

Agree Now or Later? Strategic Vertical Bargaining with Limited Capacity and Symmetric Information

Wei Zhang

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

Reza Ahmadi

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

Sriram Dasu

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

Abstract

In this paper we study dynamic strategic bargaining in business-to-business markets wherein a seller and a buyer possess symmetric information and negotiate over the price of a product for which the seller has a limited capacity. Motivated by the microprocessor market, our model offers an explanation for the occasional observations of delayed price agreement in this market, which cannot be explained by existing theories, and sheds light on how to make tactical decisions such as when to propose, reject, and accept counter-offers. In our model, the buyer can choose to wait if the seller rejects the counter-offer; as a countermeasure, the seller threatens to sell part of the limited capacity to other buyers who may possibly arrive later; the probability of the arrival of other buyers is updated while waiting. The subgame perfect equilibrium predicts that (1) it is credible for the buyer to wait when the gain from the trade is low and that (2) the seller should encourage the focal buyer to wait when selling to other buyers is more profitable. If prices are always settled without delay, the seller can lose more than 10% of its revenue according to our numerical study. We also investigate how to optimize the posted price given the anticipated equilibrium of bargaining and show that incorrect anticipations of the time of agreement lead to ineffective prices, which can be 8% lower than the optimal price.

[*Keywords*: business-to-business; price bargaining; dynamic game; delayed agreement]

1 Introduction

In many business-to-business (B2B) markets, such as markets for raw materials, airplanes (Garvin 1991), medical devices (Grennan 2013), and microprocessors (Zhang et al. 2015), prices are normally determined by negotiations between buyers and suppliers. Understanding how firms form agreements on price is essential to understanding supply chain activities and performances in these markets. As known to the public, in order to obtain ideal prices, firms use various bargaining tactics such as waiting, introducing competition, forming coalition, etc, in different circumstances. While the academia has developed some good understanding of why and how these bargaining tactics are used in various B2B markets (e.g., Nagarajan and Bassok, 2008, Aydin and Heese, 2014, and Feng et al., 2015), we are still short of knowledge regarding why firms sometimes fail to reach an agreement until the last minute even when they understand each other very well.

A recent example involves Apple and Taiwan Semiconductor Manufacturing Company (TSMC) in the microprocessor market. According to *The Wall Street Journal*, Apple and TSMC have collaborated since July, 2014 on designing and testing the 14/16-nanometer A9 processor that should power 2015 iPhones and iPads.¹ Hundreds of TSMC engineers were sent to Apple's headquarters to work on the project. Apple finally decided to award nearly one third of the A9 processor orders to TSMC around April, 2015, after wandering between GlobalFoundries and TSMC for more than half a year.² However, it is reported that the price was still unsettled for this deal even in August, 2015 when TSMC already had its 14/16-nanometer FinFET Process capacity ready and the new iPhone would be launched shortly.³ According to the August article on DigiTimes.com, Apple requested price cuts from both its A9 processor suppliers, Samsung and TSMC, but TSMC was inclined to refuse to drop prices given that its 16-nanometer FinFET process had been adopted by a number of downstream buyers such as Avago, Freescale, Nvidia, MediaTek, etc, who could be the potential users of its capacity.

In fact, such delays of price agreement are not unusual in the microprocessor market, although they do not happen frequently. For this research, we interacted with managers of a major semiconductor company and obtained a data set that contains sales (billing) record of this company over

¹<http://www.wsj.com/articles/tsmc-starts-shipping-microprocessors-to-apple-1404991514>

²<http://appleinsider.com/articles/15/04/15/apple-makes-last-minute-decision-to-use-tsmc-for-30-of-a9-chip-orders-for-next-iphone>

³<http://www.digitimes.com/news/a20150811PD217.html>

Table 1: Summary Statistics for the 130 Observations

	Mean	Median	S.D.	Min	Max
Total Purchase Qty.	3.01E6	1.22E5	9.45E6	155	8.84E7
Highest Price (US\$)	143.25	84	216.42	7	1469
Waiting Time (days)	69.05	7	154.14	7	892

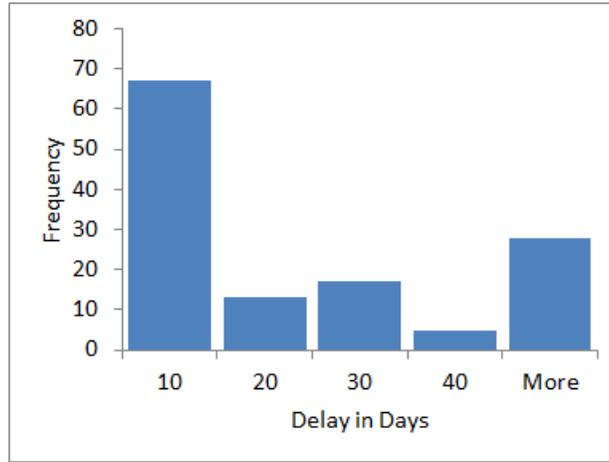
a three-year period from early 2009 to early 2012.⁴ During this period, we find 130 observations wherein a customer started price negotiation for a product but did not settle the price immediately. The waiting time or delay of pricing is measured by the recorded time gap between the start of negotiation and the determination of price. Table 1 shows some summary statistics for these observations and Figure 1 shows the distribution of waiting time. We can see that the average and the median waiting time before the price is settled are 69 and 7 days, respectively. Note that the company interacts with the major customers repetitively for multiple generations of products, so they know each other quite well.

Why do firms sometimes fail to reach an immediate agreement when they have symmetric information? It seems that no clear answer can be found in the literature. Traditional strategic bargaining models that build on Rubinstein (1982) often assume information symmetry and predict immediate agreement. The intuition is that actual waiting is not necessary because the consequence can be rationally expected and buyers just need to show the threat of waiting if it is credible. A key assumption for this stream of research is that delay is costly and it hurts both parties. Another stream of research on strategic bargaining assumes information asymmetry and thus that waiting or delaying the negotiation, although might be costly for a negotiator, can benefit him- or herself by signalling his or her low willingness to pay or “patience” (e.g., Admati and Perry 1987 and Cramton 1992). A third stream on strategic bargaining (e.g., Ma and Manove, 1993) assumes that waiting can make the opponent more “impatient” because there may not be enough time to make a counteroffer after the delay due to the long communication lead time. Nevertheless, the long communication lead time is not an appropriate assumption in B2B markets such as the microprocessor market because advanced technologies nowadays enable fast communications.

In this research, we focus on the microprocessor market and try to find an explanation to our observation in this specific setting, despite that delays of agreement are frequently observed in many

⁴The information was recorded when the sale was made (or the contract was signed).

Figure 1: Histogram of Waiting Time



other practical situations where information should be symmetric (Friedenberg 2014). If delay of price agreement turns out to be a rational result rather than a mistake in the microprocessor market, a follow-up question is that under what circumstances and for how long should firms wait? At the tactical level, when should firms propose, reject, and accept counter-offers in the process of strategic bargaining?

To answer these questions, we should first understand the typical selling process of a microprocessor product. In this market, a seller normally has a fixed capacity due to the long lead time to set up production facilities. Potential buyers arrive sequentially, request certain quantities that are determined in advance, and negotiate for the price.⁵ Normally, buyers have some flexibilities regarding when to secure the supply as long as their productions are scheduled at a later time. Therefore, if a counter-offer is rejected, a buyer may choose to wait. The seller may be forced to cut price after waiting if the demand turns out to be low, but the seller may also be able to sell the capacity to other buyers during this time. If the capacity becomes insufficient, the buyer may be forced to accept partial fulfillment because it is normally infeasible to use another product as supplement due to differences in design or specifications and it is costly to find an alternative source to supply the same product in a short period. In certain cases the target quantity may be changed, but it should be costly for the buyer to modify the production plan. If the bargaining breaks down,

⁵To produce a final product, an OEM buyer normally needs numerous different components, so it is difficult to manipulate the purchase quantity of a particular component.

the buyer re-designs its product and switches to another seller for a similar component. In sum, in the microprocessor market, waiting can be a risky strategy for buyers, and the seller's bargaining position depends on the expected revenue of selling to subsequent buyers.

Motivated by the microprocessor market, we study in this paper a strategic bargaining model with symmetric information. While considering that buyers use the tactic of waiting, we assume that the seller can counteract by threatening to sell a portion of the capacity to other buyers. We focus on situations wherein buyers and the seller negotiate over the price only once. Using this model, we learn that waiting can sometimes benefits both the buyers and the seller. From a buyer's perspective, waiting can force the seller to re-evaluate the prospect of future demands and thus the buyer can possibly obtain a lower price later, nevertheless, at the risk of receiving a smaller capacity. We find that it is a risk worth taking for the buyer when the final product is not very profitable such that switching to an alternative product will lead to a loss. Interestingly, waiting can sometimes benefit the seller as well, because it awards the seller the opportunity to gain higher profit by selling the capacity to more profitable buyers. When the seller prefers to settle the price later and waiting is credible for a buyer, the seller should reject any counter-offer and encourage the buyer to wait. This is when waiting should actually happen.

The rest of this paper is organized in the following way. We present a brief literature review in Section 2. In Section 3 we construct a dynamic model of B2B price bargaining; we analyze this model and obtain the main results in Section 4. We then explore the property of the negotiated price in equilibrium in Section 5, and we show that our theory can offer an explanation for the empirical observation of non-monotonic price-quantity relationship documented in the recent literature. Later in Section 6, we discuss how to optimize the posted price for the seller and the consequences of failing to anticipate the result of price negotiation. In Section 7, we study a modified model wherein the seller allocates her capacity on the go to multiple major customers. Lastly, we conclude the paper in Section 8. All the proofs are in the Appendix.

2 Related Literature

This paper is mainly related to three streams of research in the literature. The first stream is about bargaining with delay of agreement, the second is multilateral bargaining in supply chains, and the

third is posted pricing with the presence of both price-takers and bargainers.

Delays in reaching agreement are frequently observed in practice and studied in the literature. Cramton (1984) is among the earliest to study the phenomenon that trade often occurs after costly delay, and it is suggested that the need for learning each other's valuation under incomplete information results in rejections and thus the delay. Later, Admati and Perry (1987) and Cramton (1992) proposed slightly different models in which it is assumed that bargainers can signal the strength of their bargaining positions by delaying prior to making an offer. In the latest paper of this stream, Feng et al. (2015) predict delay in price-quantity contract settlement in supply chains where the demand information is known only to the buyer. In all these papers, there is an infinite horizon, and the delay of agreement is driven by information asymmetry. Roth et al. (1988) considered bargaining with deadlines and they documented some experimental evidences of last-minute agreements, which initiated further theoretical work. Based on this observation, Ma and Manove (1993) proposed a model of continuous-time, alternating-offer with a deadline and symmetric information. They assume that players can decide when to make an offer or counter-offer, but only after an exogenous, random delay due to information transmission and processing. Their model predicts that players adopt strategic delay early in the game, make and reject offers later on, and reach agreements late in the game if at all. Notice that the player who makes an offer closer to the deadline is less likely to be rejected, so the delay in their model is driven by the desire to obtain a stronger bargaining power. We also consider a continuous-time, alternating-offer game with a deadline and symmetric information. However, our model has a different game structure, in which the bargainer can make a counter-offer immediately and threaten to wait if the seller rejects. As a countermeasure, the seller can threaten to sell part of the capacity to price-takers whose arrival is random and shapes the value of the capacity in a way. We show that sometimes it is credible for the bargainers to wait and it is also better for the seller to reject the bargainers and encourage them to wait till the last minute. Hence, our theory adds to the literature that explains the last-minute agreements. In a completely different modeling framework from ours, Friedenber (2014) provides a novel understanding of delay in reaching agreements based on the idea of strategic uncertainty.

The literature of multilateral bargaining in B2B markets or supply chains is emerging in recent years and this stream of research normally focus on how the profit is allocated among supply chain members. Depending on the game structure that is used, there are mainly two stream of research

in this literature. One stream assumes that the multilateral bargaining happens sequentially and considers the impact of bargaining sequence and coalitions. In an assembly chain setting, Nagarajan and Bassok (2008) consider suppliers who form multilateral bargaining coalitions and compete for a position in the bargaining sequence. Nagarajan and Susic (2008) study the stability of coalition in assembly models. Different from the sequential models, in which bargaining power is manifested in the position in the sequence, simultaneous multilateral negotiation models focus on the static equilibrium in which the profit allocation is determined by the contribution of a negotiator to the entire system. Guo and Iyer (2013) compare simultaneous and sequential multilateral bargaining games in a channel system that has one manufacturer and two competing retailers. They show that the manufacturer's choice of timing of bargaining, i.e., simultaneous versus sequential bargaining, should depend on the dispersion in retail prices. Dukes et al. (2006) and Lovejoy (2010) use simultaneous bargaining model to study the impact of channel or chain structure. In a retail setting, Aydin and Heese (2015) study an assortment problem of a retailer that engages in simultaneous bilateral negotiations with all manufacturers for a given assortment. In this paper, we first focus on a subgame in a sequential bargaining framework and then we modify our model to investigate simultaneous bilateral bargaining with all the buyers. We supplement this literature by explicitly modelling the dynamics of a negotiation and offering explanations to both the observed delay of agreement and the non-monotonic price-quantity relationships empirically identified in Zhang et al. (2015).

Price-takers and bargainers coexist in many B2B and retail markets, and it is important for sellers to take both types into consideration when optimizing the posted price. Gill and Thanassoulis (2009) study a static model wherein they find that an increase in the proportion of bargainers can lower the consumer surplus overall. In contrast, Kuo et al. (2011) study a dynamic model wherein they focus on the dynamic pricing problem and the role of limited inventory when a firm sells to both price-taking and haggle-prone customers. Our study focuses on selling to a major business customer and the role of the price-takers is to serve as the outside option for the seller, shape the value of the limited capacity, and help the seller carry out the threat. Although our model is dynamic, we consider a static pricing problem.

3 A Dynamic Model of B2B Price Bargaining

In our base model, we assume that buyers arrive sequentially for a product and we focus on the problem of bargaining with one of the buyers. In Section 7, we will extend the model to consider simultaneously bargaining with multiple buyers who share the seller's capacity. We devised the model to capture the following features: capacity is limited; the seller and buyers learn over time the demand potential; buyers can threaten to wait if their counter-offers are rejected; and the seller can threaten to sell the capacity to other buyers. Hence, the seller and buyers use different bargaining tactics. Using this framework, we try to understand the dynamics of the bargaining process and the time it takes to settle the price. We consider a one-shot bargaining event because multiple transactions and repeated bargaining will unnecessarily complicate the analysis. In fact, one-shot price bargaining is common in practice; for example, when fixed-price contracts are used.

Seller \mathcal{A} (she) introduces the new product to a group of OEM buyers and sets the posted price p_A . Based on forecasts provided by different customers, \mathcal{A} builds up a finite capacity which is to be allocated to customers sequentially according to their sequence of determination of adoption or the sequence of "arrival". The seller has K units of the capacity left when \mathcal{C} (he), a major customer, arrives and agrees to adopt this product. We define it as time 0 when customer \mathcal{C} approaches seller \mathcal{A} and asks for Q units of the capacity. For simplicity, we assume \mathcal{C} accepts partial fulfillment of his requirement and \mathcal{A} tries to fulfill as much as possible.

We denote r as \mathcal{C} 's marginal payoff before subtracting the procurement cost. As an outside option, \mathcal{C} could also adopt an alternative product from a different seller. Let \bar{c} represent the net marginal cost of buying from the alternative supplier in order to keep the same margin r . As we discussed earlier, \mathcal{C} could adopt only one product. In other words, if \mathcal{C} accepts partial fulfillment of his order, \mathcal{C} could not use alternative products from other sellers for the shortage.

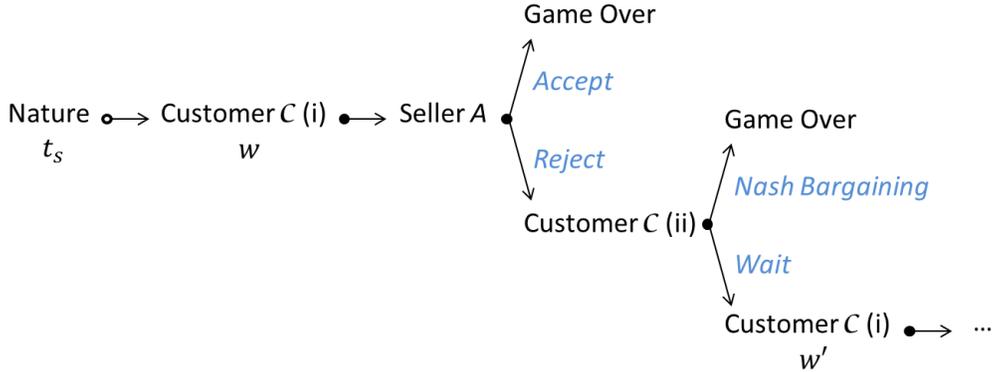
Given the allocated capacity, the price is negotiated. We assume that all information between \mathcal{A} and \mathcal{C} is symmetric, which is due to their repetitive interactions and information exchange after \mathcal{C} 's arrival. The negotiation starts with \mathcal{C} proposing a counter-offer w against the posted price. As a common tactic, \mathcal{C} threatens to walk away from the negotiation and wait if the counter-offer is rejected by \mathcal{A} . Denote $T > 0$ as the deadline that is required by \mathcal{C} 's production schedule to finalize the purchase. Hence, the buyer has some flexibility to manipulate regarding when to close the deal

and thus his threat is *possible*. However, his threat is *credible* only when he has the best incentive to carry it out, say, when the customer expects to get a lower price and sufficient capacity later. As a countermeasure, \mathcal{A} threatens to sell part of the remaining capacity to other buyers if \mathcal{C} waits. For the sake of tractability, we simplify the model by assuming that the demand from other buyers during the time interval $[0, T]$ is a stochastic shock of fixed size $s \in (0, K]$. In addition, we assume that all subsequent buyers pay the posted price p_A . Note that the results won't be qualitatively different if we assume that the value of capacity depends on the expected average price paid by subsequent buyers which we call the reference price. When information is symmetric, \mathcal{C} never pays higher than the reference price; hence, \mathcal{A} 's threat is always credible because subsequent buyers always contribute higher expected revenue. Given the counter-offer, \mathcal{A} chooses to accept or reject.

Due to the uncertainty, other buyers may not appear during $[0, T]$. Suppose the demand shock is realized at time t_s , which follows a general probability density function $\lambda(t)$ with support $[0, \infty)$. Let $\Lambda(t) := \int_0^t \lambda(x)dx$ be the cumulative distribution function and $\bar{\Lambda}(t) := 1 - \Lambda(t)$ be the complement. Accordingly, $\Lambda(t) = \Pr\{t_s \leq t\}$ represents the probability of having the demand shock by time t . The arrival of demand shock, of course, should depend on p_A . Rather than scaling $\lambda(t)$, we model the dependence of demand on price by scaling the time horizon. Denoting the original time t_o , we define $T(p_A) := T_o \cdot (1 - \alpha p_A)$ and $t(p_A) = T(p_A) \cdot \frac{t_o}{T_o}$ for any p_A . Notice that, T is defined once the posted price is known at time 0, so the model is well specified. By doing so, we maintain the basic logic: as p_A increases, $\Lambda(t)$ decreases. In other words, the time at which the demand shock is realized is stochastically decreasing in p_A . Also, we avoid assuming an explicit functional form for λ , which allows us the flexibility to extend the model to bargaining with multiple customers. For convenience, we will write $\Lambda(t)$ as Λ_t .

Figure 2 lays out the sequence of events. Let $x \wedge y$ denote $\min\{x, y\}$. We know that the transaction quantity is $Q \wedge K$ if the demand shock has not occurred and is $Q \wedge (K - s)$ if otherwise. At any particular time t , if \mathcal{C} 's counter-offer is rejected and he does not wait, we assume that the two parties will be engaged in a Nash bargaining (Nash 1950) and the bargaining solution will be given by $\bar{p}_t = \arg \max_p (d_t^C - p)^\gamma (p - d_t^A)^{1-\gamma}$ where $\gamma \in [0, 1]$ is \mathcal{C} 's relative bargaining power and d_t^i denote the disagreement price for $i \in \{\mathcal{A}, \mathcal{C}\}$ at time t . Actually, the assumption of a Nash bargaining solution is plausible, even in our strategic setting. According to Binmore et al. (1986), a Nash solution can approximate the equilibria of strategic models when the interval between consecutive

Figure 2: Sequence of Events



counteroffers approaches zero, which—because we use a continuous time model—is the case here. However, if $p_A \leq \bar{c}$, it is not credible for \mathcal{C} to switch (i.e., $d_t^C = d_t^A = p_A$) and thus he has to pay the posted price.

If the capacity has not been sold out by time T , seller \mathcal{A} can sell what is left to buyers who arrive afterwards. Or, she can downgrade the product to supplement her supply of lower grade products.⁶ Let $\pi(x)$ denote the salvage value of \mathcal{A} 's residual capacity x after time T . In general, we assume that $\pi(x)$ is increasing in x . To make sure there exists a Nash bargaining solution, we assume $\pi(K) \leq p_A \cdot K$.

4 Model Analysis

The objective of our analysis is to characterize the conditions for \mathcal{C} 's threat to be credible, obtain the best response of \mathcal{A} to \mathcal{C} 's counter-offer, and derive the negotiated price in equilibrium. We assume that both \mathcal{A} and \mathcal{C} aim to maximize their respective expected payoffs.

4.1 The Bargaining Process

If \mathcal{C} walks away and waits during the bargaining process, he may return with a counter-offer at any time. Thus, suppose \mathcal{C} starts or resumes the bargaining process at time t . First of all, if the demand shock has already occurred ($t_s \leq t$), then it is automatically credible for \mathcal{C} to wait from t to T , because \mathcal{A} 's capacity level will no longer change. Since time discounting is negligible

⁶This is normal in semiconductor and high-tech supply chains.

and both parties' outside options are fixed in this case, it is equivalent to bargaining at time T . Without loss of generality, we assume that \mathcal{C} will buy at time T . Let $p_h := \bar{p}_t|_{t_s \leq t}$ denote the Nash bargaining solution, or NB price, at time t given that $t_s \leq t$. Note that $p_h = p_A$ if $p_A \leq \bar{c}$, and $p_h = \gamma d_t^A|_{t_s \leq T} + (1 - \gamma)\bar{c}$ if $p_A > \bar{c}$, where $d_t^A|_{t_s \leq T}$ stands for the disagreement price for \mathcal{A} at time t given $t_s \leq T$.

Next, we consider the case in which demand shock has yet to occur ($t_s > t$). Suppose there exists a subgame perfect equilibrium (SPE) such that *at any time* $t \in [0, T]$, \mathcal{C} pays $\omega(t)$ if $t_s > t$, and \mathcal{C} can credibly wait from t to $\tau(t)$ in case of rejection. Given $W := \{\omega(t) : t \in [0, T]\}$, the price trajectory in SPE, let $J_z^t(W)$ denote \mathcal{C} 's expected payoff of buying at time $z \geq t$ when he is at time t and $t_s > t$. Thus, we have

$$\tau(t) = \sup \left\{ \arg \max_{z \in [t, T]} J_z^t(W) \right\}, \quad (1)$$

which is a mapping from $[0, T]$ to $[0, T]$. Given $\Gamma := \{\tau(t) : t \in [0, T]\}$, we can compute W accordingly. Hence, an SPE exists if there exists a fixed "point" for $\Gamma(W)$ or $W(\Gamma)$. In the following, we derive $W(\Gamma)$ and then $J_z^t(W)$.

At any time t , if $\tau(t) = t$, then it is not credible for \mathcal{C} to wait and we have that $\omega(t)$ equals the NB price given $t_s > t$: $\bar{p}_t|_{t_s > t}$. However, if $\tau(t) > t$, then \mathcal{C} would offer the lowest price that is acceptable for \mathcal{A} . Recursively and compactly, we have (see the appendix for the derivation)

$$\omega(t) = \begin{cases} \bar{p}_t|_{t_s > t}, & \tau(t) = t; \\ \left(1 - \Lambda_{\tau(t)}^t\right) \omega(\tau(t)) + \Lambda_{\tau(t)}^t [p_h + (p_A - p_h) \theta(K)], & \tau(t) > t, \end{cases} \quad (2)$$

where $\Lambda_{\tau}^t := \Pr \{t_s \leq \tau | t_s > t\} = \frac{\Lambda_{\tau} - \Lambda_t}{1 - \Lambda_t}$ is the Bayesian updated distribution of the demand shock given $t_s > t$, and $\theta(K) := 1 - \frac{Q \wedge (K-s)}{Q \wedge K} \in [0, 1]$ represents the percentage of \mathcal{C} 's procurement quantity that is threatened by the demand shock.

Now, we show how to compute $\bar{p}_t|_{t_s > t}$. If $p_A \leq \bar{c}$, we have $\bar{p}_t|_{t_s > t} = p_A$; if $p_A > \bar{c}$, the Nash bargaining leads to $\bar{p}_t|_{t_s > t} = \gamma d_t^A|_{t_s > t} + (1 - \gamma)\bar{c}$. To obtain $d_t^A|_{t_s > t}$, we first derive that $d_T^A|_{t_s > T} = \frac{\pi(K) - \pi((K-Q)^+)}{K \wedge Q}$ and $d_T^A|_{t_s \leq T} = \frac{\pi(K-s) - \pi((K-s-Q)^+)}{(K-s) \wedge Q}$. Note that $x^+ := \max\{0, x\}$. Second, if customer \mathcal{C} walks away at any time t , the seller will do the following: sell s units to other buyers if the demand shock occurs and salvage the residual capacity at time T . As shown in the

appendix, the lowest price the seller would accept is

$$d_t^A|_{t_s>t} = \Lambda_T^t [\theta(K)p_A + (1 - \theta(K)) d_T^A|_{t_s\leq T}] + (1 - \Lambda_T^t) d_T^A|_{t_s>T}. \quad (3)$$

Next, we define $p_l := \bar{p}_T|_{t_s>T}$. After obtaining $W(\Gamma)$, we proceed to get

$$J_z^t(W) = \Lambda_z^t (r - p_h) (Q \wedge (K - s)) + (1 - \Lambda_z^t) [r - \omega(z)] (Q \wedge K). \quad (4)$$

It is the weighted sum of two possible outcomes, which depend on whether demand shock occurs while \mathcal{C} waits. By definition, the threat to wait at t is credible if and only if $\tau(t) = \sup \{ \arg \max_{z \in [t, T]} J_z^t(W) \} > t$. Before we solve \mathcal{C} 's optimal timing problem and the SPE in the next section, we offer a preliminary observation.

Proposition 1. *NB price $\bar{p}_t|_{t_s>t}$ is decreasing in t if $\theta(K)p_A + (1 - \theta(K)) d_T^A|_{t_s\leq T} > d_T^A|_{t_s>T}$, increasing in t if $\theta(K)p_A + (1 - \theta(K)) d_T^A|_{t_s\leq T} < d_T^A|_{t_s>T}$, and time-invariant if otherwise.*

Intuitively, it is not credible for \mathcal{C} to wait if $\bar{p}_t|_{t_s>t}$ is increasing or invariant in t , considering the possibility of a demand shock that depletes \mathcal{A} 's capacity. It can be checked that $\theta(K) = 0$ if $K > s + Q$. Hence, if \mathcal{A} has sufficient capacity and $d_T^A|_{t_s\leq T} > d_T^A|_{t_s>T}$, then $\bar{p}_t|_{t_s>t}$ is decreasing in t and it may be credible for \mathcal{C} to wait. Otherwise, p_A will play a role in determining the dynamics of $\bar{p}_t|_{t_s>t}$. If capacity is moderate but p_A is high, it may also be credible for \mathcal{C} to wait.

4.2 The Customer's Problem

Since we assume \mathcal{C} 's purchase quantity is fixed, his objective is to maximize his expected payoff, $J_\tau^t(W)$, given P and W by optimizing τ . Assuming $\omega(t)$ is differentiable, we have the first order derivative

$$\frac{d}{d\tau} J_\tau^t(W) = \frac{\lambda_\tau \cdot (Q \wedge K)}{1 - \Lambda_t} \left[\omega(\tau) - \omega'(\tau) \cdot \frac{1 - \Lambda_\tau}{\lambda_\tau} - [1 - \theta(K)] \cdot p_h - \theta(K) \cdot r \right]. \quad (5)$$

If the maximum of $J_\tau^t(W)$ occurs at $\tau^* \in (t, T)$, then we must have $\frac{d}{d\tau} J_\tau^t(W)|_{\tau=\tau^*} = 0$. If multiple maximizers are present in (t, T) , then by definition the supremum of the set of solutions, τ^* , is the waiting destination. Since $\frac{\lambda_\tau \cdot (Q \wedge K)}{1 - \Lambda_t} > 0$, the first order condition can be reduced to

$\omega(\tau) - \omega'(\tau) \cdot \frac{1-\Lambda_\tau}{\lambda_\tau} = [1 - \theta(K)] \cdot p_h + \theta(K) \cdot r$, which indicates that τ^* is irrelevant of t . If the maximizer is not in (t, T) , then it is either $\tau(t) = t$ or $\tau(t) = T$. In any case, as long as it is not optimal to buy now, τ^* is irrelevant of t . In other words, $\frac{d}{dt}\tau^* \equiv 0$. Thus, if any offer is rejected at any time $t < \tau^*$, \mathcal{C} would wait until τ^* . Although it is possible that for certain $t' > \tau^*$ we have $\tau(t') > t'$ when $J_\tau^t(W)$ is not monotone in τ , it is not relevant when we stand at $t \leq \tau^*$. Incorporating the above observations, we obtain the following important result.

Theorem 1. *In any SPE, $\frac{\partial \omega(t)}{\partial \tau^*} = 0$ and thus $\omega(t) \equiv \bar{p}_t|_{t_s > t}$.*

The theorem says that \mathcal{C} always pays the NB price no matter when he buys and he cannot gain advantage by waiting. Based on this theorem, we obtain $W := \{\omega(t) : t \in [0, T]\}$, and can solve for τ^* by plugging W into (5). In particular, we have $\frac{d}{d\tau} J_\tau^t(W) = \frac{\lambda_\tau}{1-\Lambda_t} \cdot (Q \wedge K) \cdot \theta(K) \cdot (p_A \wedge \bar{c} - r)$ when we plug $\omega'(\tau)$ into (5). Accordingly, we characterize SPEs as follows.

Corollary 1. *If $K \geq Q + s$, then $\frac{d}{d\tau} J_\tau^t(W) = 0$ and thus $\tau(t) = T$; if $K < Q + s$ and $p_A \wedge \bar{c} < r$, then $\tau(t) = t$; if $K < Q + s$ and $p_A \wedge \bar{c} \geq r$, then $\tau(t) = T$.*

The result shows that if \mathcal{A} 's capacity is absolutely sufficient (i.e., $K \geq Q + s$), then \mathcal{C} can credibly wait at any time. It is because the capacity that is needed by \mathcal{C} is not under any threat. Mathematically, when $K \geq Q + s$ we have $\frac{d}{d\tau} J_\tau^t(W) = 0$ and \mathcal{C} 's expected purchase cost is a constant ex ante. Otherwise, the credibility of \mathcal{C} 's threat largely depends on the relative values of the posted price p_A , the marginal cost of the alternative product \bar{c} , and the marginal revenue r . If the end product is very profitable for the customer, then the threat of waiting is never credible because the customer would have huge loss if the capacity is reduced. If the product is not very profitable such that the profit would be negative when paying the posted price or when switching to the alternative product, then getting a greater discount is so important that waiting till the last minute is credible. In short, the SPE depends greatly on the capacity level and the level of the customer's profitability.

4.3 The Seller's Problem

Given the SPE, seller \mathcal{A} then posts p_A at $t = 0$ and decides on how to react to \mathcal{C} 's initial counter-offer in order to maximize her expected revenue in equilibrium, Π_A . In this section, we analyze the

best reaction to \mathcal{C} 's counter-offer. Let Π_A^0 and Π_A^T denote \mathcal{A} 's expected revenue when \mathcal{C} decides to buy at time 0 and T , respectively, given $\{K, Q, s\}$. We have

$$\begin{aligned} \Pi_A^0 &= \bar{p}_0 \cdot K \wedge Q + \Lambda_T \cdot [p_A \cdot s \wedge (K - Q)^+ + \pi((K - s - Q)^+)] \\ &\quad + \bar{\Lambda}_T \cdot \pi((K - Q)^+), \text{ and} \end{aligned} \quad (6)$$

$$\begin{aligned} \Pi_A^T &= \Lambda_T [p_A \cdot s \wedge K + p_h \cdot Q \wedge (K - s) + \pi((K - s - Q)^+)] \\ &\quad + \bar{\Lambda}_T [p_l \cdot K \wedge Q + \pi((K - Q)^+)]. \end{aligned} \quad (7)$$

Through algebraic manipulations, we find the following important result:

$$\Pi_A^0 - \Pi_A^T = \begin{cases} 0 & \text{if } p_A \leq \bar{c}; \\ (\bar{c} - p_A) \cdot (1 - \gamma) \cdot \Lambda_T \cdot (K \wedge Q) \cdot \theta(K) & \text{if } p_A > \bar{c}. \end{cases} \quad (8)$$

Therefore, we always have $\Pi_A^0 \leq \Pi_A^T$, which is surprising. In other words, within our modelling framework, \mathcal{A} should never strictly prefer that they close the deal at time 0. The intuition is that *splitting the entire pie* (i.e., the total expected payoff generated by the full capacity) *at time 0 is worse for \mathcal{A} than splitting the residual pie* (i.e., the total expected payoff generated by the residual capacity) *at time T* . In particular, \mathcal{A} would prefer that \mathcal{C} waits until T if $\Pi_A^0 < \Pi_A^T$, which is possible when $p_A > \bar{c}$ and $\theta(K) > 0$. This is because \mathcal{C} would not pay more than \bar{c} , and \mathcal{C} 's waiting awards \mathcal{A} the opportunity to gain higher profit and thus better utilize the capacity. Furthermore, it is possible to let \mathcal{C} wait when his threat is credible. Therefore, when $\Pi_A^0 < \Pi_A^T$ and $\tau(0) = T$, the best reaction for seller \mathcal{A} is to reject \mathcal{C} 's initial counteroffer and encourage him to wait. We restate this result in the following Corollary and summarize all the results in Table 2.

Corollary 2. *When $p_A > \bar{c} \geq r$ and $K < s + Q$, the best reaction for seller \mathcal{A} is to reject \mathcal{C} 's initial counteroffer and encourage him to wait.*

Now we know that the credibility of customers' threat to wait can actually sometimes benefit the seller. To help the seller exploit this result, we explicitly provide the bargaining tactics in Table 2, which points out the best actions for the seller to take in different situations.

- When the end product is very profitable for \mathcal{C} (i.e., $p_A \wedge \bar{c} < r$), it is not credible for him to

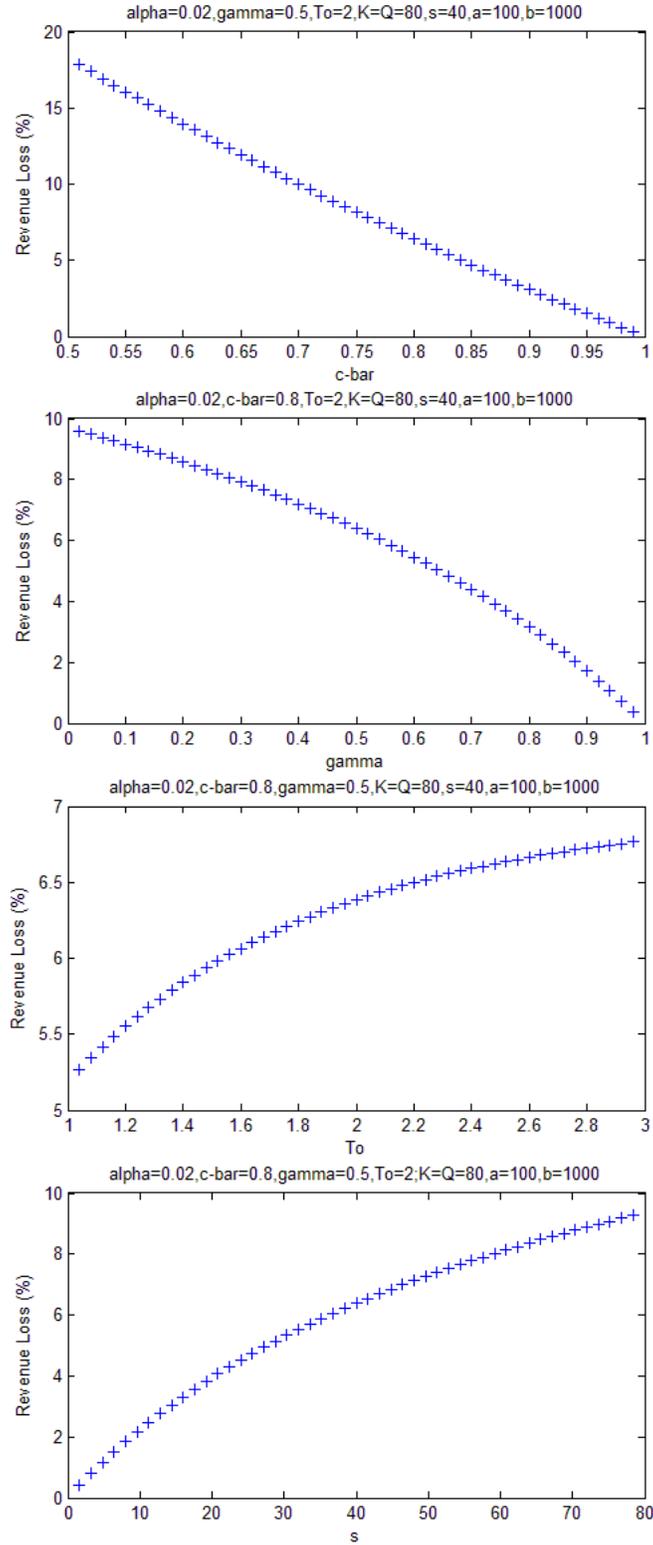
Table 2: The Time Preference versus the Credibility of Waiting

Scenarios	$p_A \leq \bar{c}$ or $K \geq s + Q$ (i.e., component product has low price or high capacity)	$p_A > \bar{c}$ and $K < s + Q$ (i.e., component product has high price and low capacity)
$p_A \wedge \bar{c} < r$ (i.e., final product has high margin)	$\Pi_A^0 - \Pi_A^T = 0$ and $\tau(0) = 0$ Best action for \mathcal{A} : Negotiating at $t = 0$	$\Pi_A^0 - \Pi_A^T < 0$ and $\tau(0) = 0$ Best action for \mathcal{A} : Negotiating at $t = 0$
$p_A \wedge \bar{c} \geq r$ (i.e., final product has low margin)	$\Pi_A^0 - \Pi_A^T = 0$ and $\tau(0) = T$ Best action for \mathcal{A} : Accepting $\omega(0)$ or negotiating at $t = T$	$\Pi_A^0 - \Pi_A^T < 0$ and $\tau(0) = T$ Best action for \mathcal{A} : Rejecting any counteroffer and negotiating at $t = T$

wait and thus it is not possible for \mathcal{A} to ask him to wait. As a result, the seller should engage in the price bargaining immediately.

- When the end product is not very profitable and switching is costly for \mathcal{C} (i.e., $r \leq p_A \leq \bar{c}$), it is credible for him to wait and \mathcal{A} can either accept $\omega(0)$ as the initial counteroffer or wait till the last minute.
- When the capacity is not absolutely sufficient (i.e., $K < s + Q$), the end product is not very profitable, and switching is not costly (i.e., $r \leq \bar{c} < p_A$), it is credible for \mathcal{C} to wait and \mathcal{A} should reject any counteroffer and wait till the last minute. In this case, if the seller does not encourage the buyer to wait but settle the price at the beginning, we show by numerical studies that the seller can lose more than 10% of its revenue. The results are summarized in Figure 3. We can see that the loss decreases with \bar{c} because the larger \bar{c} the higher price \mathcal{C} pays at the beginning. The loss also decreases with γ because the larger γ the more the revenue depends on \mathcal{A} 's own outside options but not the negotiated price. In addition, the loss increases with T_o and s , because the larger the values the higher expected revenue \mathcal{A} can obtain from the price-takers.

Figure 3: Revenue Loss for Suboptimal Bargaining Strategy



Notice that the results presented here are important not only because they tell the seller how to negotiate with the customer in different situations, but also because they show the seller what to expect *ex ante*. As we will see later, the optimal posted price depends on the timing of negotiation and thus failing to anticipate the right outcome leads to ineffective pricing.

5 Non-Monotonic Price-Quantity Relation

In this section, we explore the property of the negotiated price in equilibrium. Zhang et al. (2015) point out that the negotiated price can be a non-monotonic function of the purchase quantity in B2B markets. They find some empirical evidences in the semiconductor industry and they offer a model to explain that phenomenon. They consider the case wherein buyers don't wait. Here, we show that our model can predict a similar price-quantity relation, but we consider the case wherein buyers can strategically wait during negotiations.

5.1 Decomposition of the Negotiated Price

According to Theorem 1 and Corollary 1, customer \mathcal{C} would either buy at $t = 0$ or $t = T$ in equilibrium. Hence, there could be three possible outcomes for the negotiated price: $\bar{p}_0|_{t_s > 0}$, $\bar{p}_T|_{t_s > T}$, and $\bar{p}_T|_{t_s \leq T}$. In the following, we check how each of them changes with Q . For simplicity, we assume that the value of the residual capacity $\pi(\cdot)$ is increasing concave. However, notice that concavity is not required to obtain any of the previous results.

We use the following parameter setting. The capacity level K is fixed. To make sure that the seller's threat is effective, we assume she picks an s such that $K < Q + s$. (Note that we assume the total size of the demand shock is large enough.) We set $Q + s = 100$. In addition, we set $p_A > \bar{c}$ and $\pi(x) = \max_p p \cdot \min\{x, a - bp\}$. In preparation, we define

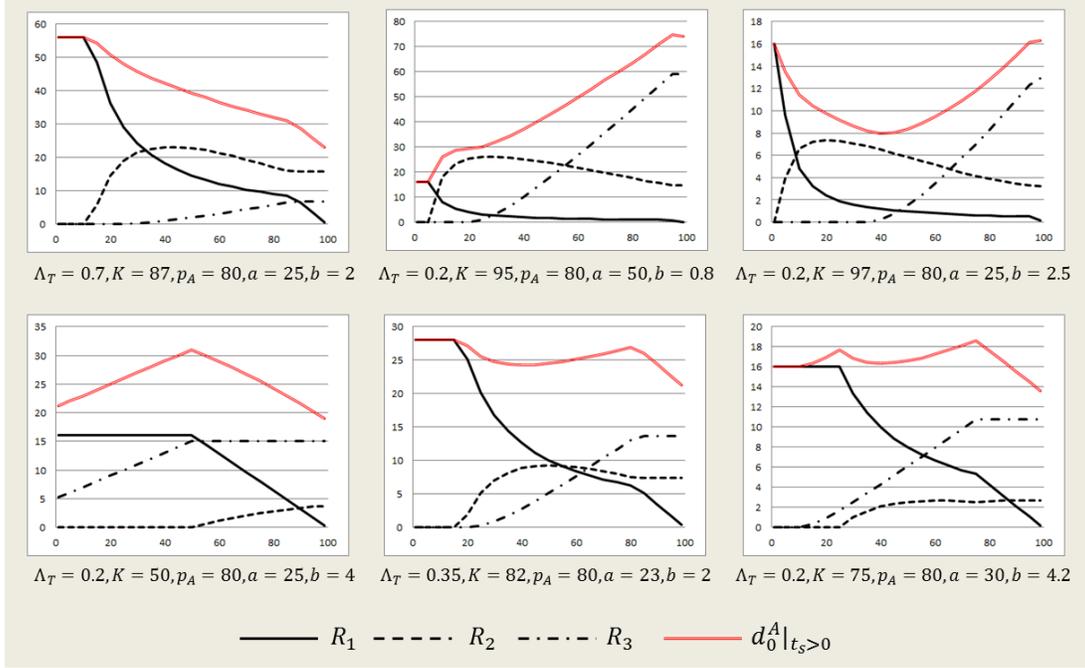
$$R_1 = \Lambda_T \cdot p_A \cdot \theta(K) = \Lambda_T \cdot p_A \cdot \frac{K \wedge Q - (K - s) \wedge Q}{K \wedge Q}; \quad (9)$$

$$R_2 = \Lambda_T \cdot d_T^A|_{t_s \leq T} \cdot [1 - \theta(K)] = \Lambda_T \cdot \frac{\pi(K - s) - \pi((K - s - Q)^+)}{K \wedge Q}; \quad (10)$$

$$R_3 = (1 - \Lambda_T) \cdot d_T^A|_{t_s > T} = (1 - \Lambda_T) \cdot \frac{\pi(K) - \pi((K - Q)^+)}{K \wedge Q}. \quad (11)$$

According to (3), $d_0^A|_{t_s > 0}$ can be expressed as the sum of R_1 , R_2 , and R_3 . Intuitively, if the

Figure 4: The Impact of Purchase Quantity Q

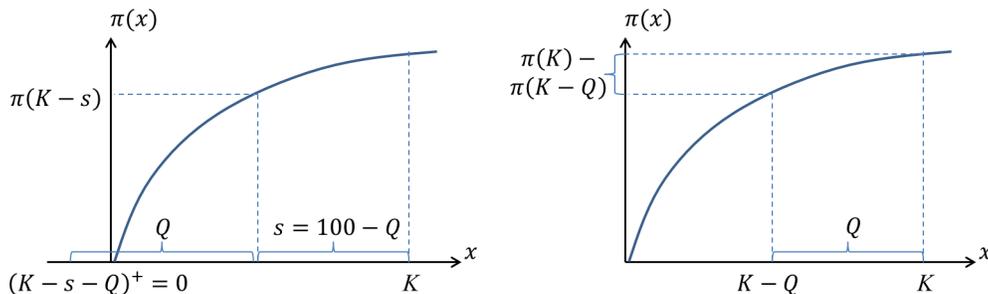


bargaining breaks down and customer \mathcal{C} walks away, there could be three possible outcomes for the amount of capacity $K \wedge Q$ that could be sold to customer \mathcal{C} . First, if the demand shock occurs, a fraction $\theta(K)$ of it will be sold to the price takers. Second, if the demand shock occurs, a fraction $1 - \theta(K)$ of it will be salvaged at time T . Third, if the demand shock does not occur, all of it will be salvaged at time T . The expected revenue obtained from each outcome is R_1 , R_2 , and R_3 , respectively. As a result, we can write

$$\bar{p}_0|_{t_s>0} = (1 - \gamma) \cdot \bar{c} + \gamma \cdot (R_1 + R_2 + R_3). \quad (12)$$

Therefore, how $\bar{p}_0|_{t_s>0}$ is influenced by Q depends on how R_1 , R_2 , and R_3 change with Q . In Figure 4, we plot $d_0^A|_{t_s>0}$, R_1 , R_2 , and R_3 against Q . We can see that R_1 is decreasing in Q . This is intuitive because as Q increases, the seller can use a smaller s to create a sense of “scarcity” for her threat and the revenue from the price takers will be lower. Next, according to Figure 4, R_2 and R_3 are not decreasing in Q , which causes the non-monotonicity of the negotiated price. The reason is that π is increasing and concave: as the residual capacity increases, the seller has more and more

Figure 5: Illustration of Increments in Salvage Value



flexibility to optimize the salvage value, but only up to a certain scale. Hence, as Q increases, the marginal salvage value diminishes and the seller's opportunity cost in this dimension increases. Figure 5 illustrates the effect of Q on $\pi(K - s) - \pi((K - s - Q)^+)$ and $\pi(K) - \pi((K - Q)^+)$ while holding $Q + s = 100$ constant.

Now let's consider the negotiated price at time T . Under the assumption of $K < s + Q$, we have $\bar{p}_T|_{t_s > T} = (1 - \gamma) \cdot \bar{c} + \gamma \cdot R_3 / (1 - \Lambda_T)$ and $\bar{p}_T|_{t_s \leq T} = (1 - \gamma) \cdot \bar{c} + \gamma \cdot \pi(K - s) / (K - s)$. Hence, we know immediately that $\bar{p}_T|_{t_s > T}$ is increasing (because R_3 is increasing) in Q and $\bar{p}_T|_{t_s \leq T}$ is decreasing in Q . They can actually be viewed as special cases of $\bar{p}_0|_{t_s > 0}$. This also suggests that only transactions for which the prices are negotiated at the beginning can display a non-monotonic price-quantity relation.

5.2 Explorations

If our model is correct, we can use it to explore possible price-quantity relations by setting different values for $\{\Lambda_T, K, p_A, a, b, \gamma\}$. After extensive numerical experiments, we show in Figure 4 that six patterns describe the price curve: (1) decreasing, (2) increasing, (3) V-shaped, (4) Λ -shaped, (5) \smile -shaped, and (6) M-shape.

The six curves are all linear combinations of R_1 , R_2 and R_3 . Key factors determining the final shape of the curves are K (capacity level), Λ_T (probability of the demand shock occurring prior to T) and π (salvage value determined by a and b). Specifically, K determines the location of the "kinked" points on R_1 , R_2 and R_3 , while Λ_T and π determine the relative weight of each component. The irregularities complicate the bargaining process and make the bargaining outcome hard to predict

without an analytic model.

Considering the intuition behind these patterns, we notice that Λ_T literally represents the probability of the demand shock. It is actually a measure of the number of potential customers for the product; the greater number of potential customers, the greater the probability of demand shock during the selling season. As illustrated by the first graph in Figure 4, when Λ_T is high ($= 0.7$) and π is small, responding to demand shock is the major outside option for the seller. In the second graph, Λ_T is low ($= 0.2$) but π is large, so R_3 , the value of selling to the salvage market after T , dominates. In the third graph, both Λ_T and π are small, so that neither of the two sources of revenue dominates, and the mixed effect generates a non-monotonic curve. The explanations for the other three patterns are similar but more complex because the effect of K is involved.

6 Posted Price Optimization

Given that the output of the negotiation could be complicated, it should be difficult to make decisions that would influence the negotiated price. In this section, we investigate the impact of the timing of price negotiation. For simplicity, we focus on the case of selling to one bargainer who is also the only major buyer. Later, we will consider the case of selling to multiple bargainers.

We focus on the case wherein the cost of the alternative product is low enough for the bargainer so that the posted price is never taken. We know from our analysis that the bargainer would settle the price at either the beginning or the deadline in equilibrium. In either case, the seller's expected revenue can be written as $C^I + C^{II} \cdot \Lambda_T + C^{III} \cdot p_A \cdot \Lambda_T$, where C^I , C^{II} , and C^{III} are constants that are independent of p_A but depend on the timing of negotiation. Particularly, it is easy to check that $C^{II} < 0$ and $C^{III} > 0$. Then, by restricting Λ to be log-concave, we can produce the following result. Note that the condition of $\Lambda(t)$ being log-concave in t is not restrictive. Many common distributions such as normal, logistic, uniform, and exponential distributions satisfy this condition.

Proposition 2. *If $\Lambda(t)$ is log-concave in t , the optimal price p_A^* uniquely solves*

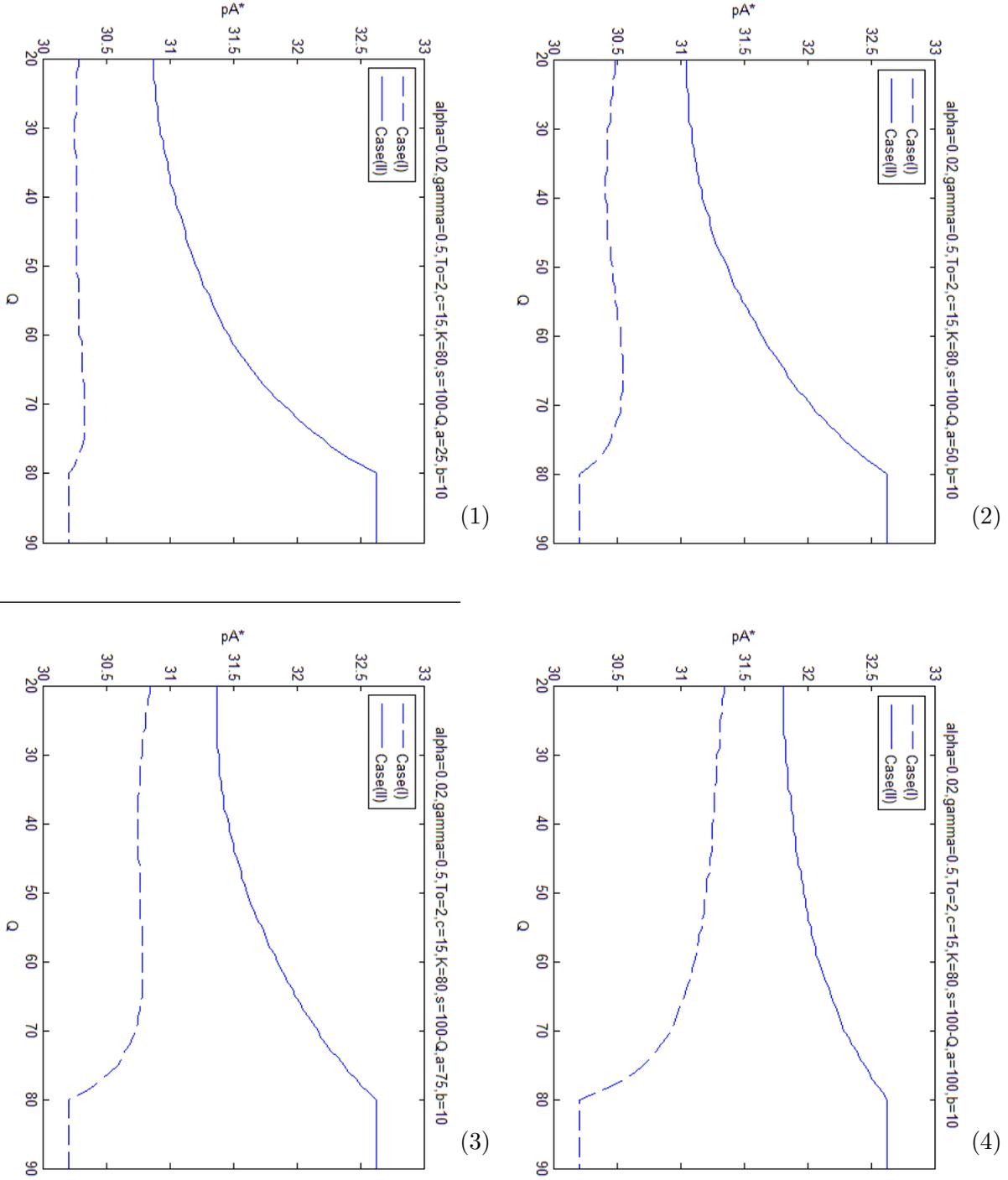
$$p_A = \frac{\Lambda(T(p_A))}{\lambda(T(p_A))} \cdot \frac{1}{\alpha T_o} - \frac{C^{II}}{C^{III}}. \quad (13)$$

What can we learn from this model? How could things go wrong if managers mistakenly

estimated the timing of negotiation and the negotiated prices? How would the posted price be affected? Here we numerically explore how the optimal posted price should be when the bargainer settles the price at different time points. In case (I), the price is negotiated at $t = 0$. In case (II), the price is negotiated at $t = T$. We also consider four different cases for the residual value function $\pi(x) = \frac{1}{b} \cdot (a - x \wedge \frac{a}{2}) \cdot (x \wedge \frac{a}{2})$ by setting four different values for a which measures the size of the residual market. In addition, we assume that Q is known when setting the price. In this way, we rule out the impact of uncertainty in the end market. We can also view such an optimal price as “hindsight-optimal”. Lastly, we fix the capacity level relative to the total potential demand to be $K/(Q + s) = 0.8$. The results of our numerical study are summarized in Figure 6. We have the following observations.

First of all, the optimal posted price should depend on the purchase quantity of the bargainer and the timing of agreement. In particular, the quantity can impact the optimal posted price in different ways, depending on when the agreements are reached. If the major buyer waits, the optimal posted price increases with the quantity; otherwise, the price should decrease with the quantity in general, although sometimes the relation could be non-monotonic. As s becomes smaller and price-takers become less important, the seller should focus more on the price bargaining. Recall that the seller splits the entire pie with the bargainer if they bargain at time 0. If the bargainer doesn’t wait, the seller should reduce the price as s shrinks in order to maximize the pie. If the bargainer waits, the seller should increase the price in order to maximize the expected revenue from price-takers. In addition, note that only the residual capacity will be sold to the price-takers if the bargainer makes the purchase at the beginning; otherwise, more capacity will be sold to the price-takers. Hence, the optimal price for case (II) should be higher than in case (I), although the chance of demand shock will be reduced. From the results, we can learn that failing to anticipate the right timing of negotiation will lead to ineffective pricing. In particular, if the seller doesn’t notice that she can benefit from customers’ waiting and always anticipates settling the price at time 0, then she may underprice the product.

Figure 6: The Optimal Posted Price Versus Purchase Quantity



7 Bargaining with Multiple Buyers

The base model is reasonable if the major buyers arrive sequentially and buyers do not factor in decisions of other buyers. Now, we will consider a scenario in which the seller negotiates with multiple buyers who consider each other's decision. In this case, we assume that the seller does not rely on the price-takers to implement her threat; instead, she can utilize the competition for the limited capacity among the major buyers. We will characterize a Nash equilibrium in this scenario.

We know from the base model that the credibility of a customer's waiting threat depends on K , $Q + s$, p_A , \bar{c} , and r . In this modified model, s represents the total demand from other customers. Therefore, the credibility of threat only depends on the total demand $\mathbb{T}\mathbb{D}$ but not individual demand Q . Hence, it is natural to focus on an equilibrium wherein all customers buy at the same time. Now suppose all other customers buy at time τ . It is equivalent to setting $\lambda(t) = 0$ for $\forall t \in [0, \tau) \cup (\tau, T]$ in the base model. Then according to Corollary 1, customer i is indifferent to $\forall t \in [0, T]$ if $K \geq \mathbb{T}\mathbb{D}$. Or, if $K < \mathbb{T}\mathbb{D}$ and $p_A \wedge \bar{c} < r$, then customer i can buy at any $t \in [0, \tau]$, because $\lambda(t) = 0$ for $\forall t \in [0, \tau)$; thus, $\frac{d}{dt} J_t^0 = \lambda_t(Q \wedge K)\theta(K)(p_A \wedge \bar{c} - r) = 0$. If $p_A \geq p_B$, similar arguments suggest that customer i can buy at any $t \in [\tau, T]$. Consequently, we have the following result.

Lemma 1. *If all other customers buy at τ , buying at τ is weakly dominant for i in any case.*

Since time discounting is ignored in our problem, the result is identical for any $\tau \in [0, T]$. Without loss of generality, we focus on the equilibrium that all strategic customers buy at time $\tau = 0$, and such a Nash equilibrium exists.

Lemma 2. *All customers buying at time 0 constitutes a Nash equilibrium.*

In this scenario, the dependence of T on p_A is meaningless because the demand shock is no longer relevant. As a result, we model the dependence of demand on p_A in the following way: the distribution of $\mathbb{T}\mathbb{D}$ under p'_A first order stochastically dominates the distribution under p''_A if $p'_A < p''_A$. Additionally, we assume $\pi(x) = \max_p p \cdot (x \wedge [\mathbb{T}\mathbb{D} \cdot (a - bp)])$. Then, the seller's pricing

problem can be described as follows:

$$\max_{p_A} \quad \mathbf{E} [(K \wedge \mathbb{T}\mathbb{D}) \cdot \bar{p} + \pi ((K - \mathbb{T}\mathbb{D})^+)] \quad (14)$$

$$s.t. \quad f_i = Q_i / \mathbb{T}\mathbb{D}, \quad (15)$$

$$\bar{p} = \sum f_i \cdot p_i, \quad (16)$$

$$p_i = (1 - \gamma_i) \cdot \bar{c} + \gamma_i \cdot w_{-i}^A \quad \forall i \in I_C, \quad (17)$$

$$w_{-i}^A = \bar{p}_{-i} \left(1 - \frac{[K - \sum_{j \neq i} Q_j]^+ \wedge Q_i}{K \wedge Q_i} \right) + \frac{\pi \left([K - \sum_{j \neq i} Q_j]^+ \right) - \pi ([K - \mathbb{T}\mathbb{D}]^+)}{K \wedge Q_i} \quad \forall i \in I_C, \quad (18)$$

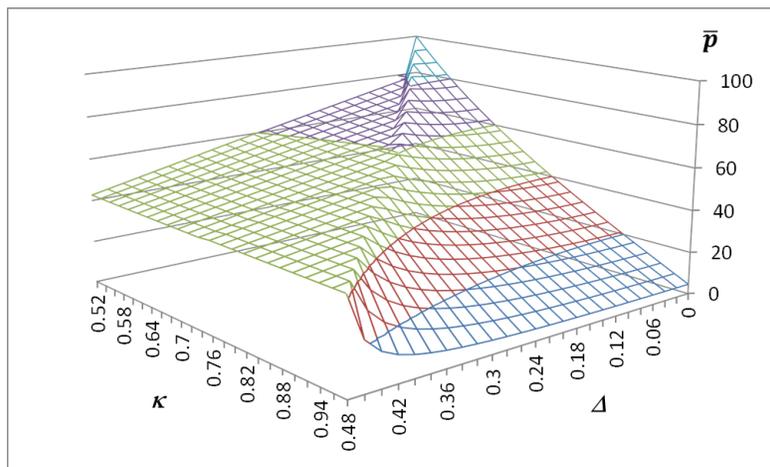
$$\bar{p}_{-i} = \sum_{j \neq i} Q_j \cdot p_j / \sum_{j \neq i} Q_j \quad \forall i \in I_C. \quad (19)$$

The expectation in (14) is with regard to $\mathbb{T}\mathbb{D}$. Parameter \bar{p} is the average price received by all customers, as defined in (16). Next, (17) describes the Nash bargaining price for customer i , which depends on the bargaining outcomes for all other customers. The seller's disagreement price (w_{-i}^A) in (18) is based on (3), in which $\Lambda_T = 1$ and p_A is replaced by \bar{p}_{-i} , the average price received by all other customers. An implicit assumption is that every customer receives the same service level when capacity is insufficient. Together, constraints (16) through (19) determine the Nash equilibrium given p_A . Apparently, there is no closed form solution for this problem, though the optimal price, p_A^* , can be determined using simulations.

In search of managerial insights, we consider a special case concerning two customers. To rule out the effect of exogenous bargaining power but focus on the impact of demand share, we consider two customers with equal exogenous bargaining power, γ , but different demand shares, f_1 and f_2 , which are known in advance. Furthermore, we introduce $\Delta \in [0, 0.5)$ to measure the demand-share asymmetry between the two customers and, without loss of generality, let $f_1 = 0.5 + \Delta$ and $f_2 = 0.5 - \Delta$. Therefore, we can write $\bar{p} = f_1 p_1 + f_2 p_2$, and we proceed to see how \bar{p} is affected by Δ and $\kappa = K / \mathbb{T}\mathbb{D}$, the measure of capacity level. Based on our formulation, the space of capacity level can be divided into three segments given f_1 and f_2 :

- High level is defined as $\kappa \geq 1$, which means the capacity is sufficient to satisfy both customers.

Figure 7: Relationship among \bar{p} , Δ , and κ



- Medium level is defined as $\max\{f_1, f_2\} \leq \kappa < 1$, which means the capacity is sufficient to satisfy the larger customer.
- Low level is defined as $\kappa < \max\{f_1, f_2\}$, which means the capacity can not satisfy the larger customer.

As we show, the impact of Δ on \bar{p} is closely related to κ .

Proposition 3. *If $\kappa \geq 1$, then \bar{p} is decreasing in $\Delta \geq 0$.*

This result says that, *when the manufacturer's capacity level is high, she can benefit from customer demand share symmetry.* This is because the salvage value function is increasing concave, and any increase in asymmetry in customer demand shares will drag down the value of the manufacturer's outside option.

Next, we check the impact of Δ on \bar{p} in the cases of medium and low capacity levels; closed-form solutions for the equilibrium prices (p_1^e and p_2^e) are provided in the appendix. Since it is difficult to see analytically how \bar{p} changes with Δ when the capacity level is medium or low, we use computational study to explore the relationships among \bar{p} , Δ , and κ . In Figure 7, the 45 degree line corresponds to $\kappa = f_1$. It can be seen that \bar{p} strictly decreases with Δ when $\kappa < f_1$ but increases with Δ when $\kappa \geq f_1$. In other words, *demand-share asymmetry benefits the manufacturer when her capacity level is medium, while symmetry benefits her when her capacity level is low.* Changes in (18), which describes the manufacturer's outside option, cause the directional changes in the

impact of Δ . The reason for this is that there generally is a kinked point on the price curve at $f_1 = \kappa$, and the price received by customer 1 (p_1^e) decreases faster with f_1 when the capacity level is low than when the capacity level is medium. Meanwhile, the price received by customer 2 (p_2^e) generally increases as f_2 decreases. Consequently, the increase of p_2^e has larger impact on \bar{p} when $\kappa \geq f_1$, while the decrease of p_1^e has larger impact on \bar{p} when $\kappa < f_1$.

Knowing that customer demand-share asymmetry can benefit or harm the seller under different capacity levels, we discuss the effect of asymmetry on the seller's optimal price, $p_A^*(\Delta)$. We know that in this modified model, p_A only affects the total demand and thus equivalently the level of capacity. According to Figure 7, we know that \bar{p} in general decreases in the capacity level, regardless of Δ . Hence, the optimal posted price does not depend on demand-share asymmetry.

8 Summary and Insights

In this paper, we build a dynamic bargaining model with stylized components to study the dynamics of price negotiation in B2B markets. In particular, buyers can make counteroffers and threaten to wait if their prices are rejected by the seller; as a countermeasure, the seller can threaten to sell the capacity to other buyers. We solve the subgame perfect Nash equilibrium for the dynamic game and characterize the credibility of buyers' threat of waiting as well as the negotiated price in different situations. Further, we analyze the seller's payoff in various situations and obtain the optimal bargaining strategy for the seller. In addition, we use the model to show that both the negotiated price and the optimal posted price can be a non-monotonic function of the customer's purchase quantity. By extending the model to the case of bargaining simultaneously with two customers, we showed that asymmetry in customer size could either benefit or harm the seller, depending on her capacity level.

Theoretic Contribution

Our model is different from the traditional cooperative bargaining models and strategic bargaining models. Cooperative models are static and unable to capture the process of negotiation. Traditional strategic models incorporate the process of negotiation, but they normally predict immediate agreements given symmetric information; furthermore, they normally do not incorporate

strategic threats commonly used in practical negotiations. In contrast, our bargaining model allows us to capture the asymmetric threats used by different parties in a dynamic bargaining process. The model predicts delay of agreements without assuming information asymmetry or exogenous delay of counteroffers. In the traditional alternating-offer models, a player has to wait till the next period to propose a counteroffer. Exogenous delay of counteroffers can lead to trivial delay of agreements because the player who proposes the counteroffer at the last moment has full bargaining power. However, we assume that it takes no time to propose a counteroffer unless the buyer decides to wait purposely, so the delay of agreements in our model is fully due to customers' threats to wait.

Implications for Bargaining Strategy

We learn from our model that the credibility of customers' threat to wait can actually sometimes benefit the seller. This is true when (1) the available capacity is not absolutely sufficient (i.e., not sufficient to satisfy both the focal buyer and the occasional demand), (2) the end product is not very profitable for the customer, and (3) switching is not costly. In this case, it is credible for the customer to wait and the seller should reject the counteroffer and encourage the customer to wait till the last minute. In this way, the seller can obtain higher expected payoff. Although the result is surprising, it is reasonable in that the three conditions indicate that it is a "mismatch" for the seller and the customer: the product is not profitable for the customer and the customer does not rely heavily on the seller. Knowing such results can make a huge difference for sales managers. In particular, if the seller does not know when to encourage the buyer to wait and always settle the price at the beginning, we show that the seller can lose more than 10% of its revenue.

Implications for Posted Pricing

We show that the optimal posted price depends on the timing of negotiation and thus failing to anticipate the right outcome leads to ineffective pricing. In particular, if the seller doesn't notice that she can benefit from customers' waiting and always anticipates settling the price at time 0, she may underprice the product. As shown by our numerical examples, the price can be set at 8% lower than the optimal level. In addition, given that the price-quantity relationship can be non-monotonic, assuming a simple structure can cause serious problems for the seller when making important decisions *ex ante*.

Future Research

In future research, it might be interesting to model customers' cost of switching as a stochastic process, which is a common consideration when competing sellers may drop prices and new technology may be frequently launched. Another interesting problem is that whether the seller should set the posted price and let buyers offer the first price that sets the anchor for the negotiation. There have been different opinions in practice regarding this issue, it might be a fruitful area for research.

References

- [1] Admati, A.R., M. Perry. 1987. Strategic Delay in Bargaining. *Review of Economic Studies* 54(3) pp.345–364.
- [2] Aydin, G., H.S. Heese. 2015. Bargaining for an Assortment. *Management Science* 61(3) pp.542–559.
- [3] Binmore, K., A. Rubinstein, A. Wolinsky. 1986. The Nash Bargaining Solution in Economic Modelling. *RAND J. of Economics* 17(2) pp.176–188.
- [4] Cramton, P.C. 1984. Bargaining with Incomplete Information: An Infinite-Horizon Model with Two-Sided Uncertainty. *Review of Economic Studies* 51(4) pp.579–593.
- [5] Cramton, P.C. 1992. Strategic Delay in Bargaining with Two-Sided Uncertainty. *Review of Economic Studies* 59(1) pp.205–225.
- [6] Dukes, A.J., E. Gal-Or, K. Srinivasan. 2006. Channel Bargaining with Retailer Asymmetry. *Journal of Marketing Research* 43(1) pp.84–97.
- [7] Feng, Q., G. Lai, L.X. Lu. 2015. Dynamic Bargaining in a Supply Chain with Asymmetric Demand Information. *Management Science* 61(2) pp.301–315.
- [8] Friedenber, A. 2014. Bargaining under Strategic Uncertainty. Working paper, Arizona State University.

- [9] Garvin, D.A. 1991. Boeing 767: From Concept to Production (A). Harvard Business School Case 688-040.
- [10] Gill, D., J. Thanassoulis. 2009. The Impact of Bargaining on Markets with Price Takers: Too Many Bargainers Spoil the Broth. *European Economic Review* 53(6) pp.658–674.
- [11] Grennan, M. 2013. Price Discrimination and Bargaining: Empirical Evidence from Medical Devices. *American Economic Review* 103(1) pp.145–177.
- [12] Guo, L., G. Iyer. 2013. Multilateral Bargaining and Downstream Competition. *Marketing Science* 32(3) pp.411–430.
- [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] Ma, C.A., M. Manove. 1993. Bargaining with Deadlines and Imperfect Player Control. *Econometrica* 61(6) pp.1313–1339.
- [16] Nagarajan, M., Y. Bassok. 2008. A Bargaining Framework in Supply Chains: The Assembly Problem. *Management Science* 54(8) pp.1482–1496.
- [17] Nagarajan, M., G. Sošic. 2008. Coalition Stability in Assembly Models. *Operations Research* 57(1) pp.131–145.
- [18] Nash, J. 1950. The Bargaining Problem. *Econometrica* 18(2) pp.155–162.
- [19] Rubinstein, A. 1982. Perfect Equilibrium in a Bargaining Model. *Econometrica* 50(1) pp.97–110.
- [20] Roth, A.E., J.K. Murnighan, F. Schoumaker. 1988. The Deadline Effect in Bargaining: Some Experimental Evidence. *American Economic Review* 78 pp.806–823.
- [21] Zhang, W., S. Dasu, R. Ahmadi. 2015. Higher Prices for Larger Quantities? Non-Monotonic Price-Quantity Relations in B2B Markets. Working paper, University of California, Los Angeles.

Appendix

Derivation of $\omega(t)$ Given $\tau(t) > t$: If \mathcal{A} accepts \mathcal{C} 's offer w at t , then

$$\Pi_A^{accept} = (K \wedge Q)w + \hat{V} [(K - Q)^+],$$

wherein $\hat{V} [(K - Q)^+] = \Lambda_\tau^t [s \wedge (K - Q)^+ p_A + \pi((K - s - Q)^+)] + (1 - \Lambda_\tau^t) \cdot v_\tau((K - Q)^+)$. If \mathcal{A} rejects, then \mathcal{C} would wait until τ and a deal is made in equilibrium. At t ,

$$\begin{aligned} \Pi_A^{reject} &= \Lambda_\tau^t [(s \wedge K) \cdot p_A + Q \wedge (K - s)^+ p_h + \pi((K - s - Q)^+)] \\ &\quad + (1 - \Lambda_\tau^t) [Q \wedge K \cdot \omega(\tau) + v_\tau((K - Q)^+)]. \end{aligned}$$

We get $\omega(t)$ by solving out w from $\Pi_A^{accept} = \Pi_A^{reject}$. \square

Derivation of $d_t^A|_{t_s > t}$: If A accepts \mathcal{C} 's offer w at t , then $\Pi_A^{accept} = (K \wedge Q)w + \bar{V} [(K - Q)^+]$, wherein $\bar{V} [(K - Q)^+] = \Lambda_T^t [s \wedge (K - Q)^+ p_A + \pi((K - s - Q)^+)] + (1 - \Lambda_T^t) \pi((K - Q)^+)$. If A rejects, then the bargaining breaks down and $\Pi_A^{reject} = \Lambda_T^t [s \wedge K \cdot p_A + \pi((K - s)^+)] + (1 - \Lambda_T^t) \pi(K)$. We get $d_t^A|_{t_s > t}$ by solving out w from $\Pi_A^{accept} = \Pi_A^{reject}$. \square

Proof of Proposition 1: $\bar{p}_t|_{t_s > t}$ is linear in $d_t^A|_{t_s > t}$, and how $d_t^A|_{t_s > t}$ is related to t can be easily obtained by looking at (3) and noticing that Λ_T^t is decreasing in t . \square

Proof of Theorem 1: By (2) and $\tau(t) = \tau^* = \tau(\tau^*)$, we have $\omega(t) = (1 - \Lambda_{\tau^*}^t) \bar{p}_{\tau^*}|_{t_s > \tau^*} + \Lambda_{\tau^*}^t [p_A - p_A^\delta (1 - \theta(K))]$. Taking the first order derivative of $\omega(t)$ with respect to τ^* , we have

$$\frac{\partial \omega(t)}{\partial \tau^*} = -\frac{\lambda_{\tau^*}}{1 - \Lambda_t} \bar{p}_{\tau^*}|_{t_s > \tau^*} + (1 - \Lambda_{\tau^*}^t) \frac{\partial \bar{p}_{\tau^*}|_{t_s > \tau^*}}{\partial \tau^*} + \frac{\lambda_{\tau^*}}{1 - \Lambda_t} [p_A - p_A^\delta (1 - \theta(K))].$$

We know that $\bar{p}_t|_{t_s>t} = (1 - \gamma) \cdot p_A + \gamma \cdot d_t^A|_{t_s>t}$. Taking the first order derivative of $\bar{p}_t|_{t_s>t}$, we get

$$\begin{aligned}
\frac{d\bar{p}_t|_{t_s>t}}{dt} &= \gamma \frac{\lambda_t(\Lambda_T - 1)}{(1 - \Lambda_t)^2} [\theta(K)p_A + (1 - \theta(K)) d_T^A|_{t_s \leq T} - d_T^A|_{t_s > T}] \\
&= \frac{\lambda_t}{1 - \Lambda_t} [\gamma(\Lambda_T^t - 1) (\theta(K)p_A + (1 - \theta(K)) d_T^A|_{t_s \leq T}) + \gamma(1 - \Lambda_T^t) d_T^A|_{t_s > T}] \\
&= \frac{\lambda_t}{1 - \Lambda_t} [\gamma d_t^A|_{t_s > t} - \gamma (\theta(K)p_A + (1 - \theta(K)) d_T^A|_{t_s \leq T}) + (1 - \gamma)p_A - (1 - \gamma)p_A] \\
&= \frac{\lambda_t}{1 - \Lambda_t} [\bar{p}_t|_{t_s > t} - \gamma d_T^A|_{t_s \leq T} - \theta(K) (\gamma p_A - \gamma d_T^A|_{t_s \leq T}) - (1 - \gamma)p_A] \\
&= \frac{\lambda_t}{1 - \Lambda_t} [\bar{p}_t|_{t_s > t} - p_h - \theta(K) (p_A - (1 - \gamma)p_A - \gamma d_T^A|_{t_s \leq T})].
\end{aligned}$$

Therefore, $\frac{\partial \omega(t)}{\partial \tau^*} = \frac{\lambda_{\tau^*}}{1 - \Lambda_t} \theta(K) [(1 - \gamma)p_A + \gamma d_T^A|_{t_s \leq T} - p_h] = 0$. Then $\omega(t) = \lim_{\tau^* \searrow t} \omega(t) = \bar{p}_t|_{t_s > t}$. \square

Proof of Corollary 1: According to (2), we have $\omega'(t) = \frac{\lambda_t}{1 - \Lambda_t} [\omega(t) - (1 - \theta(K))p_h - p_A\theta(K)]$, and thus

$$\frac{d}{d\tau} J_\tau^t(W) = \frac{\lambda_\tau}{1 - \Lambda_t} (Q \wedge K) \theta(K) (p_A - R_s).$$

The results follow immediately. \square

Proof of Proposition 2: \mathcal{A} 's expected revenue can be written as

$$\Pi_A = C^I + C^{II} \cdot \Lambda(T(p_A)) + C^{III} \cdot p_A \cdot \Lambda(T(p_A)).$$

Then we take the first order derivative with respect to p_A and get

$$\begin{aligned}
\frac{\partial \Pi_A}{\partial p_A} &= -C^{II} \cdot \lambda(T(p_A)) \cdot \alpha T_o + \sum C^{III} \cdot \Lambda(T(p_A)) \\
&\quad - C^{III} \cdot p_A \cdot \lambda(T(p_A)) \cdot \alpha T_o \\
&= \lambda(T(p_A)) \cdot \left[\frac{\Lambda(T(p_A))}{\lambda(T(p_A))} \cdot C^{III} - C^{II} \cdot \alpha T_o - C^{III} \cdot p_A \cdot \alpha T_o \right] \\
&= \lambda(T(p_A)) \cdot G(p_A)
\end{aligned}$$

Suppose there exists p_A^* that solves $G(p_A) = 0$. Because Λ is log-concave, we know $\frac{\Lambda(T(p_A))}{\lambda(T(p_A))}$ is decreasing in p_A . As a result, $G(p_A) > 0$ when $p_A < p_A^*$ and $G(p_A) < 0$ when $p_A > p_A^*$. Therefore, p_A^* is unique and thus Π_A is quasi-concave given that $\lambda(T(p_A)) > 0$. The result follows. \square

Proof of Lemma 1: As stated in the text, if $K < \mathbb{T}\mathbb{D}$ and $p_A < p_B$, then $\frac{d}{dt}J_t^0 = 0$ for $\forall t \in [0, \tau)$, so by continuity of J_t^0 we have $J_t^0 = \text{constant}$ for $\forall t \in [0, \tau]$. If $K < \mathbb{T}\mathbb{D}$ and $p_A \geq p_B$, then $\frac{d}{dt}J_t^0 = 0$ for $\forall t \in (\tau, T]$. Again, by continuity of J_t^0 we have $J_t^0 = \text{constant}$ for $\forall t \in [\tau, T]$. The result follows. \square

Proof of Lemma 2: According to Lemma 1, customer i has no incentive to deviate given all other customers buy at $\tau = 0$. This is irrelevant to customer's demand size, so the argument applies to all customers. \square

Proof of Proposition 3: Notice that we can write $\pi(X) = \mathbb{T}\mathbb{D} \cdot \mathbb{C}(x)$, for which \mathbb{C} is an increasing concave function with constant parameters and $x = X/\mathbb{T}\mathbb{D}$. Given $\kappa \geq 1$, we have $w_{-i}^A = [\mathbb{C}(\kappa - f_{-i}) - \mathbb{C}(\kappa - 1)]/f_i$. Hence, $p_i = (1 - \gamma) \cdot p_A \wedge p_B + \gamma \cdot [\mathbb{C}(\kappa - f_{-i}) - \mathbb{C}(\kappa - 1)]/f_i$. Because $f_1 + f_2 = 1$, we have $\bar{p} = f_1 p_1 + f_2 p_2 = (1 - \gamma) \cdot p_A \wedge p_B + \gamma \cdot [\mathbb{C}(\kappa - f_1) + \mathbb{C}(\kappa - f_2)] - 2\gamma \cdot \mathbb{C}(\kappa - 1)$. Due to the concavity of \mathbb{C} , for $\Delta > 0$, we have $\mathbb{C}'(\kappa - 0.5 - \Delta) > \mathbb{C}'(\kappa - 0.5 + \Delta)$, so $\partial \bar{p} / \partial \Delta = \gamma \cdot [-\mathbb{C}'(\kappa - 0.5 - \Delta) + \mathbb{C}'(\kappa - 0.5 + \Delta)] < 0$. \square

2-Bargainer Equilibrium Prices for Medium and Low Capacity levels: According to (17) and (18), if $f_1 \leq \kappa < 1$, then we can solve for the equilibrium prices,

$$p_i^e = \frac{f_{-i} \cdot (1 - \gamma) \cdot p_A \wedge p_B + \frac{f_{-i}}{1 - \kappa} \cdot \mathbb{C}(\kappa - f_{-i}) + \gamma \cdot \mathbb{C}(\kappa - f_i)}{\frac{f_i \cdot f_{-i}}{\gamma \cdot (1 - \kappa)} - \gamma \cdot (1 - \kappa)}, \quad i = 1, 2;$$

if $0 < \kappa < f_1$, then we have

$$\begin{cases} p_1^e = \frac{(1 + \gamma \cdot (\kappa \wedge f_2) / \kappa) \cdot (1 - \gamma) \cdot p_A \wedge p_B + \gamma \cdot \mathbb{C}([\kappa - f_2]^+) / \kappa}{1 - \gamma^2 \cdot (\kappa \wedge f_2) / \kappa}, \text{ and} \\ p_2^e = \frac{(1 + \gamma) \cdot (1 - \gamma) \cdot p_A \wedge p_B + \gamma^2 \cdot \mathbb{C}([\kappa - f_2]^+) / \kappa}{1 - \gamma^2 \cdot (\kappa \wedge f_2) / \kappa}. \end{cases} \quad \square$$