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

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.

We run three nonlinear regressions based on the following equation:

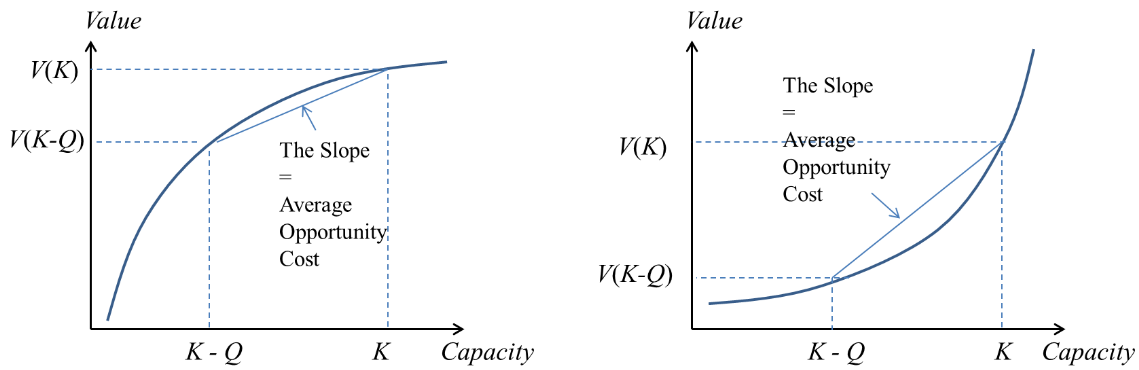
$$ED_{ij} = \mathbb{I}\left\{\hat{c}_{ij} < 1\right\} \cdot \left[1 - \beta_{ij} \cdot \vartriangle \hat{v}_{ij} - (1 - \beta_{ij}) \cdot \hat{c}_{ij}\right] + \hat{\epsilon}_{ij}. \tag{7}$$

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 degrees of freedom (d.f.), model (I) has the highest R^2 and the lowest sum of squared residuals (SS). If we assume that $\hat{\epsilon}$ is normally distributed, we can use the corrected Akaike's Information Criterion (AIC_C) value (Akaike 1981) to compute the evidence ratio and compare the models.[14] This shows us that model (I) is 2.26×10^{15} times more likely to be correct compared with model (II) and 239 times more likely compared with model (III). The evidence is overwhelmingly in favor of model (I) and we conclude that model (I) is the best model among the three.

Other models can be derived by combining models (II) and (III). However, a serious collinearity problem would exist if we let β depend on $r4ds$ in (II) or let $\hat{c}$ depend on $r4ds$ in (III). Hence, these models cannot provide better explanations. Lastly, when we add $r4ds$ as a linear part of $\hat{c}$ in model (I), we find that the estimated coefficient is not significant. Thus it is reasonable to assume that $\hat{c}$ is not correlated with $r4ds$.

[14] The corrected AIC value is $AIC_C = N \cdot \ln\left(\frac{SS}{N}\right) + \frac{2 \cdot J \cdot N}{N - J - 1}$, where N is the number of observations and $J - 1$ is the number of parameters in the model. The evidence ratio, $\Pr\{\text{model (I) is correct}\} / \Pr\{\text{model (II) is correct}\}$, is $\exp\left[\left(AIC_C^{(II)} - AIC_C^{(I)}\right)/2\right]$.

Figure 4: The Intuition Behind Proposition 1



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 show some examples of the price curve.

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 $V_i(K)$ in place of $V(K, p, \psi_i)$. We know the relationship between the price w_i and quantity Q_i depends on the sign of the first-order derivative of w_i in (5) with respect to Q_i:

$$\frac{\partial w_i}{\partial Q_i} = \frac{\beta_i}{Q_i}\left[V_i'(K-Q_i) - \frac{V_i(K) - V_i(K-Q_i)}{Q_i}\right]. \tag{8}$$

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 observe quantity discount, our simple model requires the value function to be convex. Figure 4 illustrates the intuition behind Proposition 1. Note that given buyer i's outside option and bargaining power, w_i depends on the seller's average opportunity cost of selling Q_i units. We 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 cannot not be simply convex or concave in order to observe a non-monotonic price-quantity relation. In fact, we show that a value function that is a combination of convex and concave functions 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 this property *convex-concave*. 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.

It is reasonable to expect the value function to be convex-concave or *S-shaped*. When capacity is very low, the seller is unlikely to fulfill any buyer's need and will have to salvage. As capacity increases, the likelihood of satisfying some buyers' needs increases. When capacity is very high, it may exceed demand and the seller may once again have to salvage a portion.

However, this result alone is not satisfactory, because it cannot explain the pricing pattern we observe in the data. The preceding discussion was based on perturbing the purchase quantity of a single buyer with a fixed-value function. If there are multiple buyers, the seller may update her 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. In the following, we extend our analysis 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 may 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 $Q_i = \rho^*(K, p, D_i, \beta_i, \bar{c}_i, \psi_i) \cdot D_i$ can differ for different parameter values. As a result,

$$
\begin{aligned}
V_{i-1}(K) = \delta_i \cdot s \cdot K + (1-\delta_i) \cdot \Bigg\{ & \int_{K/\eta}^{+\infty} V_i(K)\, dF(D|\psi_{i-1}) \\
& + \int_0^1 \int_0^{K/\eta} \int_p^{+\infty} \left[p\rho^* D + V_i(K-\rho^* D)\right] dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \\
& + \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} \left[\beta V_i(K) + (1-\beta) V_i(K-\rho^* D)\right] dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \\
& + \int_0^1 \int_0^{K/\eta} \int_{c_L}^{p} (1-\beta)\,\bar{c}\rho^* D dG(\bar{c}|\psi_{i-1})\, dF(D|\psi_{i-1})\, dB(\beta|\psi_{i-1}) \Bigg\}. \qquad (9)
\end{aligned}
$$

Clearly, 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

$$\nu(p, \psi_{M-1}) \;=\; p - s - \mathbf{E}_{\beta,\bar{c}}\left[(p-(1-\beta)\bar{c} - \beta s) \cdot \mathbb{I}\{\bar{c} \leq p\}\,|\psi_{M-1}\right]. \qquad (10)$$

The key idea of the proof for the lower bound is the following. First, setting $\rho = 1$ in (9) leads to a lower bound of $V_{i-1}(K)$. We then utilize the fact that the lower bound is a separable function of D and we iteratively plug the lower bound into (9). We complete the proof by induction. The proof for the upper bound is similar, except that we use $\rho^* \leq 1$. We have $\mathbf{E}_{\beta,\bar{c}}\left[\nu(p,\psi_{i-1})|\psi_{i-2}\right] = \nu(p,\psi_{i-2})$ and $\int_0^{+\infty}\int_0^{\lambda K} DdF(D|\psi_{i-1})\, dF(D'|\psi_{i-2}) = \mathbf{E}\left[\mathbf{E}\left[D \cdot \mathbb{I}\{D \leq \lambda K\}\,|\psi_{i-1}\right]|\psi_{i-2}\right] = \int_0^{\lambda K} DdF(D|\psi_{i-2})$.

Theorem 1. *For any $1 \leq i \leq M$ and $K \geq 0$,*

$$V_{i-1}(K) \geq sK + \left(1 - \delta_i^l\right) \cdot \nu(p, \psi_{i-1}) \cdot \int_0^{K \cdot \lambda^{M+1-i}} DdF(D|\psi_{i-1}), \qquad (11)$$

$$V_{i-1}(K) \leq sK + (1 - \delta_i^u) \cdot \nu(p, \psi_{i-1}) \cdot \int_0^{K/\eta} DdF(D|\psi_{i-1}), \qquad (12)$$

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

Because $\lim_{K\to+\infty} F_0(K-\lambda K) = 1$, 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 very 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 constant: $\delta = \frac{\lambda_0}{\lambda_0+\lambda_b}$. 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 $\mathbf{E}\left[V_i(K)|\psi_{i-1}\right] = \mathbf{E}\left[\mathcal{R}\left(K, \{D_n, \beta_n, \bar{c}_n, t_n - t_{n-1}\}_{n=i+1}^{\infty}\right)|\psi_{i-1}\right] = V_{i-1}(K)$. Leveraging this property of the value function, we obtain an upper bound and a lower bound for $V_{i-1}(K)$. In preparation, let

$$H_i(K) = \mathbf{E}\left[p - (p-(1-\beta)\bar{c})\cdot \mathbb{I}\{\bar{c}\le p\}|\psi_{M-1}\right]\cdot \int_0^{K/\eta} D dF(D|\psi_{i-1}), \tag{13}$$

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

$$h_i(K) = \mathbf{E}\left[V_i\left([K-D_i]^+\right)|\psi_{i-1}\right]/V_{i-1}(K). \tag{14}$$

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*

$$\frac{s\cdot K + \frac{1-\delta}{\delta}\cdot H_i(K)}{1+\frac{1-\delta}{\delta}\cdot[1-h_i(K)]\cdot(1-\mathbf{E}[\beta|\psi_{i-1}]\cdot G(p|\psi_{i-1}))} \le V_{i-1}(K) \le s\cdot K + \frac{1-\delta}{\delta}\cdot H_i(K). \tag{15}$$

If $\frac{\partial}{\partial K}V_i(K)\ge s$ 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 (15), respectively. We have that $L_{i-1} = U_{i-1}/(1+Z_i)$, where $Z_i = \frac{1-\delta}{\delta}\cdot[1-h_i(K)]\cdot(1-\mathbf{E}[\beta|\psi_{i-1}]\cdot G(p|\psi_{i-1}))$. 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 result implies 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, $\mathbf{E}[\beta|\psi_{i-1}]$, and $G(p|\psi_{i-1})$. 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. The condition $\frac{\partial}{\partial K} V_i(K) \geq s$ should be satisfied by definition because we assume that the seller can always salvage the capacity at marginal value s. Thus, s should be the lowest marginal value for $V_i(K)$.

6.5 Discussion

Note that $H_i(K)$ can be written in the form of $a' \cdot \int_0^{K \cdot a''} DdF(D|\psi_{i-1})$, where a' and a'' are parameters independent of K. Thus, the upper and lower bounds given by Theorem 1 and 2 can all be written in the form of $a \cdot s \cdot K + a' \cdot \int_0^{K \cdot a''} DdF(D|\psi_{i-1})$, where a, a', and a'' are parameters independent of K. Moreover, we learn from the proof of Theorem 1 that the value function of a single-period problem takes a similar form. In essence we are 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 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 bounds are all convex-concave.

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

Any other unimodal distribution may also generate S-shaped bounds, and it is quite natural to expect a unimodal demand distribution. Therefore, the value function is very likely to be *convex-concave* as described in Proposition 2 given that the bounds are tight enough.

6.6 Plotting the Price Curve

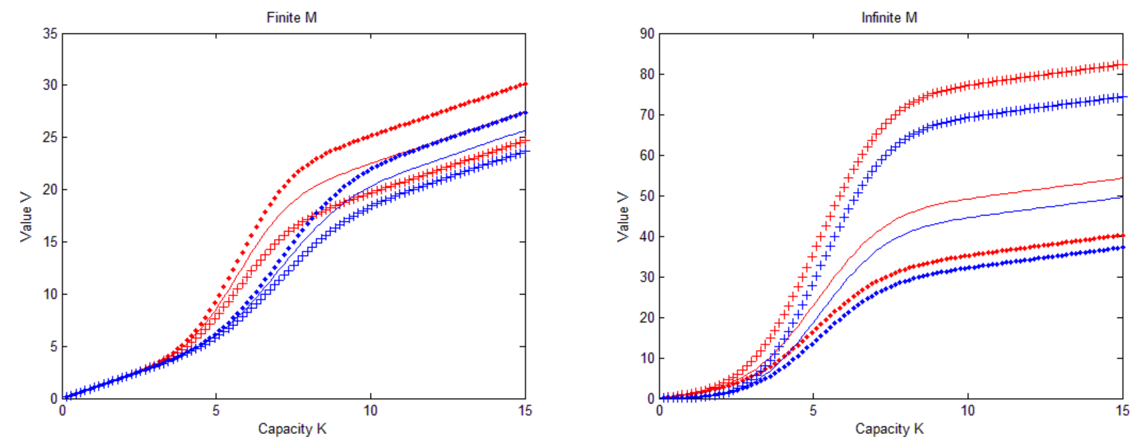
From (5) we know that the negotiated transaction price w_i is a linear combination of both parties' outside options. The seller's outside option is the average opportunity cost of selling Q_i, which depends on Q_i, capacity K_i, and the value function. In this section, we first investigate the performance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

Without loss of generality, we consider the seller 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, and 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 the 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. After 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, that is, larger s_m means more concentrated demand. This way of updating implies that if buyer i is very large, the rest of the buyers are likely to be small, especially when the seller 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 $h_i(K) \approx 1 - s \cdot \mathbf{E}\left[D_i|\psi_{i-1}\right]/V_{i-1}\left(K\right)$. 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.

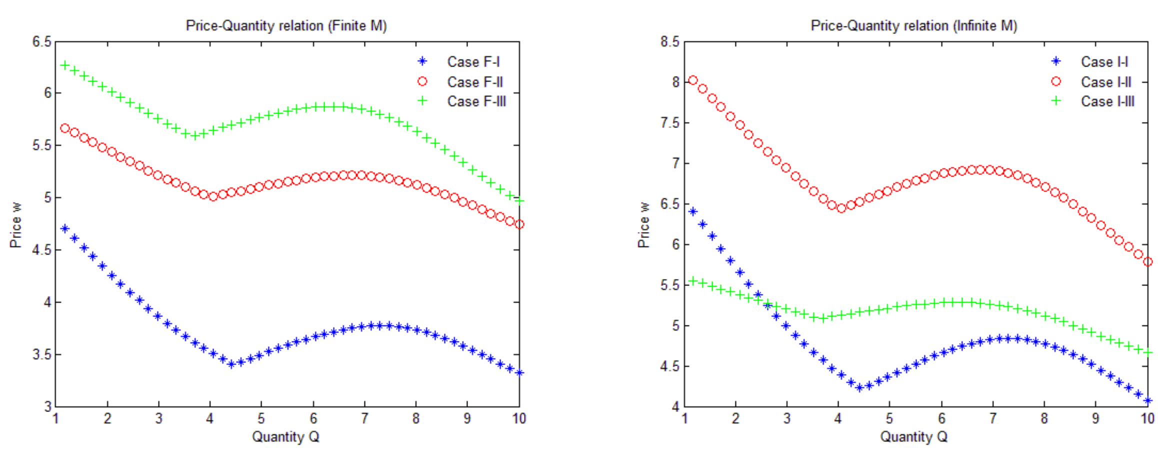
In Figure 6, we use the upper bounds in each case (of finite vs. infinite M) as the approximation of the value function and generate the negotiated price for three different scenarios. With an S-shaped value function, we obtain a price-quantity curve in all scenarios that is reversed-N-shaped, which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

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 $\delta = 0.4$.

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: $\beta_i = 0.5$; $\delta = 0.6$; $s_m = 1.1$; $C_V = 0.3$.

7 Managerial Implications

According to our model, the reason some buyers are receiving lower discounts than those 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 that may be too much for small buyers but not enough for large buyers. This logic is reflected in the S-shaped value function, as the seller's marginal opportunity cost of selling to a buyer (or marginal loss of value) increases the most with the purchase quantity when the remaining capacity gets close to the mean demand of the next buyer. In this section, we begin by discussing the capacity allocation decision for the seller given the non-monotonic price-quantity relation. Basically, the seller needs to avoid mid-sized transactions by controlling the capacity allocation. Next, we discuss how the posted price should be set given its influence on the selling process.

7.1 Dynamic Capacity Rationing

In this setting the seller should control the capacity that is allocated to each buyer while staying in a range that is acceptable to the buyer. 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 $r_i < V_i'(K - Q_i)$.*

A rationing decision depends on the remaining capacity level, purchase quantity, demand distribution, and the buyer's profit margin before subtracting the cost of this product. If we hold r_i constant, the allocated capacity should be reduced when $K - Q_i$ is close to the anticipated mean of the demand from the next buyer; otherwise, the seller should sell as much as possible. This enables the seller to reduce the likelihood of losing the next major buyer due to insufficient capacity. Due to the shape of the value function, $V_i'(K - Q_i)$ is likely to be higher 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. Clearly the seller should reserve capacity for buyers who are willing to pay more. Overall, we can conclude that the allocation decisions cannot be based solely on capacity levels and purchase quantities of the current buyer.

7.2 Posted-Price Optimization

When the posted price is determined, 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. Hence, incorrect anticipations of the average discount (or essentially the opportunity cost of selling to each buyer) 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 $p^* = \arg\max_p V_0(\kappa, p)$. Readers are referred to the electronic companion for details.

7.3 Implications for Other Industries

There are other industries that resemble the semiconductor industry. In the travel, airline, and hotel industries, limited capacities have to be sold within a limited period. Customers in these industries include bulk buyers such as travel agencies and resellers who also negotiate prices. Findings from our study may carry over to these businesses. There are, however, a few important differences between the semiconductor industry and the others. In the semiconductor industry, the obsolescence date is stochastic while for the other industries it is deterministic. 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 the 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, because selling mid-sized quantities remains costly for sellers given fixed capacities. Theoretically, Lemma 1 and Eq. (6) still hold even if both price and quantity are determined via Nash bargaining. Therefore, the formulation of the value function as well as the subsequent analysis remains unchanged.

8 Concluding Remarks

In this data-driven research, we studied the price-quantity relationship in a B2B market in which the product life cycle is short and prices are set through one-shot negotiations. Using data from the microprocessor market, we found that the price can be a non-monotonic function of the purchase quantity. Contrary to our intuition, larger quantities can actually lead to higher prices. In particular, we observed an N-shaped discount curve in the data. In our empirical analysis we controlled for a number of factors that offer competing explanations for our observation. We also showed that some of the omitted variables do not alter the key observation.

We developed and analyzed a stylized model to delve into this phenomenon to understand the underlying rationale. Our analysis reveals that it is fairly plausible for the price-quantity relation to be non-monotonic. If the value of residual capacity is convex-concave we can then observe a non-monotonic price-quantity curve. Although we normally assume that the value function is increasing and concave, our model shows that this is not necessarily 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, that is, an N-shaped discount curve.

There are many limitations to our work. Our analysis assumes a static Nash bargaining model with fixed capacity. Future research may obtain different price-quantity relations by considering dynamic strategic bargaining and capacity levels that diminish with time. Also it was impossible for us to consider and eliminate all other explanations for the N-shaped curve. It is possible that certain behavioral biases may also result in this pricing pattern. We sincerely hope that our work stimulates additional theoretical and empirical research.

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.