Carnegie Mellon
Tepper
SCHOOL OF BUSINESS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
INDUSTRIAL ADMINISTRATION
(OPERATIONS MANAGEMENT AND MANUFACTURING)

Titled

"ESSAYS IN SERVICE OPERATIONS MANAGEMENT"

+

Presented by
Michele Dufalla

Accepted by

Alan Scheller-Wolf

Chair: Prof. Alan Scheller-Wolf

April 21, 2014

Date

Approved by The Dean

Robert M. Dammon

Dean Robert M. Dammon

4/22/14

Date

Essays in Service Operations Management

by

Michele Dufalla

Submitted to the Tepper School of Business
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
Operations Management & Manufacturing

at

Carnegie Mellon University

May 1, 2014

Dissertation Committee
Professor Sunder Kekre
Professor Alan Scheller-Wolf (Chair)
Associate Professor Nicola Secomandi
Associate Professor Rein Vesilo
Associate Professor Param Vir Singh

Abstract

In this dissertation, I discuss three problems within service operations management: identifying situational attributes that lead to positive customer outcomes under a Twitter-based customer service framework; the conditions for finite delay of first-in-first-out multiserver systems when confronted with integral loads; and the relative performance of different bargaining mechanisms for a seller of finite perishable inventory, with a further investigation of the consequences of modeling private information.

First, we consider a large telecommunications company that provides customer support over Twitter. Using 10 months of service data, we apply model selection techniques to develop an ordinal logistic regression model assessing the probability that a given customer service interaction will result in a positive, neutral or negative resolution as determined by the customer's sentiment expression. Our model incorporates customer, service and network explanatory attributes. We find that customers are less likely to experience a positive final sentiment as time passes, that is, those cases later in the 10 month period studied are less likely to experience positive resolution. This suggests that there is a drop-off in the likelihood of more positive resolution, but that this effect levels off. This finding may indicate a shift by the customer service team to harder to resolve cases as the program matures.

Next, we consider conditions for finite expected delay in FIFO multiserver queues with integral loads. Scheller-Wolf and Vesilo (2006) find necessary and sufficient conditions for a finite r$^{\text{th}}$ moment of expected delay in a FIFO multiserver queue, assuming a non-integral load and a service time distribution belonging to class $\mathcal{L}_1^{\beta}$. Removing the non-integral load assumption results in a gap between the identified necessary and sufficient conditions, as discussed by Foss (2009). We decrease the size of this gap through the application of domain of attraction results. Specifically, we find a stricter necessary condition for a GI/GI/K-server system with integral ρ that is more restrictive than those in the literature.

Finally, we consider the problem of a seller with a finite supply of perishable inventory. We consider four price setting mechanisms: seller posted price, buyer posted price, split-the-difference, and the neutral bargaining solution. We rank the value of these different mechanisms analytically and numerically in the context of the symmetric uniform trading problem from the perspective of the seller. While the ordering of the mechanisms remains the same as compared to the infinite horizon case studied in the literature, we use a model analogous to the infinite horizon case to find numerically that the relative value of the split-the-difference mechanism increases when the seller ultimately faces a deadline to complete the sales. The split-the-difference mechanism becomes more valuable as

the ratio of available inventory to time remaining increases because it is more likely to result in a sale than the seller posted price mechanism. In general, modeling private information is more challenging for the split-the-difference and neutral bargaining solution mechanisms than for the two posted price mechanisms. To assess the importance of this added complication, we quantify the effect of modeling private information when computing the seller's opportunity cost and find that while private information makes only a small difference in the neutral bargaining solution case, this modeling choice makes a large difference in the split-the-difference case when the seller is weak.

Acknowledgments

Firstly and foremostly, I thank my dissertation committee: a group characterized by superhuman patience, generosity of time and unvarying kindness. Specifically, thanks to Alan Scheller-Wolf, for chairing and for years of careful, clear and complete explanations; to Sunder Kekre, for always knowing what to try next and for his insights into practice; to Nicola Secomandi, for his extensive modeling knowledge and steadfast attention to detail; to Rein Vesilo, for demystifying large deviations and effective comments; and to Param Vir Singh, for his helpful suggestions and outside viewpoint.

For the second chapter, I thank Baohung Sun for her excellent choice modeling course and advice, Liye Ma for his advice, Swapneel Desai for data acquisition, Yingda Lu for his modeling ideas, and Bo Lin for his programming skills. Additionally, I am grateful to Maria Ferreyra, Peter Boatwright, Mark Schervish, Howard Seltman, and Patrick Foley for their generous help and perspectives in detangling a few thorny statistical issues. For the third chapter, I remain in awe of Mor Harchol-Balter's Performance Modeling course and book.

The formatting for this document was provided by a template by Steven Gunn and Sunil Patel from http://www.latextemplates.com.

For years of fun and help in many different ways, I thank my friends Yangfang Zhou, Tinglong Dai, Emre Nadar, Aabha Verma, Elvin Çoban, Masha Shunko, Paul Enders, John Turner, Alp Akçay, Xin Fang, Canan Güneş, Erkut Sönmez, Xin Wang, Ying Xu, Vince Slaugh and Gaoqing Zhang. A very special thanks to Lawrence Rapp, who ensured that every administrative aspect of my time at Tepper ran perfectly smoothly!

I'm forever indebted to my parents for their limitless patience and generosity, specifically to my father Michael Dufalla for his infinite support in all matters and to my mother Penny Suwak for her willingness to listen and unwillingness to sugarcoat the truth. Finally, my thanks to my sister Nicole Dufalla for her unflinching honesty and cartoonish hijinks, to my sister Jacqueline Dufalla for her fathomless calm and rapier wit, and to Mahdi Cheraghchi for his unwavering support and much-needed perspective.

Contents

List of Figures

List of Tables

For my father:
Nie poddawaj się!

Chapter 1

Introduction

Service operations management is a broad subfield incorporating many different management problems. In this dissertation, I focus on three: identifying situational attributes that lead to positive customer outcomes under a Twitter-based customer service framework, the conditions for finite delay of first-in-first-out multiserver systems when confronted with heavy-tailed service times, and the relative performance of different bargaining mechanisms for a seller of finite perishable inventory, with a further investigation of the consequences of modeling private information.

Social media sites, including Twitter, provide companies with a unique window into the minds of their customers, as well as a way to monitor (and influence) their reputation in real-time. While social media has obvious marketing and brand management applications, many companies are now branching out into providing customer service directly through the venues where customers complain. This has the dual advantages of providing customer service in a customer-chosen outlet while potentially allowing the company to protect its reputation. However, while elements of traditional customer service have been studied extensively, the applicability of these elements to new, social media based forms of customer service is unclear.

In chapter 2, we consider a large telecommunications company that provides customer support over Twitter. In order to identify which factors are most important for customer satisfaction when administering customer support over Twitter, we use model selection techniques on 10 months of service data to develop an ordinal logistic regression model assessing the probability that a given customer service interaction will result in a positive, neutral or negative resolution as determined by the customer's sentiment expression. Our model incorporates customer, service and network attributes.

We find that customers are less likely to experience a positive final sentiment as time passes, that is, those cases later in the 10 month period studied are logistically less likely to experience positive resolution. This suggests that there is a drop-off in the likelihood of more positive resolution, but that this effect levels off. This finding may indicate a shift by the customer service team to harder to resolve cases as the program matures.

The noise present in the data suggests that Twitter-based customer service analysis is not as straightforward as it might seem. While much data is available, difficulties in sentiment coding, situation classification heterogeneity and the determination of customer sentiment from Twitter comments contribute to noise. Advance cooperation in experimental design may alleviate some of these issues, and should be considered in future work.

Variability has long been the enemy of short wait times, so it is no surprise that heavy-tailed service times create special challenges for the analysis of queuing systems. Computer network traffic is often heavy-tailed (Willinger et al. [76] empirically demonstrates self-similarity in LANs), underscoring the importance of meaningful system analysis, but simulation of heavy-tailed distributions is difficult due to the extreme rarity of extremely large times (e.g. Nguyen and Robert [57]). Fortunately, in chapter 3 we are able to extend the analytic conditions for finite expected delay in these systems.

While Scheller-Wolf and Vesilo [67] find necessary and sufficient conditions for a finite r$^{\text{th}}$ moment of expected delay in a FIFO multiserver queue, assuming a non-integral load and a service time distribution belonging to class $\mathcal{L}_1^\beta$, removing the non-integral load assumption results in a gap between the identified necessary and sufficient conditions, as discussed by Foss [27]. Specifically, Scheller-Wolf and Vesilo [67] show that for a FIFO multiserver queue, the r$^{\text{th}}$ moment of expected delay $E[D^r]$ will be finite if $E[S^{1+(\frac{r}{(K-k)})}]$ is finite, where S represents the service time distribution, K is the number of servers in the system, $k = \lfloor\rho\rfloor \leq k{+}1 \leq K$, k integral and load $\rho := E[S]/E[T]$, with T representing the interarrival time distribution. This is also a necessary condition if $k < \rho < k+1$ or if $k+1 = K$, and $S \in \mathcal{L}_1^\beta$, $1 < \beta < \infty$, $\beta = (s - \lfloor\rho\rfloor + \alpha)/(s - \lfloor\rho\rfloor$, $\alpha \geq 1$ (meaning that $E[S] < \infty$ and if $S_1, ..., S_m$ are i.i.d random variables distributed as S, then $E(S^\beta) = \infty$ implies $E(min(S_1, ..., S_m)^{m\beta}) = \infty$.

Through the application of domain of attraction results, we decrease the size of this gap. Specifically, we find a stricter necessary condition for a GI/GI/K-server system with integral $\rho = R$: the r$^{\text{th}}$ moment of expected delay $E[D^r]$ will be infinite if $E[S^{1+(\frac{r}{\alpha(K-k)})}]$ is infinite, which occurs when the shape parameter of the service time distribution $\alpha < \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{r}{K-R}}$.

Negotiation is important for both business to business and business to customer sales. Consider a seller with a finite inventory of perishable goods being visited by a series of potential buyers. The seller could propose a take-it-or-leave-it price for a unit of inventory. Alternately, the buyer could propose a take-it-or-leave-it price. Both buyer and seller could state prices, then "split-the-difference". Finally, the buyer and seller could engage in a face-to-face negotiation, the result of which can be represented by Myerson's [53] neutral bargaining solution (NBS).

In chapter 4, we rank the value of these four different mechanisms analytically and numerically in the context of Chatterjee and Samuelson's [15] symmetric uniform trading problem, from the perspective of the seller. We demonstrate that a seller posted price always performs at least as well for the seller as the neutral bargaining solution, which always performs at least as well as a buyer posted price, and that splitting-the-difference always performs at least as well as a buyer posted price. While this ordering of the mechanisms remains the same as compared to the infinite horizon case studied in the literature, we find numerically, in an analogous model, that the relative value of the split-the-difference (STD) mechanism increases as we move to a situation where the seller faces a deadline to complete the sales. The higher the ratio of available inventory to time remaining becomes, the more valuable the STD mechanism becomes, because it is more likely to result in a sale. While STD lacks the simplicity of a buyer or seller posted price, it is a relatively easy to implement method for automated bargaining, making this a practical option for a seller.

Incorporating private information into the model adds an additional layer of complexity. We quantify the importance of modeling private information when computing the seller's opportunity cost under the STD and NBS mechanisms. While using a simplified model that calculates the seller's opportunity cost using public information may be an acceptable approximation for the NBS mechanism, it produces substantially different results than the private information case when STD mechanism is used by a strong seller.

We proceed first by identifying situational attributes that lead to positive customer outcomes under a Twitter-based customer service framework in Chapter 2. In Chapter 3, we find conditions for finite delay of first-in-first-out multiserver systems when confronted with heavy-tailed service times. In Chapter 4, we assess the relative performance of different bargaining mechanisms for a seller of finite perishable inventory, with a further investigation of the consequences of modeling private information. Finally, in Chapter 5, we summarize our findings and avenues for future work.

Chapter 2

Selecting a Model for Twitter-Based Customer Service Quality Metrics

2.1 Introduction

While elements of traditional customer complaint management have been studied extensively, the applicability of these elements to new, social media based forms of customer service is still being examined. Many companies monitor social networking sites, including Twitter, to gauge public opinion and to identify problems. Other companies, such as Dell, Whole Foods Market and Jet Blue Airways, go further and use social media based customer support teams to attempt to assist customers over these media [41]. Advice for such uses is given in the popular press (for examples please see [34] [31]), but have not been exhaustively studied. In this work, we hope to identify the driving factors behind successful Twitter-based customer complaint remediation.

The public nature of social media adds an interesting complication to the conventional service interaction. Within Twitter, a user has "friends" and "followers." When a user's friend writes a tweet, that user will be able to see the tweet on his homepage. While all public tweets are viewable by all Twitter users, a user will select friends to assemble a curated page of tweets that interest him. Similarly, a "follower" is a user that subscribes to the tweets of another user. Consider figure 2.1. John likes to read the tweets his friends, Samantha, Nancy and Lady GaGa. Nancy and Ted like to read John's tweets. Notice that Nancy is both John's friend and follower, while other users may be only a friend or only a follower of a user. Please note that despite the terminology used by Twitter, a "friend" is not necessarily a mutual designation or indicative of any other relationship between users.

This chapter is joint work with Sunder Kekre.

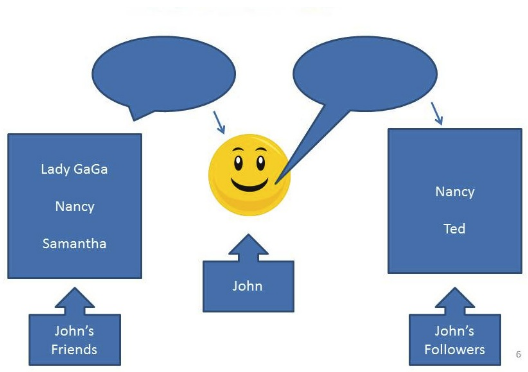


FIGURE 2.1: John's Twitter network. Messages posted by John's friends are collected for him to read. John's followers read messages posted by John.

This public interaction results in the need for the company to consider its brand image. Online complaint resolution becomes both necessary and extremely important. Kapland [41] and Griffin [34] both emphasize the need to respond quickly and appropriately to both positive and negative comments.

We consider one large telecommunications company. This company has a staff dedicated to monitoring tweets containing the names of the company or its products. Those tweets are evaluated and responded to when appropriate. Basic support advice can be provided this way. For more complicated problems, the support staff can refer a customer to a URL where they can access additional service assistance. To address the problem of which metrics are important under this new service regime, we analyzed Twitter customer service interactions from a period of 10 months. Using half the data set, a model incorporating customer, service and network attributes was built. The relevant variables are summarized in Figure 2.2. A test of this new model on the second half of the data set suggests that cases later in the data set are less likely to have a more positive resolution, but that this effect levels off. This may be the result of an expansion of the social media customer service group addressing more complicated cases as time progresses.

The difficulty in obtaining significant parameter values may stem from experiment design, but we propose several changes that could be made in future research to ameliorate this problem.

2.2 Literature

There is a large and varied literature on customer complaint management. One excellent general resource is the book "Complaint Management: The Heart of CRM." [70]

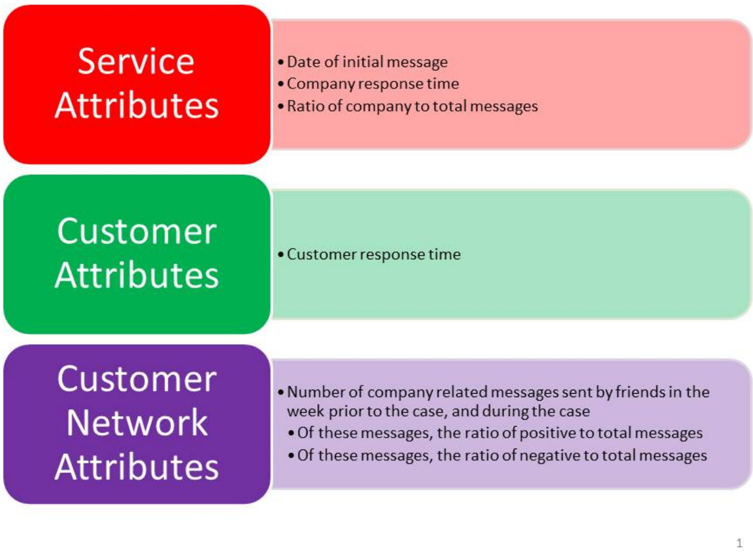


FIGURE 2.2: Service, customer and network attributes included in the model.

While this book does not discuss social media, it does detail most important elements of complaint resolution, including techniques, analysis and firm improvement.

Before progressing farther, it would be appropriate to consider the numerous benefits of complaint resolution, beyond merely solving one particular customer's problem. Adamson [2], and Fundin and Bergman [30] discuss the usage of complaint information for the improvement of company products and services. Complaints offer a window into customer perception, and insights derived from the valuable feedback can be used to make the firm more competitive. Berry and Parasuraman [7], Bosch and Enriquez [10], and Faed [23] each discuss possible systems for the appropriate collection, management and analysis of the necessary data for this task. One financial consideration is the value of customer retention. In a pair of papers, Fornell and Wernerfelt [25] [26] model the customer complaint process and consider the cost of lost customers due to unvoiced complaints. Also within the realm of defensive marketing, Ruyter and Brack [21] note the legal ramifications of appropriate complaint management for the reduction of liability.

Despite these benefits, not all firms make customer complaint management a priority. In one study, Gulas and Larsen [35] found that 29.2% of communications (a mixture of complaints and compliments) sent to a variety of companies were left unanswered. However, the same study found that this response rate was unrelated to the companies' returns on investment.

One large opportunity and possible pitfall related to complaint management is the role of customer word of mouth. Chelminski and Coulter [16] find that customers are more likely to complain to their friends rather than the company. To make matters worse, Casado et al. [11] find that customers who complain to the company and then receive inadequate service are more likely to both leave the company and complain to friends.

Given the consequences of ignoring or mismanaging customer complaints, many studies try to identify the characteristics of "good" customer complaint management. Blodgett and Anderson [9] use survey data from dissatisfied retail customers to explore a customer complaint Bayesian network model. This model incorporated store loyalty, store type, a customer's attitude towards complaining, a customer's belief about the controllability of the problem to determine how likely a customer would be to complain, and then how satisfied the customer would be in his or her problem resolution, and then considered the word of mouth and future loyalty effects of the result. Interestingly, after successful complaint resolution, newly happy customers exhibited a 46% probability of positive word of mouth. This result certainly suggests that for our paper, the company faces a great opportunity to convert Twitter complainers into Twitter promoters. A simpler loyalty focused model (without WOM effects) is provided by Andreassen [4]. Another loyalty focused model (also without WOM effects) now incorporating overall customer satisfaction is provided by McQuilken et al. [51]. Focusing on a deeper level, Davidow [20] provides a framework clarifying both the current state and proposed future directions of research focused on the actual mechanics of customer complaint response in the areas of timeliness, facilitation, redress, apology, credibility, attentiveness and interaction. Chan and Ngai [14] also focus on the mechanics of the company's complaint response, although from the perspective of justice and fairness theory. Finally, reaching into electronic customer service, Murphy et al. [52] examine company emails used for hotel customer service, categorizing each e-mail in dimensions of personalization, responsiveness, reliability and tangibility.

Extending into the online realm, several studies examine how companies should manage online customer complaints posted in forum-type environments (as opposed to microblogging arenas like Twitter - for an overview of Twitter network characteristics, see Java et al. [40]). Cho and Fjermestad [17] provide a brief overview of the scant literature surrounding online complaint management. Cho et al. [18] characterize the nature of complaints in a variety of online complaint forums. Harrison-Walker [36] performs a similar analysis and also considers the customer experience that led to the online posting of a complaint, with an excellent series of managerial recommendations. Lee and Lee [45] approach the online forum feedback response issue from a customer trust perspective. Lee and Song [46] consider the word of mouth implications of customer feedback and company response on online forums, investigating the risks of defensive, accomodative and no action strategies by the company. Bach and Kim [6] also look at company response and management of customer word of mouth in an online forum using a case study comparing one company's "proactive" approach with another's "defensive" approach. Branching from online forums into social media, Jansen et al. [39] consider WOM implications of Twitter.

This chapter attempts to span these areas by again examining the actual mechanics of customer complaint response (as in, which attributes lead to a positive complaint resolution experience), but within a Twitter-based WOM regime. Ma et al. [48] examine sentiment change for customers using the same data set as this work. However, while this chapter focuses on customer, service and network attributes that result in sentiment change during a firm intervention using ordinal logistic regression, Ma et al. [48] use a dynamic choice model to consider a customer's sentiment change over time, influenced by the sentiments in that customer's network, with firm intervention as an exogenous factor. The two approaches use different modeling strategies and different granularities to answer different questions.

2.3 Data

The complete data set includes 1,149,851 Twitter messages (tweets) collected by a large telecommunications company between February 13 and December 22, 2010. The tweets included in the set are all public messages, not private "direct messages", mentioning the company or its products by name. The company organized these messages into 12,625 cases. Some tweets were responded to by customer service representatives offering assistance. Not all tweets were assigned a case, and many cases contain multiple tweets.

2.3.1 Case Selection

Cases varied dramatically in their content. To ensure a basic level of comparability among important metrics, chosen cases fulfilled the following requirements:

1. The first message in the case was written by a customer (i.e. a profile name other than that of the company).

2. At least one message in the case was written by the company's customer support team.

3. The final message in the case was written by the customer.

4. The customer's list of friends and followers was publicly accessible at the time of inquiry.

5. The customer had fewer than 5000 friends and fewer than 5000 followers.

6. The initial message was posted after the first week of data collection (so that a full week of background network information is available).

7. The two sentiment coding tools discussed in Section 2.3.2 did not disagree on the initial and final sentiments expressed by the customer (i.e. one tool did not code the initial sentiment as "positive" while the other indicated that it was "negative")

8. The company responded to each customer message in the case within one week, on average.

The first two conditions make sure that the examined cases follow a pattern of customer complaint then customer service response. These conditions are necessary, as many included cases either do not include a customer service exchange, or include an exchange in a format impractical to analyze (such as an exchange initiated by customer service). The third condition is necessary to allow us to determine the ultimate outcome of the exchange. We are interested in the ultimate outcome of the interaction, so this final customer sentiment is needed. The fourth, fifth and sixth conditions allow us to examine the effects of comments within a customer's network of friends (see Section 2.3.3). The seventh condition allows us to exclude cases that were likely to have misclassified sentiments. The eighth condition is intended to identify and remove atypical cases where the company and customer do not seem to have an interactive experience.

706 cases fulfilled these requirements. The breakdown of these 706 cases by initial and final sentiments for the model selection and test data partitions can be found in Appendix A.1 in Tables A.1 and A.2. Descriptives for the variables ultimately included in the model can be found in Appendix A.2.

2.3.2 Sentiment Recoding

The company used an automated coding system to classify the content of messages as positive, negative or neutral. Not all messages were classified. For consistency, we initially used the Stanford Twitter Sentiment Classification API bulk classification service (previously available at http://twittersentiment.appspot.com/api/bulkClassify, now available at http://help.sentiment140.com/api) to categorize messages in the data set (including messages not included in any case, for use in determining network effects in Section 2.3.3) into positive, negative and neutral sentiment. Most messages are assigned a "neutral" coding. It appears that only very positive and very negative messages are coded as such. Manual inspection reveals that the categorization is far from perfect, however the size of the data set renders other methods of categorization impractical. For a discussion of how this tool works and its advantages, please see Go et al. [33].

In an attempt to improve the coding of the messages, we also used the free version of SentiStrength2.2 (available at http://sentistrength.wlv.ac.uk/) to code the messages. SentiStrength assigns each tweet a negativity score from -5 to -1 and a positivity score of +1 to +5. Combining these scores gives a net score of -4 to +4 indicating the polarity and strength of a tweet's sentiment. For the purpose of comparison with the Stanford Twitter Sentiment Classification tool, we considered tweets with scores between -4 and -2 to be negative, tweets with scores between -1 and +1 to be neutral, and tweets with scores between +2 and +4 to be positive. For a discussion of how this tool works and its advantages, please see Thelwall et al. [73] [72].

A tweet scored as "positive" by both tools was coded as "positive." A tweet scored as "positive" by one tool and "neutral" by the other was coded as "positive." A tweet scored as "neutral" by both tools was coded as "neutral." A tweet scored as "negative" by both tools was coded as "negative." A tweet scored as "negative" by one tool and "neutral" by the other was coded as "negative." Finally, a tweet was coded as inconclusive in the case that one tool indicated "negative" sentiment while the other tool indicated "positive" sentiment. As a consequence, those cases where either the initial or final sentiment was deemed inconclusive were removed.

To benchmark the performance of this sentiment coding scheme, we compared the sentiment assigned by the above algorithm with human-coding and found a 58% match. This match rate is similar to those found in other contexts (see the baseline (non-machine-learning) results in Pang et al. [60]).

2.3.3 Network Effects - User Friends

Twitter users have followers and friends. Consider Twitter user JohnSmith. JohnSmith's "followers" are the Twitter users who "subscribe" to his tweets. This means that when JohnSmith posts a tweet, his followers will see his tweet on a page containing the tweets of their "friends". Similarly, JohnSmith's "friends" are those users to whose tweets JohnSmith "subscribes." When making customer service decisions, the number of followers a user has can be used as a metric to determine influence. If a company is concerned about complaints propagating through a network, they may wish to give special consideration to those users who have many followers. The number of followers a user has was recorded in the original data set for this reason. Not recorded in the original data set is the number of friends that a user has.

In order to learn about complaint propagation through the network, we compiled a list of friends for each user studied. This information is available for many, though not all, Twitter users. We excluded from the study those users whose list of Twitter friends we were unable to access. In addition to those with private information, this includes those users whose profile names include spaces, which interfered with the script we used to retrieve the friend lists. Additionally, the API for friend retrieval limits returns to 5000 results so we removed cases with users with exceptionally high friend or follower counts.

After retrieving a list of user friends, we were able to find instances of users being exposed to the company related comments of others. To incorporate this effect into our model, we added a variable counting the number of positive tweets about the company a user would have seen in the week before his initial complaint and during the period of the case. Similar variables were created for neutral and negative tweets. An additional variable was created to indicated the sum of positive, neutral, negative and inconclusive messages seen during this time period. Lu et al. [47] find that recent posting activity is an important component in opinion leadership in the context of online forums, so it is not unreasonable to believe that those tweets sent most recently would be more influential in the current context. In addition to these magnitude measures, variables were created for the proportion of positive and negative messages in a customers network. Tweets written by the customer support team were ignored, as it seems likely a customer only added the company's customer support team as a friend once the customer service interaction was underway. Promotional tweets by the company were also ignored for the same reason.

Huberman et al. [37] show that while users have many friends, they interact with relatively few. (Also of interest is a study by Cha et al. [13] showing that a high follower count may not be an indicator of influence.) To account for different levels of Twitter attention, we established a new set of variables. First, we counted the number of positive comments about the company made by those who were both friends and followers of a user in the week before the customer's initial message and during the period of the case. For the previous example in Figure 2.1, we would only consider Nancy's comments when assessing John's exposure. Similar variables were created for neutral and negative messages. In addition, we counted the number of positive comments about the company made by those who were friends but not followers for a user in the previous week. Similar variables were created for neutral and negative messages. In addition to these magnitude measures, variables were created for the proportion of positive and negative messages in a customers network.

2.4 Model

Our model examines the importance of several operations metrics in the context of tweet based customer support. We attempt to identify the driving factors behind customer sentiment change. To do this, we consider the following basic interaction:

1. A customer is exposed to the messages sent by his or her Twitter friends. The comments of the customer's Twitter friends may affect the customer's future sentiment expression.

2. A customer tweets comments about the company's services or products. This tweet may be positive, neutral or negative in sentiment.

3. A customer service representative replies with assistance. The quality of this intervention may affect the customer's final sentiment.

4. Additional messages may be exchanged as the issue is resolved. The customer may still be influenced by comments within his or her network.

5. The customer tweets a last time, revealing a final sentiment of positive, neutral or negative.

We hope to determine the probabilities of different final sentiment states, given the service, customer and customer network attributes described in Section 2.4.2. To determine these probabilities, we use an ordinal logistic regression.

In SPSS, we used the ordinal logistic regression function to fit the following model:

$$\begin{aligned}
\pi_{i,positive} &= (\frac{1}{1 + exp(\alpha_{positive} - (x_i'\beta))}) \\
\pi_{i,neutral} &= (\frac{1}{1 + exp((\alpha_{neutral} - (x_i'\beta))}) - \pi_{i,positive} \\
\pi_{i,negative} &= 1 - \pi_{i,positive} - \pi_{i,neutral}
\end{aligned} \tag{2.1}$$

where $\pi_{i,j}$ is the probability of a final sentiment j (either positive, negative or neutral) for a customer i, with attributes x_i and parameters α_j and β. SPSS maximizes the log-likelihood function (plus a constant):

$$l = \sum_{i=1}^{m} \sum_{j=1}^{J-1} \left(\sum_{k=1}^{j} nklog \left(\frac{\gamma_{i,j}}{\gamma_{i,j+1} - \gamma i, j} \right) \right) - \left(\sum_{k=1}^{j+1} nklog \left(\frac{\gamma_{i,j+1}}{\gamma_{i,j+1} - \gamma i, j} \right) \right) \tag{2.2}$$

where $\gamma_{i,j}$ is "the cumulative response probability up to and including Y=j at subpopulation i", n is "the sum of all frequency weights" and m is "the number of subpopulations". [1]

For this model to be valid, the “proportional odds” assumption must hold. We used the parallel line test to confirm the validity of this assumption in our data. (UCLA Institute for Digital Research and Education [58]) As a clarification, we considered “initial sentiment” as a variable, but it was insignificant and left out of the final model. We are not modeling transitions, just the probability of a customer ending the interaction in a positive state.

To summarize, we are trying to discover which attributes change the probability that a given customer service interaction will result in a positive (or neutral or negative) final message sent by the customer. The ordinal logistic regression model allows us to calculate the probability of a positive (or negative or neutral) final sentiment, given a list of attributes of the interaction (for example, number of messages or elapsed time).

2.4.1 Model Selection

The key concern of this model is that it is unclear which of the variables (defined in Section A.3) should be included in this model. The quantity of potential variables poses a problem in several ways:

1. More variables included in a model results in lower significance for each variable, all else being equal.

2. Indiscriminately adding variables to a model greatly increases the risk of “false positive" significance results.

3. Relationships between the variables in the list could result in an inappropriate model if variables are added indiscriminately. For example, if the number of company messages, the total number of messages and the ratio of company to total messages are all added to the model, the effect of each becomes unclear. In this case, changing the number of company messages while holding the total number of messages constant necessarily changes the ratio of company to total messages. However, if all three variables are included in the model, an assumption would be made that changing the number of company messages while holding the total number of messages constant would not change the ratio of company to total messages, which is clearly untrue. For these variables, then, we could only include at most two of the three in the model.

4. We do not know if each variable has a linear relationship to the final sentiment. To explore this, we consider linear, quadratic, natural log and square root transformations of each variable. Obviously, we would not want to include both the natural log and square root transformations of a variable.

2.4.1.1 Model Selection Procedure

In order to address the above issues, we used the following procedure to move from the extended list of variables in Section A.3 to an appropriate model for analysis.

1. Randomly divided the data set into two parts. Part 1 is used for model selection (334 cases). Part 2 is saved for model testing (372 cases).

2. Using only part 1 of the data, each variable in the complete variable list is regressed alone against the dependent variable. Additionally, quadratic, square root and natural log transformations of these variables were also considered, where appropriate. (In the case where the variable may have a value of 0 (number of friends, number of followers, number of a certain type of network message, etc.), the natural log of the value of the variable plus one was taken.) Again, each model was run separately. Consequently, with the exception of the dummy variables, each variable had four possible models (linear, quadratic, natural log and square root). The linear model was assumed to be the best representation, unless one of the alternate models had an Akaike Information Criterion (AIC) value that was a least 2 less than the linear model [3]. In that case, the model with the lowest AIC was viewed to be the best representation of that variable.

3. Still using only part 1 of the data, the best representation of each variable in the list of complete variables (Appendix A.3) was considered for inclusion into a new model. Vincent Calcagno's 'glmulti' package for R was used to search through candidate models using a genetic algorithm, with the minimization of AICc as the goal [38]. Both the CLM and MASS packages were used. Due to redundant variables, some manual perturbation was used to ensure that the model was consistent with theory and not suffering from multicollinearity. These results are shown in Section 2.4.1.2.

4. A new regression was performed using the chosen model and the untouched half of the data (part 2) to obtain true significance values for the parameters and to validate the model selection. These results are shown in Section 2.5.

5. For comparison, we also ran a regression using the chosen model and the full data set (both the exploration half and the untouched test half). These results are shown in Appendix A.4.

2.4.1.2 Model Selection Procedure Results

Using R's glmulti and manual adjustment, the lowest AICc obtained was for a model including the variables for the ratio of company to customer messages, the average customer response time, the ratio of positive to total messages in the customer's network,

Table 2.1: Ordinal Logistic Regression Results for Exploratory Data

	Parameter Value	Std. Error	Sig.
$\alpha_{neutral}$	-1.721	.653	.008
$\alpha_{positive}$	.645	.646	.318
β Ratio of company to total messages	2.434	1.083	.025
β Company response time	.010	.076	.892
β Customer response time	-.115	.079	.144
β Number of company related messages in network	-.011	.010	.267
β Ratio of positive to total network messages	5.006	1.521	.001
β Ratio of positive to total network messages squared	-5.100	1.650	.002
β Ratio of negative to total network messages	-.614	.567	.278
β LN(Date of initial message)	-.245	.115	.034

the ratio of positive to total messages in the customer's network squared and the date of the customer's initial message. These variables are described in Section 2.4.2. This model had an AICc of 680.87. All variables had significant parameters (95%) in this case, except for the average customer response time. For comparison, some exploration was performed using glmulti for Bayesian Information Criterion (BIC) minimization [69] (via [12]). It is important to note that the best BIC value obtained occurred using only an intercept.

To make the model easier to interpret, we added the average company response time, the total number of messages sent within the customer's network during the period of interest and the ratio of negative to total messages in the customer's network. This increased the AICc to 685.10, indicating a significantly worse model. However, the same variables were significant as in step 3. The results of this model can be seen in Table 2.1. This model was computed in SPSS. Please note that the test of parallel lines accepted the null hypothesis that slope coefficients are the same across response categories (significance = .341), so the proportional odds assumption does not appear to be violated. Additionally, overall model significance is 0.004.

2.4.2 Variables

The variables included in the chosen model are described here. The full list of variables considered can be found in appendix A.3.

2.4.2.1 Service/Customer Attribute - Ratio of company to total messages

This variable divides the number of company messages by the total number of messages in a case (that is, the number of company messages plus the number of customer messages in a case).

2.4.2.2 Service Attribute - Company response time

This variable is the average time between a given customer message and the company's response in a case.

2.4.2.3 Service Attribute - Date

This variable indicates the date of the customer's first message.

2.4.2.4 Customer Attribute - Customer response time

This variable is the average time between a company message and the customer's response in a case.

2.4.2.5 Network Attribute - Number of company related messages in network

This variable is the total number of positive, negative, neutral and indeterminate messages involving the company or its products sent during the case, as well as in the week prior to the customer's first message, by the customer's friends.

2.4.2.6 Network Attribute - Ratio of positive to total network messages

This variable is the number of positive messages involving the company or its products sent during the case, as well as in the week prior to the customer's first message, by the customer's friends, divided by the number of company related messages in network variable. In the event that the number of company related messages in network variable had a value of 0, this variable was also coded as 0.

2.4.2.7 Network Attribute - Ratio of negative to total network messages

This variable is the number of negative messages involving the company or its products sent during the case, as well as in the week prior to the customer's first message, by the customer's friends, divided by the number of company related messages in network variable. In the event that the number of company related messages in network variable had a value of 0, this variable was also coded as 0.

2.4.2.8 Service Attribute - Date of initial message

This variable indicates the date of the customer's first message, where "1" indicates the first day in the data set (February 22). This value is subsequently incremented (e.g. Februrary 25 is "4").

2.4.3 Hypotheses

Based on our exploration of part 1 of the data set, we developed some hypotheses to test on part 2 of the data set.

2.4.3.1 Ratio of company to total messages

Hypothesis 1. A higher ratio of company to total case messages increases the probability of a more positive case resolution (holding the date, company response time, customer response time, number of network messages and the percentage of positive and negative network messages constant).

A higher company to total case message ratio indicates a higher service level. For each piece of information the customer sends to the company, he or she receives more information back. A lower company to total case message ratio would indicate that the company was providing less information to the customer. A higher service level would result in higher customer satisfaction. As an aside, because the first and last messages are always customer messages due to our case selection criteria, an otherwise 1:1 exchange (that is, the customer is always responded to by one company message, which is responded to by one customer message, etc.) would see a higher ratio of company to total messages as the total number of messages in the case increased. This could be a marker of case difficulty or complexity.

2.4.3.2 Ratio of positive to total network messages

Hypothesis 2. A higher ratio of positive to total network messages in the period between one week prior to the customer's first message and the customer's last message initially increases the probability of a more positive case resolution, but ultimately a quadratic relationship results in a penalty for a high positive ratio (holding the date, ratio of company to total messages, company response time, customer response time, number of network messages and the percentage of negative network messages constant).

Ma et al [48] found that positive sentiment expression in a customer's network led to more positive sentiment expression if the customer was already in a positive state and more negative sentiment expression if the customer was already in a negative state. Our initial analysis suggests that the effect during a customer service intervention is quadratic. During a customer service event, the customer's state is in flux. Initially, the customer had some complaint, but now resolution is possible. As the customer's network becomes more positive, the customer may absorb some of this enthusiasm. However, if the network becomes too positive, the customer's expectations may increase, resulting in a lower final sentiment when these expectations are not met.

2.4.3.3 Date of initial message

Hypothesis 3. A later chronological date decreases the probability of a more positive case resolution, although the effect stabilizes (holding the ratio of company to total case messages, company response time, customer response time, number of network messages and the percentage of positive and negative network messages constant).

Our initial analysis suggests the counter intuitive result that cases handled later in the data set are less likely to have a positive resolution. Please see Section 2.5.1 for details.

2.5 Results and Analysis

Using the untouched half of the data, the model selected in Section 2.4.1.2 produced the results shown in Table 2.2. Overall model significance is 0.066, so the model is significant at 90%. The test of parallel lines resulted in a significance of .170, indicating the the proportional odds assumption is not rejected, so ordinal regression remains appropriate. Unfortunately (but foreshadowed by the poor BIC results in Section 2.4.1.2), the parameters for the ratio of company to total messages and the ratio of positive to total network messages (and its quadratic term) are not significant in this independent sample. As

Table 2.2: Ordinal Logistic Regression Results for Test Data

	Parameter Value	Std. Error	Sig.
$\alpha_{neutral}$	-2.522	.570	.000
$\alpha_{positive}$	-.322	.555	.561
β Ratio of company to total messages	1.270	1.003	.205
β Company response time	.067	.069	.322
β Customer response time	-.050	.089	.576
β Number of company related messages in network	.009	.010	.378
β Ratio of positive to total network messages	-.526	1.338	.695
β Ratio of positive to total network messages squared	.755	1.570	.631
β Ratio of negative to total network messages	-.080	.508	.875
β LN(Date of initial message)	-.355	.104	.001

such, we are unable to address Hypotheses 1 and 2. The parameter for the natural log of the date of the initial message is significant, however, so Hypothesis 3 can be addressed.

2.5.1 Evaluation of Hypothesis 3

We find that the natural log of the date of the initial message has a parameter value of -0.355 with significance of 0.001 when the model was applied to the test data. This result confirms our hypothesis that a later chronological date decreases the probability of a more positive case resolution, although the effect stabilizes (holding the ratio of company to total case messages, company response time, customer response time, number of network messages and the percentage of positive and negative network messages constant). This effect is demonstrated in Figure 2.3, using the mean values recorded in Table A.3.

One may expect the probability of positive resolution to improve over time, as the company refines its customer service strategy. We propose three possible explanations for this counter intuitive result.

1. Brand perception may have changed over time and made customers harder to please, either through a general increase in negative sentiments, or through a general increase in positive sentiments resulting in harder to satisfy higher expectations. Although we control for the influence of positive and negative sentiments in a customer's network in the period immediately preceding (and during) a case, perhaps a long term effect is in play. To explore this possibility, the sentiments of all of the company related messages sent by the friends of the customers in the data set were analyzed. Figure 2.4 shows the percentage of the messages sent by these users classified as positive, negative and neutral during each week studied. (Percentages do not add to 1 because of a small number of messages coded as "inconclusive".) A linear regression suggests that the trends in positive

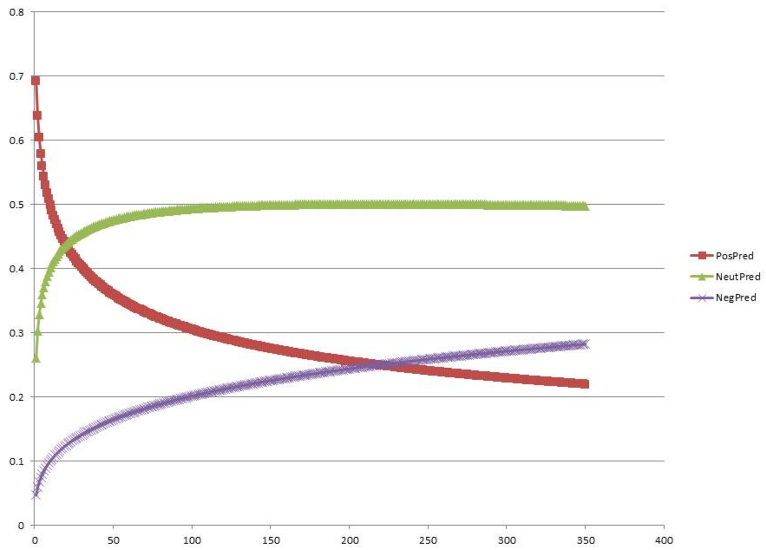


FIGURE 2.3: Probability of positive, negative or neutral final resolution for each day of the study period (February 22 is day 1).

and negative message percentages are not significant in this time period. These results seem to indicate that a general company related sentiment trend is not a driver of this result.

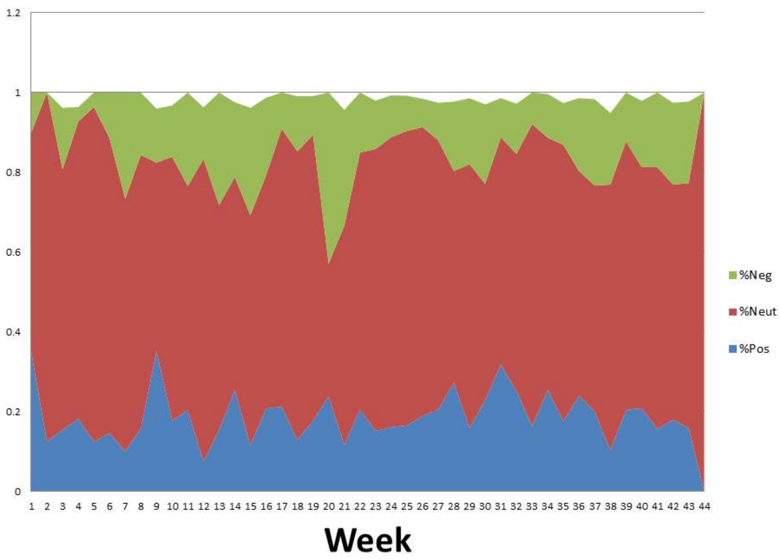


FIGURE 2.4: Percentage of messages sent by friends of the customers in the data set classified as positive, negative and neutral during each week studied.

2. The customer service group may have initially been staffed by higher quality agents, who were then replaced by a new set of lower quality agents. However, if this is the case,

we would expect to see an a reversal of this trend eventually as the new agents improved, while our model does not suggest improvement over a ten month period.

3. The customer service group may have changed its case selection methodology. Not all tweets are addressed, and the service representatives make the decision of which customers to contact. As time goes on, the customer service group becomes more confident and established. They attempt to solve harder cases, resulting in a drop off in service outcomes, but this drop off stabilizes after the transition is complete. We do not have any metrics controlling for case complexity or the nature of the problem being addressed. As such, we cannot test this hypothesis. However, this explanation would account for the decrease in the probability of more positive resolutions, as well as the leveling off period that follows.

2.5.2 Service time variables

It is of interest to note that none of the service time variables (average company response time, the company's response time to a customer's first message and the total elapsed time of a case) were selected by AIC (or BIC) to be included in the model, despite the well-known importance of wait time and customer satisfaction (please see Durrande-Moreau's [22] survey of empirical research in this area, as well as Taylor's [71] exploration of these effects). Average company response time was added manually for completeness, but was insignificant in both the model selection and hold out data regressions. Two possible explanations are immediately apparent:

1. Signal-to-noise ratio may be too high to detect these effects in our data set. Mattila and Mount [50] conducted an examination of the effect of company response times on e-mail-based complaint resolution. Unsurprisingly, they found that long company response times resulted in decreased customer satisfaction. It would be plausible to see a similar effect in social media based customer complaint situations as well.

2. Traditional service time metrics may not be important for Twitter-based customer service. Maister [49] notes that customers find waits to be shorter when they are occupied. Customers seeking resolution over Twitter do not have to wait in the same way as customers waiting for service over the phone or in person, so perhaps they are not concerned about wait time. However, a similar argument could be made for e-mail, where response times are significant. A possible difference could lie in the fact that customers who seek assistance over e-mail are actively awaiting a response, while those who are served using Twitter may or may not have initially been expecting a response. However, this would only affect the initial response time sensitivity, and not the average company response time sensitivity as once the service interaction has begun and customer would

naturally be expecting a response. Additionally, controlling for those customers who initially sought out assistance (by "mentioning" (using the "@" symbol) the company or its customer service profile) does not seem to result in significant service time metric parameters in our data set. It seems necessary to mention that Maister also indicates that wait times of uncertain duration seem longer than those of known duration, which would seem to indicate that wait time could be important in Twitter-based customer service, as there is no way for a customer to know what his or her wait might be. For completeness, it is important to note that even for traditional customer service Davidow [20] suggests that wait time may not be that significant, as long as it is "reasonable" according to the situation.

2.6 Conclusion

2.6.1 Limitations

Several limitations should be acknowledged. First, the coding of the data was imperfect. That is, even under the best of conditions categorizing messages as "positive," "negative," and "neutral" lacks nuance. In this case, the automatic coding was not perfect, often conflicting with manual coding, even under these rough categories. Combining the results of two different coding methodologies provided an important control, but the classification was still problematic. Unfortunately, the scale of data required precludes manual coding. However, the science of sentiment analysis continues to improve and this may not be as large an issue in the future.

Second, the data was extremely heterogeneous. Different products and problems were being discussed. For instance, one customer might have a small question about the operation of her phone, while another customer cannot access the internet at all. Some problems were trivial and some were serious. Additionally, the company's response type was also varied. In some cases, customers were simply referred to a URL for further assistance. Other cases involved detailed back and forth discussion between the customer and the customer support team. Again, the scale of the data causes difficulty for the categorization and control of these issues. Further research could prove illuminating here.

Third, in order to have a "final sentiment," only those cases where the customer sent a final message were considered. This decision adds a bias toward gregarious customers, and it is not hard to imagine that gregariousness would be influenced by the actual final sentiment state. Further study on this matter is required.

Fourth, in order to consider network effects, only those cases where the customer had complete network information available were considered. Hence, a few extreme customers with more than 5000 friends or followers were excluded from the analysis, as well as those with screen names containing spaces. More worrisome was the exclusion of customers with private follower lists.

Fifth, friend and follower information was time-delayed. While the company collected the number of followers at the time of service, the lists of customer friends and followers were obtained well after the end of service. Because of the dynamic nature of a customer's network, the list of friends and followers that were used to determine a customer's influence at time of service may not actually be the same friends and followers the customer had at the time of service.

Sixth, the analysis considered only a customer's first and last sentiment but not the intermediary states. Further analysis on these transitory states would be useful both for understanding the evolution of a customer's sentiment change as well as seeing the sentiment effects of different firm, customer and company attributes at a higher level of granularity.

Finally, customer satisfaction was only measured indirectly by Twitter sentiment analysis, instead of directly by survey. We have no information about the customer (due to the anonymity of Twitter), but many demographic and experience related questions could be asked in a survey. More troubling, we can only guess at a customer's customer service experience quality. For example, a polite customer's reaction may mask a substandard experience when scanning tweets, while a direct discussion with the customer could reveal his or her dissatisfaction. The opposite could also easily occur. Actual customer perception is completely ignored by our model.

The difference in granularity between this paper and that of Ma et al. may explain the difference in the quality of results between the two papers. Ma et al. [48] follows customers from message to message, only considering whether or not an intervention happens, while this paper considers an entire customer service interaction as one case. This aggregation may be responsible for the loss of explanatory power experienced.

2.6.2 Implications for Future Research

The limitations of post-hoc Twitter-based customer service analysis are severe. While sentiment coding and case classification may improve with technology, there is no fix for network data collection delays and the unavailability of customer perception information beyond better experimental design require the advance cooperation of a host company.

By planning data collection with the company before the service interactions take place, several limitations could be removed. Additionally, the framework provided by Froehle and Roth [29] could be an excellent starting point for developing a system to determine customer perceptions.

1. While sentiment coding of a customer's network may still be required, the actual customer sentiments could be determined by actually asking the customer, rather than inference.

2. Through consultation with the company, a list of common problem types could be developed. Going forward, the CSR who manages a case could code that case as a certain situation, avoiding difficult post-hoc decisions and classification difficulties.

3. Through cooperation with the company, customer friend and follower information could be collected at the time of service, rather than months later.

The problems ingrained in post-hoc Twitter customer service analysis are difficult to solve, but easy to bypass through advance planning and experimental design in conjunction with a company. Unfortunately, the information availability problems of after the fact analysis seem to proclude simple model building.

2.7 Caveat

Initial analysis was performed on the entire data set and a multinomial logistic model was developed, incorporating the ratio of company to total messages, the natural log of the number of messages sent by the company during the case, the natural log of the average company response time, the natural log of the average customer response time, the natural log of the number of followers a customer had, the natural log of the number of friends a customer had, the natural logs of the number of positive, neutral and negative messages written by a customer's friends who were also the customer's followers in the week before the initial message (but not during the period of case), the natural logs of the number of positive, neutral and negative messages written by a customer's friends who were not also the customer's followers in the week before the initial message (but not during the period of case), and the initial sentiment of the customer. All sentiments were coded using only one coding methodology (Sentiment140).

Chapter 3

Necessary Condition for Finite Delay Moments for FIFO GI/GI/K Queues with Integral Load

3.1 Introduction

Sufficient conditions for delay moments in stable GI/GI/K FIFO queuing systems, which are also necessary conditions for GI/GI/1 queues, were first established by Kiefer and Wolfowitz [42] [43]. Scheller-Wolf and Sigman [66], and later Scheller-Wolf [64], establish that these conditions are not necessary for multi-server GI/GI/K queues. By adding the condition that service times belong to class $\mathcal{L}^\beta$, as well as a requirement that traffic intensity is non-integer, Scheller-Wolf [65] was able to find necessary and sufficient conditions for GI/GI/K queues. Using methods developed by Whitt [75] and Foss and Korshunov [28], Scheller-Wolf and Vesilo [67] extended this result to service times in class $\mathcal{L}_1^\beta$, and then to the workload at different servers within a GI/GI/K system (as opposed to only delay) [68].

As mentioned above, Scheller-Wolf and Vesilo [67] show that for a FIFO multiserver queue, the r$^{\text{th}}$ moment of expected delay $E[D^r]$ will be finite if $E[S^{1+(\frac{r}{(K-k)})}]$ is finite, where S represents the service time distribution, K is the number of servers in the system, $k = \lfloor \rho \rfloor \leq k+1 \leq K$, k integral and load $\rho := E[S]/E[T]$, with T representing the interarrival time distribution. This is also a necessary condition if $k < \rho < k+1$ or if $k+1 = K$, and $S \in \mathcal{L}_1^\beta$, $1 < \beta < \infty$, $\beta = (s - \lfloor \rho \rfloor + \alpha)/(s - \lfloor \rho \rfloor)$, $\alpha \geq 1$. $S \in \mathcal{L}_1^\beta$ means that $E[S] < \infty$ and if $S_1, ..., S_m$ are i.i.d random variables distributed

This chapter is joint work with Alan Scheller-Wolf and Rein Vesilo.

as S, then $E(S^\beta) = \infty$ implies $E(\min(S_1, ..., S_m)^{m\beta}) = \infty$. Note that $\mathcal{L}_1^\beta$ includes the Pareto distribution. Consequently, there is a gap between the identified necessary and sufficient conditions for integral ρ, as noted by Foss [27].

Using domain of attraction results from Feller [24], as well as results from Scheller-Wolf [65], Scheller-Wolf and Vesilo [67], Foss and Korshunov [28] and the additional assumption that the service time distribution S has $P(X > u) \sim u^{-\alpha}$, such as with a Pareto distribution, we find a stricter necessary condition for a GI/GI/K-server system with integral $\rho = R$: the r$^{\text{th}}$ moment of expected delay $E[D^r]$ will be infinite if $E[S^{1+(\frac{r}{\alpha(K-R)})}]$ is infinite, which occurs when the shape parameter of the service time distribution $\alpha < \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{r}{K-R}}$. Please note that we are using R to denote integer values of ρ, to distinguish from the notation of k used with non-integer loads. To prove this, we first consider the case of $\rho = 1$ (Section 3.2) then generalize to higher integral ρ values (Section 3.3).

Theorem 3.1. *For a FIFO GI/GI/K queue with integral $\rho \leq K$, where $K \in N$ and $S \in \mathcal{L}_1^{\alpha+1}$ and S has $P(X > u) \sim u^{-\alpha}$, with $1 \leq \alpha \leq 2$ (such as with a Pareto distribution):*

$$E[S^{1+\frac{r}{\alpha(K-R)}}] = \infty \Rightarrow E[D^r] = \infty$$

3.2 GI/GI/K, $\rho = 1$

We begin by characterizing a FIFO GI/GI/K queue with $\rho = 1$. The work at server i at time n in a k-server system is denoted $W_{n,i}^{\{k\}}$. By definition, servers will not have negative work, accordingly $W_{n,1}^{\{k\}}, W_{n,2}^{\{k\}}, ..., W_{n,K}^{\{k\}} \geq 0$. Servers are periodically reordered so that $W_{n,i} \leq W_{n,i+1} \forall i = 1...K-1$. Consider a netput process Y, defined to be the amount of work accumulated by the system by job i. That is, $Y_i = (S_i - T_i)^+$.

We will consider lower bounds of the workload at each server in this system as time progresses. We will show that, with some probability, the delay of the system will be proportional to $x^{\frac{1}{\alpha}}$ in a deterministic arrival system (without loss of generality, as per Lemma 3.3). We can bound this probability by comparing with a system with $K-1$ servers and $\rho = \frac{1}{2}$. Finally, we find conditions on S that ensure an infinite expected delay.

3.2.1 Bounding Delay

3.2.1.1 $t = -M$

At time $-M$, we assume the system is empty. Time period M is sufficiently large that there is a probability $P_1(x)$ that the second server will accumulate more than $\frac{C}{2}x$ work during this time period.

3.2.1.2 $t = 0$

We assume, for a given x, that by time 0, server 2 contains more than $\frac{C}{2}x$ work, where C is chosen to be large enough to insure that $C > 2 + 2\epsilon E[S]^{-\frac{1}{\alpha}}$ for $\epsilon > 0$. This will occur with some probability $P_1(x)$. Consequently, servers 2 through K must each contain more than $\frac{C}{2}x$ work. The lower bounds on the work at each server are thus as follows.

$W_{0,1}^{\{K\}} \geq 0$

$W_{0,2}^{\{K\}} > \frac{C}{2}x$

...

$W_{0,K}^{\{K\}} > \frac{C}{2}x$

Now, we can find bounds for $P_1(x)$, the probability that server 2 contains more than $\frac{C}{2}x$ work using a result from Scheller-Wolf [65] .

Lemma 3.2. *(Lemma 4.3, Scheller-Wolf [65]) If two initially empty FIFO queues having K-1 and K servers, respectively, are fed by two identical interarrival and service time sequences and in addition, the queue with K-1 servers serves customers twice as fast as the queue with K servers, then for all $1 \leq l \leq K-1$ and all [arrivals] n, it holds almost surely that:*

$$W_{n,i-1}^{\{K-1\}} \leq W_{n,1}^{\{K\}} + W_{n,i}^{\{K\}}. \tag{3.1}$$

Applying Lemma 3.2 to our system with $i = 2$, we see that $W_{n,1}^{\{K-1\}} \leq W_{n,1}^{\{K\}} + W_{n,2}^{\{K\}}$ almost surely. Because $W_{n,1}^{\{K\}} \leq W_{n,2}^{\{K\}}$ by definition, we can see that $W_{n,1}^{\{K-1\})} \leq 2W_{n,2}^{\{K\}}$. Therefore $P(W_{n,1}^{\{K-1\}} > Cx) \leq P(2W_{n,2}^{\{K\}} > Cx)$. By rearranging and recognizing that the work at server 1 is equivalent to the delay of a system, we find $P(D^{\{K-1\}} > Cx) \leq P(W_{n,2}^{\{K\}} > \frac{C}{2}x)$: the probability that the work at the second server of the K-server system with $\rho = 1$ exceeds $\frac{C}{2}x$ is greater than or equal to the probability that the delay of a (K-1)-server system with $\rho = \frac{1}{2}$ exceeds Cx.

3.2.1.3 $t = x$, $x > 0$

Because x time has passed, x work has been processed at servers 2 through K since time 0. We consider $x = nE[S]$. However, during this same period of time, additional work may have arrived and been sent to server 1. There is thus a possibility that work has accumulated at server 1. To bound this possibility, we consider additional lemmas.

We are interested in activity over a time period of x. We assume that arrival times are deterministic with $T \equiv E[S]$. However, the following analysis will still hold for general interarrival times, due to the following result from Foss and Korshunov [28].

Lemma 3.3. *Consider a GI/GI/s system with stationary waiting time W. Now consider an auxiliary D/GI/s system with the same service times and deterministic arrival times with stationary waiting time W'.*

If $P\{W' > x\} \geq \bar{G}(x)$ *for some long-tailed distribution G, then*

$$\liminf_{x\to\infty} \frac{P\{W > x\}}{\bar{G}(x)} \geq 1 \tag{3.2}$$

(Foss and Korshunov Lemma 2 [28])

Now, we consider the behavior of the sum of the netput process in the GI/GI/K system with with $W_{0,2}^{\{K\}} \geq \frac{C}{2}x$ and $\rho = 1$. Work on the order of $x^{\frac{1}{\alpha}}$ will accumulate at the server with some probability, bounded by the following Lemma.

Lemma 3.4. *Let* $Y_i = S_i - T$, *where* T *is deterministic and equal to E[S], and S has* $P(X > u) \sim u^{-\alpha}$ *with* $1 < \alpha < 2$. *Then,*

$$P(\frac{\Sigma Y_i}{n^{\frac{1}{\alpha}}} > \epsilon) \sim U_\alpha(\epsilon) \tag{3.3}$$

where U is the stable distribution determined (including centering) by the characteristic function

$$\psi(\zeta) = |\zeta|^\alpha C \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)}[cos\frac{\pi\alpha}{2} \mp i(p-q)sin\frac{\pi\alpha}{2}]. \tag{3.4}$$

Proof. Follows from Feller [24]. (Although we followed Feller, similar results are discussed in Omey and Van Gulck [59] and Petrov [61] (found via [59]).) Please see Appendix B.1. □

Because interarrival time is equal to $E[S]$, n is the number of arrivals and we are considering the time period between 0 and x, we will set $n = \frac{x}{E[S]}$ in Lemma 3.4.

$$P(\frac{\Sigma Y_i}{(x/E[S])^{\frac{1}{\alpha}}} > \epsilon) \sim U_\alpha(\epsilon) \tag{3.5}$$

To lower bound the delay in the system, we will assume $W_{0,1}^{\{K\}} = 0$. Therefore, because servers 2 through K have a large amount of work, this new accumulated work is assigned to server 1. (If some of this work would have been assigned to server 2 we discard it, again yielding a lower bound.) Using equation (3.5), we see that this accumulated work is greater than $\epsilon E[S]^{-\frac{1}{\alpha}} x^{\frac{1}{\alpha}}$ with some probability $P_2(\epsilon) = U_\alpha(\epsilon) > 0$. Consequently, the updated workload at time x is as follows.

$W_{x,1}^{\{K\}} \geq \epsilon E[S]^{-\frac{1}{\alpha}} x^{\frac{1}{\alpha}}$

$W_{x,2}^{\{K\}} > (\frac{C}{2} - 1)x$

...

$W_{x,Q}^{\{K\}} > (\frac{C}{2} - 1)x$

Recall that $C > 2 + 2\epsilon E[S]^{-\frac{1}{\alpha}}$ and $\epsilon > 0$. Coupled with the fact that α must be greater than 1 because of the finite mean of the service time distribution, this choice ensures that $(\frac{C}{2} - 1)x > \epsilon E[S]^{-\frac{1}{\alpha}} x^{\frac{1}{\alpha}}$, which guarantees that $\epsilon E[S]^{-\frac{1}{\alpha}} x^{\frac{1}{\alpha}}$ will lower bound the smallest workload, and therefore, the delay of the system.

To summarize our findings so far, delay in a K-server system with $\rho = 1$ will be at least $\epsilon E[S]^{-\frac{1}{\alpha}} x^{\frac{1}{\alpha}}$ with some probability $P_2(\epsilon)$ if the work at the second server exceeds $\frac{C}{2}$, which will happen with probability $P_1(x)$ which is greater than or equal to the probability that the delay of a (K-1)-server system with $\rho = \frac{1}{2}$ exceeds Cx:

$$\begin{aligned} P(D^{\{K\}} > \epsilon E[S]^{-\frac{1}{\alpha}} x^{\frac{1}{\alpha}}) &\geq P_1(x) P_2(\epsilon) \\ &\geq P_2(\epsilon) P(W_{n,2}^{\{K\}} > \frac{C}{2} x) \\ &\geq P_2(\epsilon) P(D^{\{K-1\}} > Cx) \end{aligned}$$

Changing variables ($y = x^{\frac{1}{\alpha}}$):

$$P(D^{\{K\}} > \epsilon E[S]^{-\frac{1}{\alpha}} y) \geq P_2(\epsilon) P(D^{\{K-1\}} > C y^\alpha)$$

Multiplying by $r y^{r-1}$:

$$ry^{r-1}P(D^{\{K\}} > \epsilon E[S]^{-\frac{1}{\alpha}}y) \geq ry^{r-1}P_2(\epsilon)P(D^{\{K-1\}} > Cy^{\alpha})$$
$$\int_0^\infty ry^{r-1}P(\frac{D^{\{K\}}}{\epsilon E[S]^{-\frac{1}{\alpha}}} > y)dy \geq \int_0^\infty ry^{r-1}P_2(\epsilon)P(\frac{D^{\{K-1\}}}{C} > y^{\alpha})dy$$
$$\frac{E[S]^{\frac{1}{\alpha}}}{\epsilon}E[D^{\{K\}r}] \geq \int_0^\infty ry^{r-1}P_2(\epsilon)P(\frac{D^{\{K-1\}}}{C} > y^{\alpha})dy$$

For the next step, we use the follow result:

Lemma 3.5. *Wolff [77], page 37, for* $\gamma > 0$

$$E[X^\gamma] = \int_0^\infty \gamma u^{\gamma-1}P(X > u)du$$

Substituting $z = y^\alpha$ then applying Lemma 3.5:

$$\frac{E[S]^{\frac{1}{\alpha}}}{\epsilon}E[D^{\{K\}r}] \geq \int_0^\infty rz^{\frac{r-1}{\alpha}}P_2(\epsilon)P(\frac{D^{\{K-1\}}}{C} > z)\frac{z^{\frac{1}{\alpha}-1}}{\alpha}dz$$
$$\frac{E[S]^{\frac{1}{\alpha}}}{\epsilon}E[D^{\{K\}r}] \geq P_2(\epsilon)E[(\frac{D^{\{K-1\}}}{C})^{\frac{r}{\alpha}}]$$
$$\frac{E[S]^{\frac{1}{\alpha}}}{\epsilon}E[D^{\{K\}r}] \geq P_2(\epsilon)C^{\frac{-r}{\alpha}}E[(D^{\{K-1\}})^{\frac{r}{\alpha}}]$$
$$E[D^{\{K\}r}] \geq \frac{\epsilon}{E[S]^{\frac{1}{\alpha}}}P_2(\epsilon)C^{\frac{-r}{\alpha}}E[(D^{\{K-1\}})^{\frac{r}{\alpha}}]$$

So now we turn to $E[(D^{\{K-1\}})^{\frac{r}{\alpha}}]$, a value for which results already exist.

3.2.2 Delay in the (K-1)-server system

For information about the delay of this (K-1)-server system, we consider results from Scheller-Wolf [65].

Lemma 3.6. *(Lemma 6.6, Scheller-Wolf [65]) For a FIFO GI/GI/K queue with* $k < \rho < k+1 \leq K$, *where* $k \in N$ *and* $S \in \mathcal{L}^{\gamma+1}$ *for* $\gamma > 0$*:*

$$E[S^{1+\frac{\gamma}{K-k}}] = \infty \Rightarrow E[D^\gamma] = \infty$$

As a reminder, while the K-server system has $\rho = 1$, the (K-1)-server system has $\rho = \frac{1}{2}$. Additionally, recall that we are assuming that $S \in \mathcal{L}_1^{\alpha+1}$. Applying Lemma 3.6 with $k = 0$, $K-1$ servers and $\gamma = r/\alpha$, we see that $E[S^{1+\frac{r}{\alpha(K-1)}}] = \infty \Rightarrow E[D^{\{K-1\}\frac{r}{\alpha}}] = \infty$.

$$E[D^{\{K\}r}] \geq \frac{\epsilon}{E[S]^{\frac{1}{\alpha}}} P_2(\epsilon) C^{\frac{-r}{\alpha}} E[D^{\{K-1\}\frac{r}{\alpha}}]$$
$$E[S^{1+\frac{r}{\alpha(K-1)}}] = \infty \Rightarrow E[D^{\{K-1\}\frac{r}{\alpha}}] = \infty$$
$$E[S^{1+\frac{r}{\alpha(K-1)}}] = \infty \Rightarrow E[D^{\{K\}r}] = \infty$$

3.2.3 Conditions for infinite $E[S^{1+\frac{r}{\alpha(K-1)}}]$

Using Lemma 3.5 with $\gamma = 1 + \frac{r}{\alpha(K-1)}$ and assuming $P(X > u) \sim u^{-\alpha}$:

$$\begin{aligned}
E[X^\gamma] &= \int_0^\infty \gamma u^{\gamma-1} P(X > u) du \\
E[S^{1+\frac{r}{\alpha(K-1)}}] &= \int_0^\infty (1 + \frac{r}{\alpha(K-1)}) u^{\frac{r}{\alpha(K-1)}} P(S > u) du \\
&= (1 + \frac{r}{\alpha(K-1)}) \int_0^\infty u^{\frac{r}{\alpha(K-1)}} u^{-\alpha} du \\
&= (1 + \frac{r}{\alpha(K-1)}) \int_0^\infty u^{\frac{r}{\alpha(K-1)}-\alpha} du
\end{aligned}$$

We see that that the limit is infinite when $\frac{r}{\alpha(K-1)} - \alpha > -1$.

$$\begin{aligned}
\frac{r}{\alpha(K-1)} - \alpha &> -1 \\
\frac{r}{\alpha(K-1)} - \alpha + 1 &> 0 \\
r - \alpha^2(K-1) + \alpha(K-1) &> 0 \\
(K-1)\alpha^2 - (K-1)\alpha - r &< 0
\end{aligned}$$

So $E[S^{1+\frac{r}{\alpha(K-1)}}]$ and consequently $E[D^{\{K\}r}]$ is infinite when $(K-1)\alpha^2 - (K-1)\alpha - r < 0$, or when $\alpha < \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{r}{K-1}}$. These fall above Scheller-Wolf and Vesilo's [67] previously established lower bounds of $\frac{K+r}{K}$ when $r = 1$ as shown in the table below. This is true for all $K > 2$ and $0 < \rho < K$, although the difference between the old and new bounds decreases to 0 asymptotically. [Shown in Appendix B.2, using $R = 1$.] (When $r > \frac{K}{K-1}$, the new bounds are still valid, but fall below the previously established bounds.)

K	Previous Lower Bound	New Lower Bound
2	1.500	1.618
3	1.333	1.366
4	1.250	1.264
5	1.200	1.207

TABLE 3.1: Old and new lower bounds for α for different numbers of servers K, with $\rho = 1$

3.3 GI/GI/K, $\rho = R$, $R \geq 2$, $R < K$, $R \in N$

Now we extend the above results to FIFO GI/GI/K queues with integral $\rho \geq 2$. As before, we will consider the lower bounds of the workload at each server in this system as time progresses. We will show that, with positive probability, the delay of the system will be proportional to $x^{\frac{1}{\alpha}}$ in a deterministic arrival system (without loss of generality, as per Lemma 3.3). We can bound this probability by comparing with a system with $K-1$ servers and $\rho = \frac{R}{2}$. Finally, we find conditions on S that ensure an infinite expected delay.

3.3.1 Bounding Delay

3.3.1.1 $t = -M$

At time $-M$, we assume the system is empty. Time period M is sufficiently large that there is a probability P_1 that the $R+1^{\text{th}}$ server will accumulate more than $\frac{C}{2^R}x$ work during this time period (where C is chosen to be large enough that $C > \left(\frac{R^{R-1}-1}{\epsilon R^{R-2}(R-1)}\right)^{\alpha}(2^{R+1}RE[S])$ and $\epsilon > 0$, for reasoning explained in Section 3.3.1.8).

3.3.1.2 $t = 0$

We assume, for a given x, that server R+1 contains more than $\frac{C}{2^R}x$ work. This will occur with probability $P_1(x)$. For clarity of exposition, we will designate $\Upsilon = \frac{C}{2^R}$. Consequently, servers R+2 through K must each contain more than Υx work. The lower bounds on the work at each server are thus as follows.

$W_{0,1}^{\{K\}} \geq 0$

$W_{0,2}^{\{K\}} \geq 0$

...

$W_{0,R}^{\{K\}} \geq 0$

$W_{0,R+1}^{\{K\}} > \Upsilon x$

$W_{0,R+2}^{\{K\}} > \Upsilon x$

...

$W_{0,K}^{\{K\}} > \Upsilon x$

Now, we can find bounds for $P_1(x)$, the probability that server R+1 contains more than Υx work.

Applying Lemma 3.2 to our system with $s = K$ and $i = R + 1$, we see that $W_{n,R}^{\{K-1\}} \leq W_{n,1}^{\{K\}} + W_{n,R+1}^{\{K\}}$. Because $W_{n,1}^{\{K\}} \leq W_{n,R+1}^{\{K\}}$ by definition, we can see that $W_{n,R}^{\{K-1\}} \leq 2W_{n,R+1}^{\{K\}}$. Therefore $P(W_{n,R}^{\{K-1\}} > Cx) \leq P(2W_{n,R+1}^{\{K\}} > Cx)$. Proceeding iteratively, we find $W_{n,1}^{\{K-R\}} \leq 2W_{n,2}^{\{K-R+1\}} \leq 4W_{n,3}^{\{K-R+2\}} \leq ... \leq 2^R W_{n,R+1}^{\{K\}}$. Therefore $P(W_{n,1}^{\{K-R\}} > Cx) \leq P(2^R W_{n,R+1}^{\{K\}} > Cx)$. Please note that while the K-server system has $\rho = R$, the (K-R)-server system has $\rho = \frac{R}{2^R}$.

$$\begin{aligned} P(W_{n,1}^{\{K-R\}} > Cx) &\leq P(2^R W_{n,R+1}^{\{K\}} > Cx) \\ &\leq P(W_{n,R+1}^{\{K\}} > \frac{C}{2^R} x) \end{aligned}$$

or

$$P(D^{\{K-R\}} > Cx) \leq P(W_{n,R+1}^{\{K\}} > \frac{C}{2^R} x)$$

3.3.1.3 $t = \frac{\Upsilon}{2}x$, $x > 0$

We are interested in activity over a time period of $\frac{\Upsilon}{2}x$, $\frac{\Upsilon}{2}x = nRE[S]$. We assume that arrival times are deterministic with $T \equiv RE[S]$. These bounds will still hold for general interarrival times, as per Lemma 3.3. Because interarrival time is equal to $RE[S]$, n is the number of arrivals and we are considering the time period between 0 and $\frac{\Upsilon}{2}x$, we will set $n = \frac{\Upsilon}{2RE[S]}x$ in Lemma 3.4. (Please note that since $\rho = R$ and $E[S] = RT$, so $E[Y] = 0$ and Lemma 3.4 is still valid.)

Because server R+1 is occupied by a large job, this new accumulated work is distributed between servers 1 through R. (The event of an extremely large job, i.e. larger than the workload at server R+1, will give similar results. The reasoning for this is described at the end of this section.) Using equation (3.5), we see that this accumulated work is greater than $\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}}$ with some probability $P_2(\epsilon) = \Omega_\alpha(\epsilon) > 0$. The updated workload is then as follows:

$W^{\{K\}}_{\frac{\Upsilon}{2}x,1} + W^{\{K\}}_{\frac{\Upsilon}{2}x,2} + ... + W^{\{K\}}_{\frac{\Upsilon}{2}x,R} > \epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}}$

$W^{\{K\}}_{\frac{\Upsilon}{2}x,R+1} > \frac{1}{2}\Upsilon x$

$W^{\{K\}}_{\frac{\Upsilon}{2}x,R+2} > \frac{1}{2}\Upsilon x$

...

$W^{\{K\}}_{\frac{\Upsilon}{2}x,K} > \frac{1}{2}\Upsilon x$

This new work could be one large job, located at server R, or it could be a series of smaller jobs, exactly and evenly divisible between servers 1 through R, or the division could fall somewhere between these two extremes. In any case, $W^{\{K\}}_{0,R} \geq \frac{1}{R}\epsilon(\frac{1}{2RE[S]}\Upsilon x)^{\frac{1}{\alpha}}$, so the updated workload is no smaller than

$W^{\{K\}}_{\frac{\Upsilon}{2}x,1} \geq 0$

$W^{\{K\}}_{\frac{\Upsilon}{2}x,2} \geq 0$

...

$W^{\{K\}}_{\frac{\Upsilon}{2}x,R} \geq \frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}}$

$W^{\{K\}}_{\frac{\Upsilon}{2}x,R+1} > \frac{1}{2}\Upsilon x$

$W^{\{K\}}_{\frac{\Upsilon}{2}x,R+2} > \frac{1}{2}\Upsilon x$

...

$W^{\{K\}}_{\frac{\Upsilon}{2}x,K} > \frac{1}{2}\Upsilon x$

In the event of an extremely large job (i.e. larger than the workload at server R+1), this job will cause a reordering that results in the server R+1 becoming server R. In this case, the workload at server R will be greater than $\frac{1}{2}\Upsilon x$, which is greater than $\epsilon(\frac{\frac{1}{2RE[S]}\Upsilon}{2}x)^{\frac{1}{\alpha}}$. Similar reasoning holds for the time periods to follow.

3.3.1.4 $t = \frac{\Upsilon}{2}x + \frac{x^{\frac{1}{\alpha}}}{R}$, $x > 0$

At this point, the load remains R, but at most R-1 servers are free. Consequently, work begins accumulating linearly. This new work could be one large job, located at server

R-1, or it could be a series of smaller jobs, exactly and evenly divisible between servers 1 through R-1, or the division could fall somewhere between these two extremes. In the "worst" case for accumulation, work has been flowing into the system at a steady pace since time $t = \frac{\Upsilon}{2}x$, in very small jobs. In this event, the work will be evenly distributed among servers 1 through R-1, which will process the work as it arrives (but not completely, because there is 1 "extra" server's work arriving). At time $t = \frac{\Upsilon}{2}x + \frac{x^{\frac{1}{\alpha}}}{R}$ we know that even in this most distributed of cases, $W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R},R-1} \geq \frac{1}{R-1}\frac{x^{\frac{1}{\alpha}}}{R}$, so the updated workload is lower bounded by

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R},1} \geq 0$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R},2} \geq 0$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R},R-1} \geq \frac{1}{R-1}\frac{x^{\frac{1}{\alpha}}}{R}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R},R} \geq \frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}} - \frac{x^{\frac{1}{\alpha}}}{R}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R},R+1} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R},R+2} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R}$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R},K} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R}$$

As before, this remains true even in the event of an extremely large job.

3.3.1.5 $t = \frac{\Upsilon}{2}x + \frac{x^{\frac{1}{\alpha}}}{R} + \frac{x^{\frac{1}{\alpha}}}{R^2}$, $x > 0$

At this point, the load remains R, but at most only R-2 servers are free. Work continues accumulating linearly. This new work could be one large job, located at server R-2, or it could be a series of smaller jobs, exactly and evenly divisible between servers 1 through R-2, or the division could fall somewhere between these two extremes. In any case, $W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}],R-2} \geq \frac{2}{R-2}\frac{x^{\frac{1}{\alpha}}}{R^2}$, so the updated workload is lower bounded by

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2},1} \geq 0$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2},2} \geq 0$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2},R-2} \geq \frac{2}{R-2}\frac{x^{\frac{1}{\alpha}}}{R^2}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2},R-1} \geq \frac{1}{R-1}\frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2},R} \geq \frac{1}{R}\epsilon\left(\frac{\Upsilon}{2RE[S]}x\right)^{\frac{1}{\alpha}} - \frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2},R+1} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2},R+2} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2}$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2},K} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2}$$

At this point, we suspend the reordering of servers based on workload. Note that the workload at server R-1 is smaller than the workload at server R-2. This discrepancy is not important for the arguments to follow. The key is that there are a series of servers filled with large jobs (those labeled R+1 and higher), those filled with workloads proportional to $x^{\frac{1}{\alpha}}$ (those labeled R, R-1 and R-2), and those that may be empty (servers labeled 1 through R-3).

We will show that after sufficient time passes, you may have a situation where servers R+1 through K are filled with larger workloads than servers 1 through R (we ensure this through our choice of C). Servers 1 through R will have smaller workloads than servers R+1 through K, but their workloads will be lighter (due to the choice of C), greater than 0 (proven in Section 3.3.1.8), and proportional to $x^{\frac{1}{\alpha}}$. The precise ordering of servers 1 through R is not important, as we only need to show that one of these servers will have the smallest load, and therefore, indicates the delay of the system. As mentioned before, the arrival of a particularly large job does not change this reasoning. In the most extreme case, we would see a series of extremely large jobs, resulting in all servers being blocked except for that currently labeled R. This would then be the lightest workload server, and it is proportional to $x^{\frac{1}{\alpha}}$.

We will continue moving forward by time steps of $\frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$, where i indicates the next “free” server to be considered.

3.3.1.6 $t = \frac{\Upsilon}{2}x + \frac{x^{\frac{1}{\alpha}}}{R} + \frac{x^{\frac{1}{\alpha}}}{R^2} + ... + \frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$, $x > 0$

At this point, the load remains R, but only i servers are free, where $i \in [1, R-1]$. Work continues linear accumulation. This new work could be one large job, located at server i, or it could be a series of smaller jobs, exactly and evenly divisible between servers 1 through i, or the division could fall somewhere between these two extremes. In any case, $W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},i} \geq \frac{R-i}{i}\frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$, so the updated workload is lower bounded by

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},1} \geq 0$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},2} \geq 0$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},i} \geq \frac{R-i}{i}\frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},R-2} \geq \frac{2}{R-2}\frac{x^{\frac{1}{\alpha}}}{R^2} - ... - \frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},R-1} \geq \frac{1}{R-1}\frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2} - ... - \frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},R} \geq \frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}} - \frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2} - ... - \frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},R+1} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2} - ... - \frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},R+2} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2} - ... - \frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\frac{x^{\frac{1}{\alpha}}}{R}+\frac{x^{\frac{1}{\alpha}}}{R^2}+...+\frac{x^{\frac{1}{\alpha}}}{R^{R-i}},K} > \frac{1}{2}\Upsilon x - \frac{x^{\frac{1}{\alpha}}}{R} - \frac{x^{\frac{1}{\alpha}}}{R^2} - ... - \frac{x^{\frac{1}{\alpha}}}{R^{R-i}}$$

We can follow this reasoning until we reach time $t = \frac{\Upsilon}{2}x + \frac{x^{\frac{1}{\alpha}}}{R} + \frac{x^{\frac{1}{\alpha}}}{R^2} + ... + \frac{x^{\frac{1}{\alpha}}}{R^{R-i}} + ... + \frac{x^{\frac{1}{\alpha}}}{R^{R-2}}$, at which point servers 2 through K will all be busy.

3.3.1.7 $t = \frac{\Upsilon}{2}x + \sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n}$, $x > 0$

The updated workload is lower bounded by

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},1} \geq 0$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},2} \geq \frac{R-2}{2}\frac{x^{\frac{1}{\alpha}}}{R^{R-2}}$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},i} \geq \frac{R-i}{i}\frac{x^{\frac{1}{\alpha}}}{R^{R-i}} - \sum_{n=R-i+1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n}$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},R-2} \geq \frac{2}{R-2}\frac{x^{\frac{1}{\alpha}}}{R^2} - \sum_{n=3}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},R-1} \geq \frac{1}{R-1}\frac{x^{\frac{1}{\alpha}}}{R} - \sum_{n=2}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},R} \geq \frac{1}{R}\epsilon\left(\frac{\Upsilon}{2RE[S]}x\right)^{\frac{1}{\alpha}} - \sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},R+1} > \frac{1}{2}\Upsilon x - \sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},R+2} > \frac{1}{2}\Upsilon x - \sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n}$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n},K} > \frac{1}{2}\Upsilon x - \sum_{n=1}^{R-2}\frac{x^{\frac{1}{\alpha}}}{R^n}$$

Now, only server 1 is free.

3.3.1.8 $t = \frac{\Upsilon}{2}x + \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n}$, $x > 0$

All of the new accumulated work must be sent to server 1.

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},1} \geq x^{\frac{1}{\alpha}}(R-1)\frac{1}{R^{R-1}}$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},2} \geq x^{\frac{1}{\alpha}}\left[\frac{R-2}{2}\frac{1}{R^{R-2}} - \sum_{n=R-1}^{R-1}\frac{1}{R^n}\right]$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},i} \geq x^{\frac{1}{\alpha}}\left[\frac{R-i}{i}\frac{1}{R^{R-i}} - \sum_{n=R-i+1}^{R-1}\frac{1}{R^n}\right]$$

...

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},R-2} \geq x^{\frac{1}{\alpha}}\left[\frac{2}{R-2}\frac{1}{R^2} - \sum_{n=3}^{R-1}\frac{1}{R^n}\right]$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},R-1} \geq x^{\frac{1}{\alpha}}\left[\frac{1}{R-1}\frac{1}{R} - \sum_{n=2}^{R-1}\frac{1}{R^n}\right]$$

$$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},R} \geq x^{\frac{1}{\alpha}}\left[\frac{1}{R}\epsilon\left(\frac{\Upsilon}{2RE[S]}\right)^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{1}{R^n}\right]$$

$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},R+1} > \frac{1}{2}\Upsilon x - \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n}$

$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},R+2} > \frac{1}{2}\Upsilon x - \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n}$

...

$W^{\{K\}}_{\frac{\Upsilon}{2}x+\sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n},K} > \frac{1}{2}\Upsilon x - \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n}$

Remember that $C > \left(\frac{R^{R-1}-1}{\epsilon R^{R-2}(R-1)}\right)^{\alpha}(2^{R+1}RE[S])$ and $\epsilon > 0$. Coupled with the fact that α must be greater than 1 because of the finite mean of the service time distribution, this choice ensures that the workload at server R is greater than $C'x^{\frac{1}{\alpha}}$, where C' is a positive constant. This is proved in Appendix B.3.

To summarize our findings so far, the workload at servers 1 through R is $C'x^{\frac{1}{\alpha}}$, where C' is a constant that differs by server. Likewise servers R+1 through K have workloads no smaller than $C'x^{\frac{1}{\alpha}}$, because the workload at servers R+1 through K will be greater than that at server R. That is, $\frac{1}{2}\Upsilon x$ is greater than $\frac{\epsilon}{R}(\frac{\Upsilon x}{2RE[S]})^{\frac{1}{\alpha}}$. Consequently, we know that one of the servers 1 through R must have the smallest workload. As mentioned above, server R has a workload greater than $C'x^{\frac{1}{\alpha}}$. Servers 1 through R-1 will also have workloads greater than $C'x^{\frac{1}{\alpha}}$. Please see Appendix B.3 for a proof of this. As mentioned before, we do not establish an ordering for the workloads of servers 1 through R, except to say that all are of the format $C'x^{\frac{1}{\alpha}} > 0$. Consequently, delay in a K-server system with integral $\rho = R > 1$ will be at least $C'x^{\frac{1}{\alpha}}$ with some probability $P_2(\epsilon)$ if the work at the R+1 server exceeds Υx, which will happen with probability $P_1(x)$ which is greater than or equal to the probability that the delay of a (K-R)-server system with $\rho = \frac{R}{2^R}$ exceeds $\frac{C}{2^R}x$.

$$\begin{aligned} P(D^{\{K\}} > C'x^{\frac{1}{\alpha}}) &\geq P_1(x)P_2(\epsilon) \\ &\geq P_2(\epsilon)P(W^{\{K\}}_{n,R+1} > \frac{C}{2^R}x) \\ &\geq P_2(\epsilon)P(D^{\{K-R\}} > Cx) \end{aligned}$$

or taking $y = x^{\frac{1}{\alpha}}$

$$P(D^{\{K\}} > C'y) \geq P_2(\epsilon)P(D^{\{K-R\}} > Cy^{\alpha})$$

Multiplying by ry^{r-1} and integrating:

$$\int_0^\infty ry^{r-1}P(\frac{1}{C'}D^{\{K\}} > y)dy \geq \int_0^\infty ry^{r-1}P_2(\epsilon)P(\frac{D^{\{K-R\}}}{C} > y^\alpha)dy$$
$$\frac{1}{C'}E[D^{\{K\}r}] \geq \int_0^\infty ry^{r-1}P_2(\epsilon)P(\frac{D^{\{K-R\}}}{C} > y^\alpha)dy$$

Substituting $z = y^\alpha$:

$$\begin{aligned}\frac{1}{C'}E[D^{r\{K\}}] &\geq \int_0^\infty rz^{\frac{r-1}{\alpha}}P_2(\epsilon)P(\frac{D^{\{K-R\}}}{C} > z)\frac{z^{\frac{1}{\alpha}-1}}{\alpha}dz \\ &\geq \int_0^\infty \frac{r}{\alpha}z^{\frac{r}{\alpha}-1}P_2(\epsilon)P(\frac{D^{\{K-R\}}}{C} > z)dz \\ &\geq \frac{P_2(\epsilon)}{\alpha}E[(\frac{D^{\{K-R\}}}{C})^{\frac{r}{\alpha}}] \\ &\geq \frac{P_2(\epsilon)}{\alpha}C^{\frac{-r}{\alpha}}E[(D^{\{K-R\}})^{\frac{r}{\alpha}}]\end{aligned}$$

Therefore

$$E[D^{\{K\}r}] \geq C'\frac{P_2(\epsilon)}{\alpha}C^{\frac{-r}{\alpha}}E[(D^{\{K-R\}})^{\frac{r}{\alpha}}]$$

3.3.2 Delay in the (K-R)-server system

As a reminder, while the K-server system has $\rho = R$, the (K-R)-server system has $\rho = \frac{R}{2^R}$. Additionally, recall that we are assuming that $S \in \mathcal{L}^{\alpha+1}$. Applying Lemma 3.6 with $k = 0$ [because $0 < \frac{R}{2^R} < 1$ for $R \geq 2$], $K - R$ servers and $\gamma = r/\alpha$, we see that $E[S^{1+\frac{r}{\alpha(K-R)}}] = \infty \Rightarrow E[D^{\{K-R\}\frac{r}{\alpha}}] = \infty$.

Summarizing:

$$E[D^{\{K\}r}] \geq C'P_2(\epsilon)C^{\frac{-r}{\alpha}}E[D^{\{K-R\}\frac{r}{\alpha}}].$$

And

$$E[S^{1+\frac{r}{\alpha(K-R)}}] = \infty \Rightarrow E[D^{\{K-R\}\frac{r}{\alpha}}] = \infty.$$

So

$$E[S^{1+\frac{r}{\alpha(K-R)}}] = \infty \Rightarrow E[D^{\{K\}r}] = \infty.$$

3.3.3 Conditions for infinite $E[S^{1+\frac{r}{\alpha(K-R)}}]$

Using Lemma 3.5 With $\gamma = 1 + \frac{r}{\alpha(K-R)}$ and assuming $P(X > u) \sim u^{-\alpha}$

$$
\begin{aligned}
E[X^\gamma] &= \int_0^\infty \gamma u^{\gamma-1} P(S > u) du \\
E[S^{1+\frac{r}{\alpha(K-R)}}] &= \int_0^\infty (1 + \frac{r}{\alpha(K-R)}) u^{\frac{r}{\alpha(K-R)}} P(X > u) du \\
&= (1 + \frac{r}{\alpha(K-R)}) \int_0^\infty u^{\frac{r}{\alpha(K-R)}} u^{-\alpha} du \\
&= (1 + \frac{r}{\alpha(K-R)}) \int_0^\infty u^{\frac{r}{\alpha(K-R)}-\alpha} du
\end{aligned}
$$

We see that that the limit is infinite when $\frac{r}{\alpha(K-R)} - \alpha \geq -1$.

$$
\begin{aligned}
\frac{r}{\alpha(K-R)} - \alpha &\geq -1 \\
\frac{r}{\alpha(K-R)} - \alpha + 1 &\geq 0 \\
r - \alpha^2(K-R) + \alpha(K-R) &\geq 0 \\
(K-R)\alpha^2 - (K-R)\alpha - r &\geq 0
\end{aligned}
$$

So $E[S^{1+\frac{r}{\alpha(K-R)}}]$ and consequently $E[D^{\{K\}r}]$ is infinite when $(K-R)\alpha^2 - (K-R)\alpha - r < 0$, or when $\alpha < \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{r}{K-R}}$. These fall above above Scheller-Wolf and Vesilo's [67] previously established lower bounds of $\frac{K-R+1+r}{K-R+1}$ when $r = 1$ as shown in the table below, for all $K > 2$ and $0 < R < K$, although the difference between the old and new bounds decreases to 0 asymptotically. [Shown in Appendix B.2.] (When $r > \frac{K-R+1}{K-R}$, the new bounds fall below the previously established bounds.)

R	Previous Lower Bound	New Lower Bound
K-1	1.500	1.618
K-2	1.333	1.366
K-3	1.250	1.264
K-4	1.200	1.207

Table 3.2: Old and new lower bounds for α for different numbers of servers K, and different loads $\rho = R$

3.4 Conclusion

We have successfully extended the Scheller-Wolf and Vesilo [68] necessary conditions for finite mean delay with integral load for FIFO GI/GI/K queues, with service times belonging to class $\mathcal{L}_1^{1+\alpha}$ and service time distributions S such that $P(X > u) \sim u^{-\alpha}$ (such as Pareto), for those moments $r < 1 + \frac{1}{K-R}$, under certain conditions. Previous work established that the finite delay conditions of a system with integral ρ would be no worse than a system with $\rho + \epsilon$ and no better than a system with $\rho - \epsilon$. This work shows that the finite delay conditions will be strictly worse than a system with $\rho - \epsilon$, but whether more lenient conditions than those found for $\rho + \epsilon$ are available remains an open question.

Chapter 4

Revenue Management with Bargaining and a Finite Horizon

4.1 Introduction

Negotiation is commonplace in both business to business (B2B) and business to customer (B2C) interactions. Consider the situation of an airline. For each flight, the airline holds an inventory of seats that can only be sold before the plane takes off. They must decide how they will approach pricing, with the hope of obtaining the maximum payment for their inventory. Myerson [54] details four different price-deciding mechanisms. Applied to this example, the airline could set a price and let the buyer decide if it is acceptable (seller posted price, SPP), the airline could allow the buyer to set a price and then decide for itself if it is acceptable (buyer posted price, BPP), the airline and the buyer could each choose a price and then split-the-difference (STD), or the airline could conduct a negotiation with the buyer (represented by the neutral bargaining solution, NBS, developed by Myerson [53]). The seller must determine which of these options is expected to provide the greatest return. In this paper, we investigate the relative performance of these approaches from the seller's perspective.

Bhandari and Secomandi [8] consider the problem of a seller selling an inventory item-by-item to multiple, stochastically-arriving, non-strategic buyers over an infinite horizon sales period. The distributions of the seller's and buyer's valuations are known to both parties, but the actual valuations remain private. To model this situation, they develop a Markov decision process based on Myerson and Satterthwaite's [55] model. However, while Myerson and Satterthwaite consider a seller who has an exogenous valuation for a unit of inventory, Bhandari and Secomandi endogenously model the seller's remaining

This chapter is joint work with Nicola Secomandi.

inventory valuation as the (optimal) opportunity cost of that unit of inventory, generalizing the SPP-specific model of Das Varma and Vettas [19]. We extend the Bhandari and Secomandi [8] model to the more realistic finite horizon case. While our model is analogous, a different strategy is required for proving the structural results.

An important caveat is that in our model, the time remaining to sell the inventory is the private information of the seller. Consequently, the buyers do not have the necessary information to strategically exploit the seller's limited time. This assumption makes our model internally consistent, as each period is the same to the buyer, but is not very realistic--one possible scenario would be a seller of expiring printer ink, who would have a secret, but binding, sell-by date for the ink inventory. On a related note, the seller's inventory level is also the private information of the seller, but this situation occurs frequently in practice.

Analytically, we demonstrate that SPP always performs at least as well for the seller as NBS, which always performs at least as well as BPP, and that STD always performs at least as well as BPP. Compactly, $SPP \geq NBS \geq BPP$ and $STD \geq BPP$. More generally, we demonstrate that a mechanism with a higher interim expected utility will produce a higher value for the seller at a given time-to-go and inventory level, extending the analytical findings of Bhandari and Secomandi [8] to a finite-horizon case. Numerically, we find that the quantitative differences between the seller's optimal value function under the four considered mechanisms in Chatterjee and Samuelson's [15] symmetric uniform trading problem (SUTP) change when moving from an infinite to a finite time horizon. While in the infinite horizon case the STD mechanism can dominate the other mechanisms under extreme parameter values, we show that this same dominance occurs in the finite horizon case using much more plausible parameter values. This is an important finding because, although not as simple to use as SPP or BPP, STD is an easy to implement mechanism that can be used in a variety of practical settings. For instance, a buyer and seller could simply report their valuations to a webpage which could then return the negotiated price. The evenness of the split can even be adjusted to accommodate varying degrees of bargaining power, although we do not investigate this feature. From a broader perspective, while the NBS mechanism provides a normative representation of the outcome of face-to-face negotiation, the STD mechanism could be implemented as an automated negotiation.

Because modeling private information is more challenging for the STD and NBS mechanisms than for the SPP and BPP mechanisms, we quantify the importance of modeling private information when computing the seller's opportunity cost under the STD and NBS mechanisms. We find that modeling private information has only a small effect in the evaluation of the seller's value function in the NBS case, but has a substantial effect in STD cases where there is relatively high time-to-go and relatively little inventory. Consequently, it may be acceptable to use a simplified model that assumes voluntary disclosure of both the seller's and the buyers' valuations to compute the seller's opportunity cost under the NBS mechanism. This simplified model may also be acceptable under the STD mechanism, but only in cases where there is relatively high inventory and relatively little time-to-go.

Kuo, Ahn and Aydin [44] consider a similar problem, with a finite time horizon, but using a generalized Nash bargaining solution requiring the true valuations of both parties to be voluntarily revealed without the need to model incentive compatibility issues (complete information). They find that allowing negotiation (via the generalized Nash bargaining solution) can result in more favorable outcomes for the seller than a posted price, due to the advantage of price discrimination. Ayvaz-Cavdaroglu et al. [5] also build a model to study SPP and BPP within a finite horizon and private information setting, with the addition of unknown, non-stationary buyer and seller distributions. However, they focus only on posted pricing mechanisms.

Within the realm of mechanism comparison, Riley and Zeckhauser [62] consider the sale of a single unit of inventory and find SPP to be the best possible mechanism for the seller. Gallien [32] considers the sale of multiple units of inventory, and also finds SPP to be the best possible mechanism for the seller. However, both Riley and Zeckhauser and Gallien use a dominant equilibrium framework, which assesses a mechanism against all potential choices made by buyers, as opposed to a Bayesian Nash equilibrium framework, which assesses a mechanism against the beliefs the seller holds about the likely choices of the buyers. Significantly, STD is incompatible with the dominant equilibrium framework, and our (Bayesian Nash equilbrium based) findings suggest that STD can outperform SPP when the seller is weak. Also challenging the superiority of SPP are Wang [74] and Roth et al. [63] who demonstrate the advantages of the Nash bargaining solution (which significantly does not allow for private information).

In Section 4.2, we detail our model, which includes both a finite horizon and private information, then structurally analyze this model in Section 4.3. We discuss the relevance of modeling private information in Section 4.4. Numerical results are presented in Section 4.5. Finally, Section 4.6 includes a summary and thoughts on future work. Additional numerical results are included in Appendix C.

4.2 Model

In this section, we first consider Myerson and Satterthwaite's one-period static bargaining model [55] in Subsection 4.2.1. In Subsection 4.2.2, we discuss the application of this model to the SUTP. We then consider the different properties of mechanisms, as well as the application of Myerson's [54] four mechanisms to the SUTP. Finally, we extend these models into a new multi-period stochastic dynamic model in Subsection 4.2.3.

4.2.1 Static Bargaining Model

First, consider a static bargaining model by Myerson and Satterthwaite [55] and described by Bhandari and Secomandi [8], but with valuations scaled to [0,1], rather than the more general $[a, b]$. A risk neutral seller with one unit of inventory holds a valuation for the unit of $v_1 \in \mathcal{V}_1 = [0, 1]$. (Throughout this paper, we will use subscripts of 1 to refer to the seller, consistent with the notation in the literature.) A risk neutral buyer's valuation for the unit of inventory is $v_2 \in \mathcal{V}_2 = [0, 1]$. (We will use subscripts of 2 to refer to the buyer.) Each party knows his own valuation, but not the valuation of the other party. The buyer believes the seller's valuation to be drawn from a cumulative distribution function $F_1(v_1)$. Similarly, the seller believes the buyer's valuation to be drawn from distribution $F_2(v_2)$. $F_1(v_1)$ and $F_2(v_2)$ both have support $[0, 1]$, and both are public knowledge. An intermediary will confidentially request valuations from both the seller and buyer, and then apply some bilateral bargaining mechanism j to determine a candidate sale price. If this price is mutually acceptable, a sale will occur. Otherwise, the buyer will leave without a sale. We use $s^j(v_1, v_2) \in \{0, 1\}$ to indicate whether a sale occurs given that a buyer with valuation v_2 arrives, the seller has valuation v_1 and mechanism j is applied. We let $s^j(v_1, v_2) = 1$ if the sale occurs and 0 otherwise. Similarly, $x^j(v_1, v_2)$ will be the price paid under the same conditions. (If a sale does not occur, $x^j(v_1, v_2) = 0$.) The transfer probability under mechanism j for valuations v_1 and v_2 is $p^j(v_1, v_2) \in [0, 1]$. That is, $s^j(v_1, v_2) = 1$ with probability $p^j(v_1, v_2)$, and $s^j(v_1, v_2) = 0$ with probability $1 - p^j(v_1, v_2)$.

4.2.2 Mechanisms

The SUTP proposed by Chatterjee and Samuelson [15] follows the static bargaining model described above with $F_1(v_1) = v_1$ and $F_2(v_2) = v_2$. The following summary of four mechanisms, as well as their characteristic prices and transfer probabilities, was originally presented by Myerson [54] and Bhandari and Secomandi [8]. First, a seller may post a price that the buyer may accept or reject. In the SUTP, the corresponding transfer probability $p^{SPP}(v_1, v_2)$ will be equal to the probability that $v_2 \geq \frac{1+v_1}{2}$, and the price paid will be $x^{SPP}(v_1, v_2) = p^{SPP}(v_1, v_2)\frac{1+v_1}{2}$. Analogously, a buyer may post a price that the seller may accept or reject. In the SUTP, in this case the transfer probability $p^{BPP}(v_1, v_2)$ will be equal to the probability that $\frac{v_2}{2} \geq v_1$, and the price paid will be $x^{BPP}(v_1, v_2) = p^{BPP}(v_1, v_2)\frac{v_2}{2}$. The buyer and seller may both propose prices which would then be averaged (possibly weighted by bargaining power, however in the SUTP the two parties have equal bargaining power) to split-the-difference with a proposed "compromise" price, which the buyer and seller would then accept or reject. In the SUTP, the transfer probability $p^{STD}(v_1, v_2)$ will be equal to the probability that $v_2 \geq v_1 + \frac{1}{4}$ and the price paid will be $x^{STD}(v_1, v_2) = p^{STD}(v_1, v_2)\frac{v_1+v_2+\frac{1}{2}}{3}$. Finally, the neutral bargaining solution (Myerson [53]) corresponds to a seller posted price outcome when the seller has higher bargaining power, and a buyer posted price outcome when the buyer has higher bargaining power. In the SUTP, the transfer probability $p^{NBS}(v_1, v_2)$ will be equal to the probability that either $v_2 \geq 3v_1$ or $3v_2 - 2 \geq v_1$. The corresponding price paid will be $x^{NBS}(v_1, v_2) = p^{NBS}(v_1, v_2)\frac{v_2}{2}$ if $v_2 \leq 1 - v_1$ and $x^{NBS}(v_1, v_2) = p^{NBS}(v_1, v_2)\frac{1+v_1}{2}$ otherwise.

All four of the mechanisms described above are direct mechanisms, meaning that an intermediary produces price and transfer decisions for a buyer and seller who concurrently and confidentially report their valuations [54]. To formally understand the properties of these and other mechanisms, we can consider the interim expected utility of the seller and buyer, or the expected gap between the price and valuation of one unit for that party when negotiating with the other party. As before, the following is a summary of Myerson [54], using notation from Bhandari and Secomandi [8]. For the seller, this utility will be the expected price received less the valuation of the seller times the probability of the sale. For the buyer, it will be the valuation of the buyer times the probability of the sale minus the price paid. For ease of exposition, we define

$$\begin{aligned}
\bar{x}_1^j(v_1) &:= \int_{v_2 \in \mathcal{V}_2} x^j(v_1, \tilde{v}_2) dF_2(v_2), \\
\bar{x}_2^j(v_2) &:= \int_{v_1 \in \mathcal{V}_1} x^j(\tilde{v}_1, v_2) dF_1(v_1), \\
\bar{p}_1^j(v_1) &:= \int_{v_2 \in \mathcal{V}_2} p^j(v_1, \tilde{v}_2) dF_2(v_2), \\
\bar{p}_2^j(v_2) &:= \int_{v_1 \in \mathcal{V}_1} p^j(\tilde{v}_1, v_2) dF_1(v_1).
\end{aligned}$$

The term $\bar{x}_1^j(v_1)$ is the expected price from the seller's perspective, that is, with known seller's valuation v_1. Similarly, $\bar{x}_2^j(v_2)$ is the expected price from the buyer's perspective, with known buyer's valuation v_2. Similarly, $\bar{p}_1^j(v_1)$ and $\bar{p}_2^j(v_2)$ are the expected transfer probabilities for the seller and buyer, respectively. Using this notation, the interim expected utility for the seller is $\bar{u}_1^j(v_1) := \bar{x}_1^j(v_1) - v_1\bar{p}_1^j(v_1)$, and the interim expected utility for the buyer $\bar{u}_2^j(v_2) := v_2\bar{p}_2^j(v_2) - \bar{x}_2^j(v_2)$.

A direct mechanism is incentive compatible when the buyer and seller expect their highest utility outcomes from "reporting [their] true valuations" [54] to the intermediary, assuming truthful reporting by the other party. Formally, $\bar{u}_1^j(v_1) \geq \bar{x}_1^j(\hat{v}_1) - v_1\bar{p}_1^j(\hat{v}_1)$ $\forall v_1, \hat{v}_1 \in \mathcal{V}_1$ and $\bar{u}_2^j(v_2) \geq v_2\bar{p}_2^j(\hat{v}_2) - \bar{x}_2^j(\hat{v}_2)$ $\forall v_2, \hat{v}_2 \in \mathcal{V}_2$ [8].

A mechanism is individually rational when neither the buyer nor seller expect negative utility outcomes from the negotiation [54]. Formally, $\bar{u}_1^j(v_1) \geq 0$ $\forall v_1 \in \mathcal{V}_1$ and $\bar{u}_2^j(v_2) \geq 0$ $\forall v_2 \in \mathcal{V}_2$ [8].

Mechanisms that are "both individually rational and incentive compatible" are called *feasible* [54]. All four of the mechanisms described above are both direct and feasible, and our analytical results will focus on mechanisms with these properties. The Bayesian equilibrium of any mechanism can be represented using a direct and incentive compatible mechanism via the revelation principle, so this restriction is without loss of generality [54].

4.2.3 Stochastic and Dynamic Model

Now we extend the static case to a dynamic case. Consider a risk neutral seller with $y \in \mathcal{Y} = \{1, ..., Y\}$ units of inventory (where Y is the starting inventory), and $t \in \mathcal{T} = \{0, ..., T\}$ time periods in which that inventory may be sold, or "time-to-go" (where T is the first sale period). With 0 time-to-go, the value of the inventory is 0. Additionally, $\beta \in (0, 1]$ is the discount factor for the seller for one period. For each time period, the probability that a customer arrives is $\lambda \in (0, 1]$. That is, at most one customer may arrive per period. If either β or λ is zero, the problem becomes trivial. The quantities Y, y, T, β and λ are known only to the seller. While many realistic scenarios could feature a seller's private starting inventory, current inventory, discount factor and customer arrival rate, a private deadline T is rather rare. However, this assumption is required for the internal consistency of our model, which does not include buyers strategically exploiting the information of an impending sell-by date. The seller has a valuation for one unit of inventory v_1, but this valuation now depends on the amount of inventory remaining and time-to-go. Each buyer has a valuation for a unit of inventory $v_2 \in \mathcal{V}_2 = [0, 1]$. Each party knows his own valuation, but not the valuation of the other party. Each buyer believes the seller's valuation to be drawn from distribution $F_1(v_1)$. Similarly, the seller believes each buyer's valuation to be drawn from distribution $F_2(v_2)$. $F_1(v_1)$ and $F_2(v_2)$ both have support $[0, 1]$, and both are public knowledge. During each time period, if a buyer arrives, the seller and buyer will confidentially report their valuations to a mediator who will apply some direct and feasible mechanism j to determine a candidate sale price. If this price is mutually acceptable, a sale will occur. Otherwise, the buyer will leave without a sale. As before, we use $s^j(v_1, v_2)$ to indicate whether a sale occurs, $x^j(v_1, v_2)$ for the price paid, and $p^j(v_1, v_2)$ for the transfer probability.

To study this dynamic case, we develop a stochastic dynamic programming model. Our model is comparable to that of Bhandari and Secomandi [8], with the added complication of a finite horizon. We want to develop a model for $V_t^j(y)$, the optimal expected value of the y units of inventory remaining, given that we have t time periods left in which to sell the units.

At time t, a customer will not arrive with probability $(1-\lambda)$. In this case, the value to the seller of y units of inventory at time t is simply the value of y units of inventory at time $t-1$ discounted by β. The seller has no opportunity for action. A customer will arrive at time period t with probability λ. In this case, a price $x^j(v_1, \tilde{v}_2)$ is paid from buyer to seller (again, this price will be 0 if no sale occurs). Remember that while the seller must decide v_1, he only knows the distribution of v_2, hence we will use a tilde to indicate that $\tilde{v}_2$ is a random variable. If the sale occurs, the seller also has the discounted value of his remaining $y-1$ units to sell over $t-1$ more periods. If the sale does not occur, the seller has the discounted value of the full y units to sell over $t-1$ more periods. This result is the following stochastic dynamic programming model:

$$\begin{aligned} V_t^j(y) &= (1-\lambda)\beta V_{t-1}^j(y) \\ &\quad +\lambda \max_{v_1 \in \mathcal{V}_1} E[x^j(v_1, \tilde{v}_2) + \beta V_{t-1}^j(y-1) 1\{s^j(v_1, \tilde{v}_2) = 1\} \\ &\qquad +\beta V_{t-1}^j(y) 1\{s^j(v_1, \tilde{v}_2) = 0\}]. \end{aligned} \tag{4.1}$$

Because the seller will obtain nothing more if he has sold all of his inventory, we define $V_t^j(0) := 0$. Similarly, if the seller runs out of time, his inventory is valueless. Consequently, $V_0^j(y) := 0$. Model (4.1) can be rearranged as

$$\begin{aligned} V_t^j(y) &= (1-\lambda)\beta V_{t-1}^j(y) + \lambda \max_{v_1 \in \mathcal{V}_1} [\bar{x}_1^j(v_1) + \beta V_{t-1}^j(y-1)\bar{p}_1^j(v_1) + \beta V_{t-1}^j(y)(1-\bar{p}_1^j(v_1))] \\ &= \beta V_{t-1}^j(y) + \lambda \max_{v_1 \in \mathcal{V}_1} [\bar{x}_1^j(v_1) + \beta V_{t-1}^j(y-1)\bar{p}_1^j(v_1) - \beta V_{t-1}^j(y)\bar{p}_1^j(v_1)]. \end{aligned} \tag{4.2}$$

Finally, we will use the notation $\Delta V_t^j(y) := V_t^j(y) - V_t^j(y-1)$ to condense model (4.2):

$$V_t^j(y) = \beta V_{t-1}^j(y) + \lambda \max_{v_1 \in \mathcal{V}_1} [\bar{x}_1^j(v_1) - \beta \Delta V_{t-1}^j(y)\bar{p}_1^j(v_1)]. \tag{4.3}$$

The opportunity cost to the seller for the sale of the y^{th} unit of inventory at time t is equal to $\beta \Delta V_{t-1}^{j}(y)$. By the feasibility of mechanism j, this quantity will then be the optimal reported valuation of the seller, assuming that the opportunity cost lies within the support of $F_1(v_1)$, or $\beta \Delta V_{t-1}^{j}(y) \in \mathcal{V}_1 = [0,1]$. We prove that this is the case in Lemma 4.1.

4.3 Structural Analysis

In this section we demonstrate the ordering of value functions of the four mechanisms of interest. In order to show this, we first establish that the opportunity cost of the seller $\beta \Delta V_{t-1}^{j}(y) \in \mathcal{V}_1 \equiv [0,1]$ (Lemma 4.1), and is consequently the optimal choice for v_1 in model (4.3) (Proposition 1). Next, we demonstrate that the ordering of the seller's interim expected utilities between mechanisms corresponds to the ordering of value functions (Theorem 1). That is, a mechanism j with a higher interim expected utility than mechanism k will also have a higher value function value, for a given t and y. Combining this result with the ordering of interim expected utilities for the four mechanisms under study given by Bhandari and Secomandi [8] allows us to establish the ordering of the value functions of these mechanisms when each transaction opportunity is modeled consistently with the SUTP, for a given y and t (Proposition 2).

Lemma 4.1. *(a) Given direct and feasible mechanism j, the optimal value function $V_t^j(y)$ is weakly increasing at a non-increasing rate in inventory $\forall y \in \mathcal{Y} \cup \{0\}$. Equivalently, the function $\beta \Delta V_t^j(y)$ is nonnegative and weakly decreases in inventory $\forall y \in \mathcal{Y}$. (b) Moreover, it holds that $\beta \Delta V_t^j(y) \leq 1, \forall y \in \mathcal{Y}$.*

Lemma 4.1 is analogous to Lemma 2 in Bhandari and Secomandi [8], with the exception that our Lemma 4.1 does not require $\bar{u}_1^j(1) = 0$. The proof that follows, however, is necessarily different.

Proof. (a) First, we show that $V_t^j(y)$ is weakly increasing in inventory, using analogous reasoning to Bhandari and Secomandi [8].

As defined, $V_0^j(y) \equiv 0$ for all $y \in \mathcal{Y} \cup \{0\}$. Consequently, $\Delta V_0^j(y) \equiv 0$ for all $y \geq 1$.

We make the induction hypothesis $\Delta V_t^j(y) \geq 0$ for steps $1, ..., t-1$ and all $y \in \mathcal{Y}$. Consider time t. We define $v_{1,t}^*(y) \in \arg\max_{v_1 \in \mathcal{V}_1}[\bar{x}_1^j(v_1) - \beta \Delta V_{t-1}^j(y)\bar{p}_1^j(v_1)]$. We have

$$\begin{aligned}
\Delta V_t^j(y) &= V_t^j(y) - V_t^j(y-1) \\
&= \beta V_{t-1}^j(y) + \lambda \max_{v_1 \in \mathcal{V}_1}[\bar{x}_1^j(v_1) - \beta \Delta V_{t-1}^j(y)\bar{p}_1^j(v_1)] \\
&\quad -\beta V_{t-1}^j(y-1) - \lambda \max_{v_1 \in \mathcal{V}_1}[\bar{x}_1^j(v_1) - \beta \Delta V_{t-1}^j(y-1)\bar{p}_1^j(v_1)] \\
&= \beta \Delta V_{t-1}^j(y) + \lambda \bar{x}_1^j(v_{1,t}^*(y)) - \lambda\beta \Delta V_{t-1}^j(y)\bar{p}_1^j(v_{1,t}^*(y)) \\
&\quad -\lambda \bar{x}_1^j(v_{1,t}^*(y-1)) + \lambda\beta \Delta V_{t-1}^j(y-1)\bar{p}_1^j(v_{1,t}^*(y-1)) \\
&\geq \beta \Delta V_{t-1}^j(y) + \lambda \bar{x}_1^j(v_{1,t}^*(y-1)) - \lambda\beta \Delta V_{t-1}^j(y)\bar{p}_1^j(v_{1,t}^*(y-1)) \\
&\quad -\lambda \bar{x}_1^j(v_{1,t}^*(y-1)) + \lambda\beta \Delta V_{t-1}^j(y-1)\bar{p}_1^j(v_{1,t}^*(y-1)) \\
&= \beta[(1 - \lambda \bar{p}_1^j(v_{1,t}^*(y-1)))\Delta V_{t-1}^j(y) + \lambda \bar{p}_1^j(v_{1,t}^*(y-1))\Delta V_{t-1}^j(y-1)] \\
&\geq 0,
\end{aligned}$$

where the first inequality follows from the optimality of $v_{1,t}^*(y)$ in stage t and state y, and the second inequality from the observation that $0 \leq \beta\lambda\bar{p}_1^j(v_{1,t}^*(y-1)) \leq 1$ and the application of the induction hypothesis.

Consequently, $\Delta V_t^j(y) \geq 0$ for all $t \in \mathcal{T}$, and all $y \in \mathcal{Y}$ by the principle of mathematical induction.

Next, we show that $V_t^j(y)$ increases at a non-increasing rate in inventory. This property is trivially true in stage 0 because $V_0^j(y) \equiv 0$ for all $y \in \mathcal{Y} \cup \{0\}$.

We make the induction hypothesis $\Delta V_t^j(y) \leq \Delta V_t^j(y-1)$ for steps $1, ..., t-1$ and $\forall y \in \mathcal{Y}$. Consider time t. Proceeding as in the proof of part (a), but with respect to $\Delta V_t^j(y-1)$ yields

$$\begin{aligned}
\Delta V_t^j(y-1) &\geq \beta[(1 - \lambda \bar{p}_1^j(v_{1,t}^*(y-2)))\Delta V_{t-1}^j(y-1) + \lambda \bar{p}_1^j(v_{1,t}^*(y-2))\Delta V_{t-1}^j(y-2)] \\
&= \beta \Delta V_{t-1}^j(y-1) + \beta\lambda \bar{p}_1^j(v_{1,t}^*(y-2))[\Delta V_{t-1}^j(y-2) - \Delta V_{t-1}^j(y-1)] \\
&\geq \beta \Delta V_{t-1}^j(y-1), \qquad (4.4)
\end{aligned}$$

where the last inequality follows from the induction hypothesis and $\beta\lambda\bar{p}_1^j(v_{1,t}^*(y-1)) \in [0,1]$. We also have

$$\begin{aligned}
\Delta V_t^j(y) &= V_t^j(y) - V_t^j(y-1) \\
&= \beta V_{t-1}^j(y) + \lambda \max_{v_1 \in \mathcal{V}_1}[\bar{x}_1^j(v_1) - \beta \Delta V_{t-1}^j(y)\bar{p}_1^j(v_1)] \\
&\quad -\beta V_{t-1}^j(y-1) - \lambda \max_{v_1 \in \mathcal{V}_1}[\bar{x}_1^j(v_1) + \beta \Delta V_{t-1}^j(y-1)\bar{p}_1^j(v_1)] \\
&= \beta V_{t-1}^j(y) + \lambda \bar{x}_1^j(v_{1,t}^*(y)) - \lambda\beta\Delta V_{t-1}^j(y)\bar{p}_1^j(v_{1,t}^*(y)) \\
&\quad -\beta V_{t-1}^j(y-1) - \lambda \bar{x}_1^j(v_{1,t}^*(y-1)) + \lambda\beta\Delta V_{t-1}^j(y-1)\bar{p}_1^j(v_{1,t}^*(y-1)) \\
&\leq \beta \Delta V_{t-1}^j(y) + \lambda \bar{x}_1^j(v_{1,t}^*(y)) - \lambda\beta\Delta V_{t-1}^j(y)\bar{p}_1^j(v_{1,t}^*(y)) \\
&\quad -\lambda \bar{x}_1^j(v_{1,t}^*(y)) + \lambda\beta\Delta V_{t-1}^j(y-1)\bar{p}_1^j(v_{1,t}^*(y)) \\
&= \beta[(1-\lambda\bar{p}_1^j(v_{1,t}^*(y)))\Delta V_{t-1}^j(y) + \lambda\bar{p}_1^j(v_{1,t}^*(y))\Delta V_{t-1}^j(y-1)] \\
&\leq \beta[(1-\lambda\bar{p}_1^j(v_{1,t}^*(y)))\Delta V_{t-1}^j(y-1) + \lambda\bar{p}_1^j(v_{1,t}^*(y))\Delta V_{t-1}^j(y-1)] \\
&= \beta \Delta V_{t-1}^j(y-1), \qquad (4.5)
\end{aligned}$$

where the first inequality follows from the optimality of $v_{1,t}^*(y-1)$ in stage t and state $y-1$ and the second inequality follows from the induction hypothesis and $\lambda\bar{p}_1^j(v_{1,t}^*(y-1)) \in [0,1]$.

Now we apply (4.4) and (4.5) to bound from above the difference between $\Delta V_t^j(y)$ and $\Delta V_t^j(y-1)$:

$$\begin{aligned}
\Delta V_t^j(y) - \Delta V_t^j(y-1) &\leq \beta\Delta V_{t-1}^j(y-1) - \Delta V_t^j(y-1) \\
&\leq \beta\Delta V_{t-1}^j(y-1) - \beta\Delta V_{t-1}^j(y-1) \\
&= 0,
\end{aligned}$$

where the first inequality follows from inequality (4.5), and the second inequality from inequality (4.4). Consequently, $\Delta V_t^j(y) - \Delta V_t^j(y-1) \leq 0$ for all $t \in \mathcal{T}$, and all $y \in \mathcal{Y}$ by the principle of mathematical induction.

(b) We trivially have $\Delta V_0^j(y) = 0 \leq 1$ for all $y \in \mathcal{Y}$. We make the induction hypothesis that $\Delta V_t^j(y) \leq 1$ for all $t = 1, ..., t-1$ and all $y \in \mathcal{Y}$. In stage t we obtain

$$\begin{aligned}
\Delta V_t^j(y) &= V_t^j(y) - V_t^j(y-1) \\
&= \beta V_{t-1}^j(y) + \lambda \max_{v_1 \in \mathcal{V}_1} [\bar{x}_1^j(v_1) - \beta \Delta V_{t-1}^j(y) \bar{p}_1^j(v_1)] \\
&\quad -\beta V_{t-1}^j(y-1) - \lambda \max_{v_1 \in \mathcal{V}_1} [\bar{x}_1^j(v_1) - \beta \Delta V_{t-1}^j(y-1) \bar{p}_1^j(v_1)] \\
&= \beta V_{t-1}^j(y) + \lambda \bar{x}_1^j(v_{1,t}^*(y)) - \lambda \beta \Delta V_{t-1}^j(y) \bar{p}_1^j(v_{1,t}^*(y)) \\
&\quad -\beta V_{t-1}^j(y-1) - \lambda \bar{x}_1^j(v_{1,t-1}^*(y-1)) + \lambda \beta \Delta V_{t-1}^j(y-1) \bar{p}_1^j(v_{1,t-1}^*(y-1)) \\
&\leq \beta[1 - \lambda \bar{p}_1^j(v_{1,t}^*(y))] \Delta V_{t-1}^j(y) + \lambda \beta \Delta V_{t-1}^j(y-1) \bar{p}_1^j(v_{1,t}^*(y)) \\
&\quad +\lambda[\bar{x}_1^j(v_{1,t}^*(y)) - \bar{x}_1^j(v_{1,t}^*(y)] \\
&= \beta\{\Delta V_{t-1}^j(y) + \lambda \bar{p}_1^j(v_{1,t}^*(y))[\Delta V_{t-1}^j(y-1) - \Delta V_{t-1}^j(y)]\} \\
&\leq \Delta V_{t-1}^j(y) + \Delta V_{t-1}^j(y-1) - \Delta V_{t-1}^j(y) \\
&= \Delta V_{t-1}^j(y-1) \\
&\leq 1,
\end{aligned} \tag{4.6}$$

where the first inequality follows from the optimality of $v_{1,t}^*(y-1)$ at stage t and state $y-1$ and rearranging, the second inequality from part (a) and $\lambda \bar{p}_1^j(v_{1,t}^*(y)) \in [0,1]$, and the final inequality from the induction hypothesis. □

As in Bhandari and Secomandi [8], Lemma 4.1 and the incentive compatibility of mechanism j imply the following proposition, analogous to Proposition 1 in that paper.

Proposition 4.2. *If direct mechanism j is feasible then the seller's optimal value function satisfies the following conditions, for all t and y:*

$$V_t^j(y) = \beta V_{t-1}^j(y) + \lambda[\bar{x}_1^j(\beta \Delta V_{t-1}^j(y)) - \beta \Delta V_{t-1}^j(y) \bar{p}_1^j(\beta \Delta V_{t-1}^j(y))].$$

Proposition 4.2 combines our new stochastic dynamic program model with the interpretation of the seller's valuation v_1 as the seller's opportunity cost of one unit of inventory $\beta \Delta V_{t-1}^j(y)$ (incentive compatibility). The results of Lemma 4.1 demonstrate that this is acceptable, by showing that $0 \leq \beta \Delta V_{t-1}^j(y) \leq 1$.

Now that we have resolved the nature of our model, we can consider the circumstances under which different mechanisms can be compared. The following theorem (which is analogous to Theorem 1 in Bhandari and Secomandi [8]) relates the ordering of the seller's interim expected utilities for different mechanisms to the ordering of the value functions for those mechanisms (defined by Proposition 4.2).

Theorem 4.3. *Suppose that direct and feasible mechanisms j and k are defined on set $\mathcal{V}_1 \times \mathcal{V}_2 \equiv [0,1]^2$, and are such that the seller's interim expected utilities are ordered as $\bar{u}_1^j(v_1) \geq \bar{u}_1^k(v_1), \forall v_1 \in \mathcal{V}_1$. Then it holds that $V_t^j(y) \geq V_t^k(y), \forall y \in \mathcal{Y} \cup \{0\}$ and $\forall t \in \mathcal{T}$.*

Proof. We have $V_0^j(y) \equiv V_0^k(y) \equiv 0$, $\forall y \in \mathcal{Y} \cup \{0\}$. Thus the property trivially holds in stage 0. Make the induction hypothesis that $V_t^j(y) \geq V_t^k(y)$ for steps $1, ..., t-1$ and $\forall y \in \mathcal{Y}$. Consider stage t. Lemma 4.1 combined with the assumption on the interim expected utilities of mechanisms j and k implies

$$\begin{aligned} V_t^k(y) &= \beta V_{t-1}^k(y) + \lambda[\bar{x}_1^k(\beta\Delta V_{t-1}^k(y)) - \beta\Delta V_{t-1}^k(y)\bar{p}_1^k(\beta\Delta V_{t-1}^k(y))] \\ &\leq \beta V_{t-1}^k(y) + \lambda[\bar{x}_1^j(\beta\Delta V_{t-1}^k(y)) - \beta\Delta V_{t-1}^k(y)\bar{p}_1^j(\beta\Delta V_{t-1}^k(y))], \end{aligned}$$

which can be rearranged as

$$\lambda\bar{x}_1^j(\beta\Delta V_{t-1}^k(y)) \geq V_t^k(y) - \beta V_{t-1}^k(y) + \lambda\beta\Delta V_{t-1}^k(y)\bar{p}_1^j(\beta\Delta V_{t-1}^k(y)). \tag{4.7}$$

Using Lemma 4.1 and the feasibility of mechanism j, we obtain

$$\begin{aligned} V_t^j(y) &= \beta V_{t-1}^j(y) + \lambda[\bar{x}_1^j(\beta\Delta V_{t-1}^j(y)) - \beta\Delta V_{t-1}^j(y)\bar{p}_1^j(\beta\Delta V_{t-1}^j(y))] \\ &\geq \beta V_{t-1}^j(y) + \lambda[\bar{x}_1^j(\beta\Delta V_{t-1}^k(y)) - \beta\Delta V_{t-1}^j(y)\bar{p}_1^j(\beta\Delta V_{t-1}^k(y))], \end{aligned}$$

which rearranged is

$$\lambda\bar{x}_1^j(\beta\Delta V_{t-1}^k(y)) \leq V_t^j(y) - \beta V_{t-1}^j(y) + \lambda\beta\Delta V_{t-1}^j(y)\bar{p}_1^j(\beta\Delta V_{t-1}^k(y)). \tag{4.8}$$

Using inequalities (4.7) and (4.8) yields

$$V_t^j(y) - \beta V_{t-1}^j(y) + \lambda\beta\Delta V_{t-1}^j(y)\bar{p}_1^j(\beta\Delta V_{t-1}^k(y)) \geq V_t^k(y) - \beta V_{t-1}^k(y) + \lambda\beta\Delta V_{t-1}^k(y)\bar{p}_1^j(\beta\Delta V_{t-1}^k(y)).$$

This inequality can be rearranged as

$$\begin{aligned} V_t^j(y) - V_t^k(y) \;\; &\geq \;\; \beta[(1 - \lambda\bar{p}_1^j(\beta\Delta V_{t-1}^k(y)))(V_{t-1}^j(y) - V_{t-1}^k(y)) \\ &\qquad + \lambda\bar{p}_1^j(\beta\Delta V_{t-1}^k(y))(V_{t-1}^j(y-1) - V_{t-1}^k(y-1))] \\ &\geq \;\; 0, \end{aligned}$$

where the second inequality follows from the induction hypothesis and $\lambda\bar{p}_1^j(\beta\Delta V_{t-1}^k(y)) \in [0,1]$. By the principle of mathematical induction the property is true in all stages and states. $\square$

Now that we have a theorem to order the seller's value function under different mechanisms based on the ordering of the interim expected utilities of the seller under those mechanisms, we need to know the ordering of the mechanisms under consideration, summarized in Lemma 4.4.

Lemma 4.4. *(Bhandari and Secomandi [8] Lemma 1) For SUTP it holds that* $\bar{u}_1^{SPP}(v_1) \geq \bar{u}_1^{NBS}(v_1) \geq \bar{u}_1^{BPP}(v_1)$ *and* $\bar{u}_1^{STD}(v_1) \geq \bar{u}_1^{BPP}(v_1)$, $\forall\ v_1 \in \mathcal{V}_1$

As in the infinite horizon case, we can use Theorem 1 and Lemma 4.4 to establish value function comparisons between the four mechanisms under consideration (as done by Myerson [54]). These are the same comparisons established in the infinite horizon case in Proposition 2 of Bhandari and Secomandi [8].

Proposition 4.5. *Suppose that* $\mathcal{V}_i \equiv [0,1], \forall i \in \{1,2\}$ *and* $F_i(v_i) \equiv v_i, \forall i \in \{1,2\}$*. Then it holds that* $V_t^{SPP}(y) \geq V_t^{NBS}(y) \geq V_t^{BPP}(y)$ *and* $V_t^{STD}(y) \geq V_t^{BPP}(y), \forall y \in \mathcal{Y} \cup \{0\}$ *and* $\forall t \in \mathcal{T}$*.*

4.4 Assessing the Relevance of Modeling Private Information Under the STD and NBS Mechanisms

In our model the buyers and seller have private information about their respective marginal inventory valuations. In this private information setting, it is generally (that is, beyond the SUTP case) more challenging to obtain the STD and NBS mechanisms while it is simpler to derive the SPP and BPP mechanisms. However, these constraints may be important when calculating the seller's opportunity cost under the STD and NBS mechanisms. It is thus of interest to assess the relevance of modeling private information when computing the seller's opportunity costs under the STD and NBS mechanisms.

If the seller's and buyers' marginal inventory valuations were public knowledge, then the STD and NBS mechanisms would reduce to the Nash bargaining solution, a mechanism developed by Nash [56] to resolve a complete information, two-player, risk-neutral bargaining game. In the static case, a sale will occur under the Nash bargaining solution iff $v_1 \leq v_2$ at price $\frac{v_1+v_2}{2}$. Myerson [53] generalized the Nash bargaining solution to an incomplete information case to develop the NBS mechanism. Consequently, the NBS mechanism reduces to the Nash bargaining solution in the absence of private information. The STD obviously reduces to the Nash bargaining solution in the public information case (assuming an even split of the seller's and buyers' respective valuations).

For the SUTP, assuming no private information and that the Nash bargaining solution is used to model each negotiation, the resulting stochastic dynamic program for the seller is

$$\begin{aligned}V_t^{Nash}(y) = (1-\lambda)\beta V_{t-1}^{Nash}(y) \\ + \lambda \int_0^1 \Big\{ I\{v_2 \geq \beta \Delta V_{t-1}^{Nash}(y)\} \Big[\frac{v_2 + \beta \Delta V_{t-1}^{Nash}(y)}{2} + \beta V_{t-1}^{Nash}(y-1) \Big] \\ + I\{v_2 < \beta \Delta V_{t-1}^{Nash}(y)\} \beta V_{t-1}^{Nash}(y) \Big\} dv_2. \quad (4.9)\end{aligned}$$

where $I\{\mathcal{A}\}$ is an indicator function that equals 1 if $\mathcal{A}$ evaluates as true and 0 if $\mathcal{A}$ evaluates as false. Model (4.9) is simpler than Model (4.2) because of its simpler transactional setting.

We can use the opportunity cost $\beta \Delta V_{t-1}^{Nash}(y)$ based on the Nash bargaining solution model (4.9) in the presence of private information when using the STD and NBS mechanisms to obtain approximate value functions $U_t^{STD}(y)$ and $U_t^{NBS}(y)$ for these mechanisms as follows:

$$U_t^{STD}(y) = \beta U_{t-1}^{STD}(y) + \lambda[\bar{x}_1^{STD}(\beta \Delta V_{t-1}^{Nash}(y)) - \beta \Delta U_{t-1}^{STD}(y) \bar{p}_1^{STD}(\beta \Delta V_{t-1}^{Nash}(y))], \quad (4.10)$$

$$U_t^{NBS}(y) = \beta U_{t-1}^{NBS}(y) + \lambda[\bar{x}_1^{NBS}(\beta \Delta V_{t-1}^{Nash}(y)) - \beta \Delta U_{t-1}^{NBS}(y) \bar{p}_1^{NBS}(\beta \Delta V_{t-1}^{Nash}(y))]. \quad (4.11)$$

By comparing the value functions calculated for the SUTP under both the private and public information regimes, that is, under Model (4.9) specified with j equal to STD and NBS and Models (4.10) and (4.11), we can establish the importance of modeling private information when computing the seller's opportunity cost in the SUTP case. We demonstrate these results numerically in Section 4.5.2. This analysis might provide insights into the relevance of modeling private information beyond the SUTP case.

4.5 Numerical Results

In order to determine the significance of the comparison results in Proposition 4.5, in Subsection 4.5.1 we analyze numerical SUTP examples. Using a period length of one day, we consider a variety of parameter values: annual interest rate $r \in \{0.05, 0.1\}$, with discount factor $\beta = 1/(1+r/365)$; arrival probability $\lambda \in \{0.3, 0.6, 0.9\}$--this is the probability that a customer will arrive on a given day. In Subsection 4.5.2, we investigate the importance of modeling private information when computing the seller's opportunity cost under the STD and NBS mechanisms numerically within the SUTP. Our results are similar in all parameter combination cases, so only the $r = 0.05$ and $\lambda = 0.3$ case is discussed below. The corresponding graphs for the other parameter combinations are displayed in Appendix C.

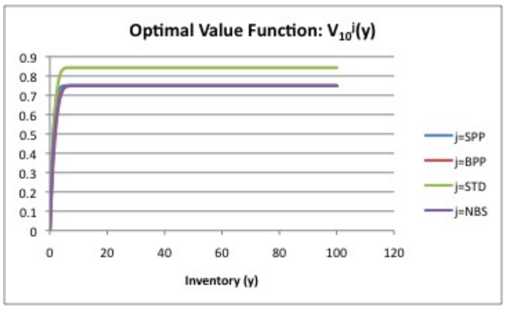
4.5.1 Optimal Value Function Comparison by Mechanism

Figure 4.1 shows the optimal value functions under the SPP, BPP, STD and NBS mechanisms at different inventory levels y and for different periods-to-go t. Each period is one day. The ordering of the mechanisms is compatible with Proposition 4.5. Specifically, for sufficiently high inventory levels and sufficiently low time-to-go, BPP, SPP and NBS perform similarly, while STD provides a higher optimal value function. At sufficiently low inventory levels with sufficiently high time-to-go, SPP outperforms STD which outperforms BPP. NBS performs similarly to BPP when the seller is in a weak position (high inventory, low time-to-go), and similarly to SPP when the seller is in a strong position (low inventory, high time-to-go). This is the expected result, given the design of the NBS mechanism.

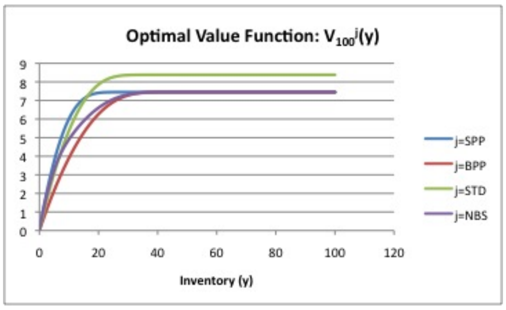
The comparisons in Proposition 4.5 are the same as the comparisons found in Bhandari and Secomandi [8]. However, in that paper, STD becomes a dominant mechanism only in extreme parameter value cases (such as $\lambda = 0.006$, $r = 0.05$ and inventory= 90--a situation where the seller has only 2.2 potential customers arriving each year, but 90 units of inventory available). Here, STD outperforms the other mechanisms for sufficiently high inventory-remaining to time-remaining ratio cases for even high λ values (see Appendix C for examples with inventory levels from 0 to 100 and $\lambda = 0.9$--a situation where the seller has as many as 328.5 potential customers each year).

For high values of t, these results converge to those found in Bhandari and Secomandi [8]. This is demonstrated in Figure 4.2 which shows the infinite horizon results from Bhandari and Secomandi next to the 5000 period-to-go results using our new model (both for the $\lambda = 0.3$, $r = 0.05$ case).

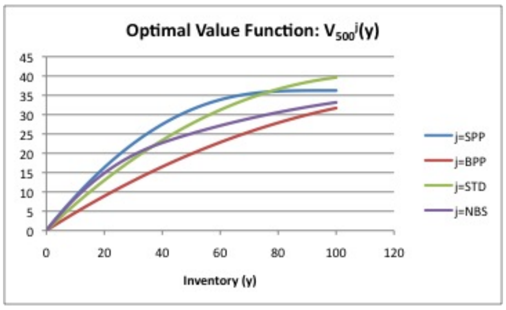
In order to better understand why STD outperforms SPP when the seller is weak (high relative inventory, low relative remaining time), we consider the quantities expected to be sold and the average price per unit expected to be received (calculated by dividing the optimal value function by the quantity expected to be sold for a given inventory level and time-to-go) for a given amount of inventory and a given remaining time-to-go. Figure 4.3 shows the expected quantity sold for different starting inventories and times-to-go under the four mechanisms. SPP results in the fewest sales, followed by STD, with BPP and NBS resulting in the most sales. The higher the remaining inventory, the higher the sales, until a “saturation point" is reached. Figure 4.4 shows the expected average price per unit expected to be received for different starting inventory levels and times-to-go



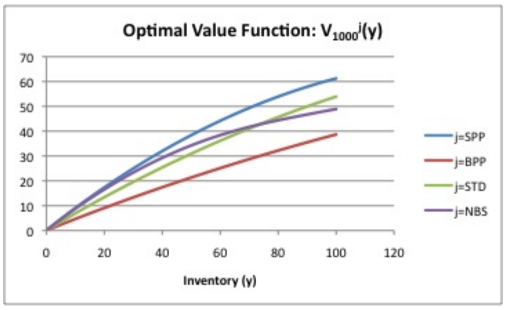
10 periods to go



100 periods to go

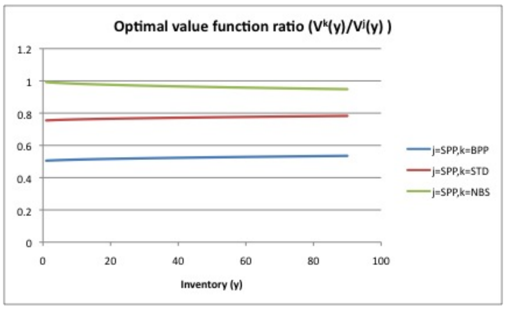


500 periods to go

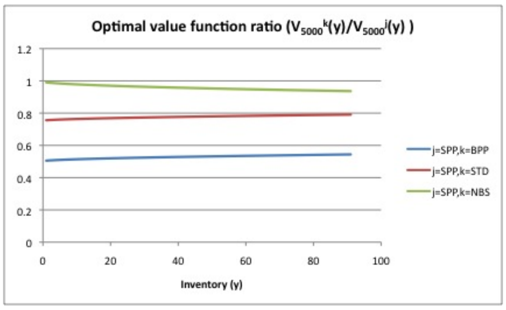


1000 periods to go

FIGURE 4.1: Optimal value function ratio $(V_t^k(y)/V_t^j(y))$ for different times-to-go, with $r = 0.05$ and $\lambda = 0.30$



Bhandari and Secomandi [8] (Figure 1)



Our model, with $t = 5000$

FIGURE 4.2: Optimal value function ratio $(V_t^k(y)/V_t^j(y))$, with $r = 0.05$ and $\lambda = 0.30$

under the four mechanisms. BPP results in the lowest average price, followed by STD, with SPP resulting in the highest price. The average price of NBS is close to that of SPP when the seller is strong, and close to that of BPP when the seller is weak (which is expected, given the design of the NBS mechanism).

If we consider the case of a weak seller (with high inventory and low time-to-go) we see that while SPP has a higher average price than STD, STD has a higher quantity sold than SPP. It appears that STD outperforms SPP for a weak seller due to quantity, rather than price, effects. By "splitting-the-difference" with the buyer, the seller receives a lower price, but has a higher likelihood of making a sale. When the seller is weak, it seems to

10 periods to go

100 periods to go

500 periods to go

1000 periods to go

FIGURE 4.3: Expected quantity sold under the four mechanisms for different times-to-go, with $r = 0.05$ and $\lambda = 0.30$

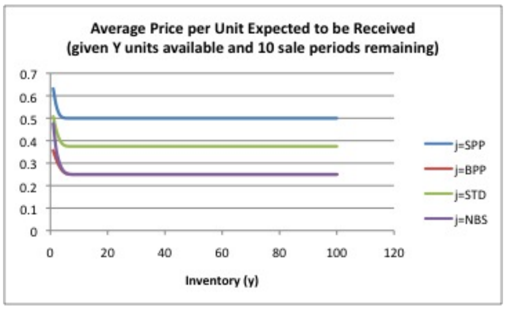
be better to be paid a smaller amount for some of the "excess" inventory, than nothing at all.

4.5.2 Effects of Modeling Private Information

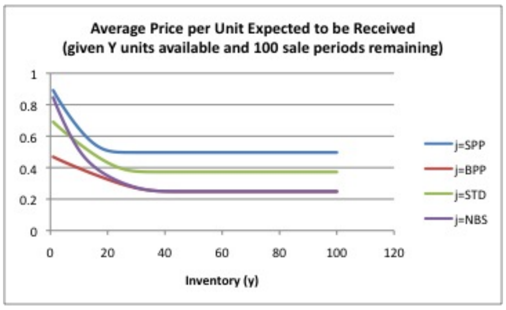
In this section, we quantify the impact of modeling private information when determining the seller's opportunity cost under the NBS and STD mechanisms.

Consider the NBS mechanism. Figure 4.5 shows the optimal value function from Model (4.2) with j equalt to NBS compared to the approximate value function from Model (4.11) at different inventory levels and times-to-go. We can see that the effect of modeling private information is rather small, a less than 5% difference for $t = 1000$ in all parameter cases studied.

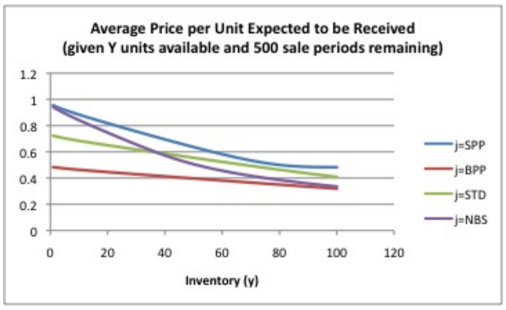
Focus on the STD mechanism. Figure 4.6 shows the optimal value function from Model (4.2) with j equal to STD compared to the approxiamte value function from Model (4.10) at different inventory levels and times-to-go. As an aside, we point out that the results for NBS and STD cannot be compared directly, due to the fact that their corresponding optimal value functions are different. We can see that the effect of modeling private information is small for high relative inventory levels, but for lower inventory levels and



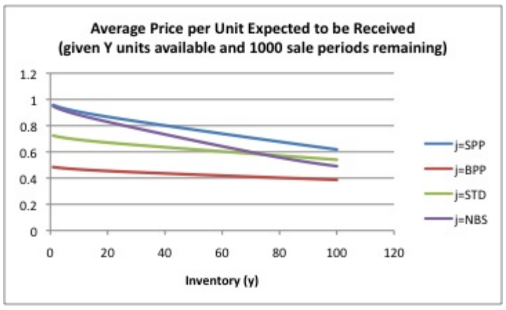
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE 4.4: Average price per unit expected to be received under the four mechanisms for different times-to-go, with $r = 0.05$ and $\lambda = 0.30$

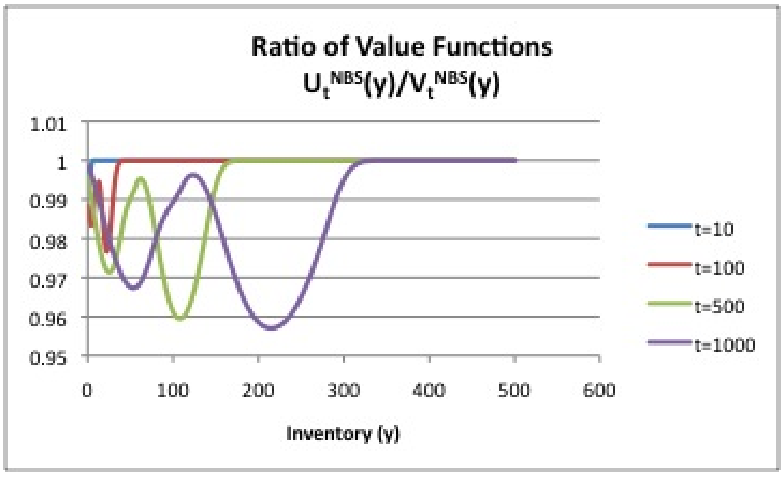


FIGURE 4.5: Ratio of approximate optimal value function to optimal value function under the NBS mechanism, with $r = 0.05$ and $\lambda = 0.30$

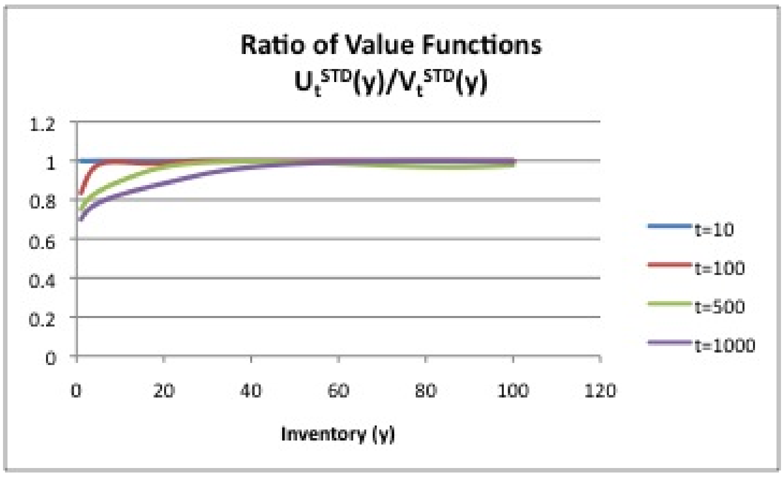


FIGURE 4.6: Ratio of approximate optimal value function to optimal value function under the STD mechanism, with $r = 0.05$ and $\lambda = 0.30$

higher times-to-go, this effect approaches a nearly 40% difference in value functions when $t = 1000$ (the results are similar in other parameter cases). In other words, it seems that the impact of modeling private information is more pronounced when the seller is strong.

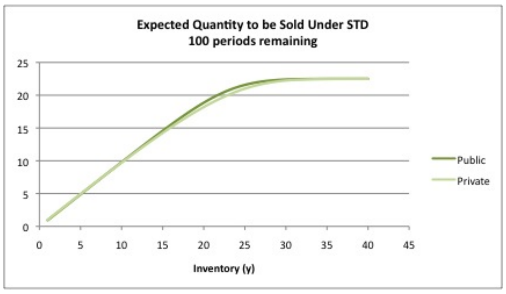
In order to develop some intuition for this finding, Figure 4.7 plots the expected quantity sold as well as the average price per unit expected to be received for Model (4.2) with j =STD and Model (4.10).

While the expected quantity sold is similar in both Models (4.3) and (4.10), the average price per unit expected to be received is higher in the Model (4.3) case. When the seller is strong, the seller's valuation for a unit of inventory is correspondingly high. In a private information setting, the seller can optimally increase his reported valuation above his true marginal valuation (this is what happens in equilibrium in the STD case; the STD mechanism considered here is the equivalent direct and feasible mechanism version of this equilibrium), leading to a higher average sales price, with less concern for lost sales due to his strong position.

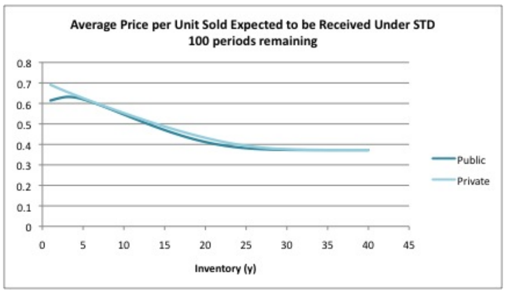
To summarize, modeling private information does not appear critical when using the NBS mechanism, or when a weak seller uses the STD mechanism. However, neglecting to model private information is not a reasonable approximation when considering a strong seller under the STD mechanism.

4.6 Conclusion

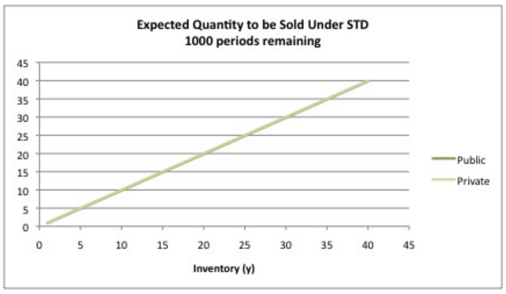
In this paper, we have extended the revenue management bargaining model developed by Bhandari and Secomandi [8] to a finite horizon setting, obtaining both structural



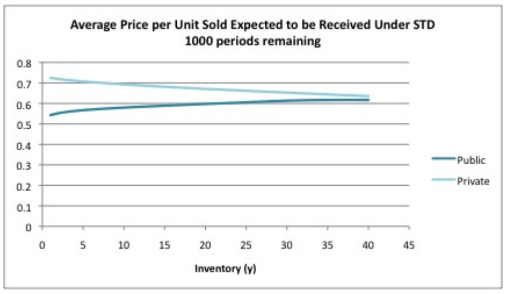
Expected quantity to be sold, 100 periods-to-go



Average price per unit expected to be received, 100 periods-to-go



Expected quantity to be sold, 1000 periods-to-go



Average price per unit expected to be received, 1000 periods-to-go

FIGURE 4.7: Expected quantities sold and average price per unit expected to be received under Model (4.2) specified for the STD mechanism and Model (4.10), for different times-to-go, with $r = 0.05$ and $\lambda = 0.30$

and numerical results. While the ordering of the SPP, BPP, STD and NBS pricing mechanisms remains consistent to the ordering in the infinite time horizon case, there are now far more (and more realistic) parameter regimes under which the STD mechanism is the most attractive option for the seller. Hypothetically, a hard deadline for sales (as in the finite horizon case) naturally emphasizes the benefits of a mechanism (such as STD) that can result in a higher probability of winning a sale before time runs out, rather than waiting for a higher price while the value of the inventory creeps towards a cliff.

We also considered the significance of modeling private information when determining the seller's opportunity cost under the NBS and STD mechanisms. While the results for the NBS mechanism are not sensitive to this modeling choice, modeling private information is critical for the STD mechanism when the seller is strong (when the seller is weak, they are again similar). As a consequence, easier to model public information constructions may serve as reasonable approximations for the NBS mechanism and the STD mechanism when the seller is weak. However, the use of such an approximation for the STD mechanism when the seller is strong would be misleading.

These results could be more robustly explored in further work examining the effects of a public deadline T, strategic buyers and non-stationary or unknown valuation distributions, as well as through numerical examples in situations beyond the SUTP, as noted in Secomandi and Bhandari [8].

Chapter 5

Conclusion

In this dissertation, we examined three different service operations problems. First, we empirically examined the provision of customer service on Twitter. Using an ordinal logistic regression model, we found that customers are less likely to experience a positive final sentiment as time passes. This finding may indicate a shift by the customer service team to harder to resolve cases as the program matures. Due to the noise in the data, future work could focus on better methods to reduce noise in data scraped from Twitter, such as better sentiment coding algorithms. Direct collaboration with a company during the data collection stage may be most rewarding, however. By obtaining direct customer satisfaction reports, as well as detailed customer and complaint information, noise would be reduced and additional important variables could be included in the analysis, leading to a better understanding of customer reactions to different Twitter-based customer service metrics.

Next, we partially extended Scheller-Wolf and Vesilo's [67] results for necessary and sufficient conditions for a finite r^{th} moment of expected delay in a FIFO multiserver queue, assuming a non-integral load and a service time distribution belonging to class $\mathcal{L}_1^{\beta}$, to the non-integral load case: we find a stricter necessary condition for a GI/GI/K-server system with integral $\rho = R$: the r^{th} moment of expected delay $E[D^r]$ will be infinite if $E[S^{1+(\frac{r}{\alpha(K-k)})}]$ is infinite, which occurs when the shape parameter of the service time distribution $\alpha < \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{r}{K-R}}$. Future work could include stricter necessary or sufficient conditions until the gap is closed, as well as further investigation of higher moment results. These results would provide further insight into the question of whether integral load systems in this class behave more like a system with slightly more or less work, or some combination thereof.

Finally, we ranked the value of four different bargaining mechanisms analytically and numerically in the context of the symmetric uniform trading problem, from the perspective of a seller of a finite inventory of perishable goods. While this ordering of the mechanisms remains the same as compared to the infinite horizon case studied in the literature, we find numerically, in an analogous model, that the relative value of the split-the-difference (STD) mechanism increases as we move to a situation where the seller faces a deadline to complete the sales. Additionally, we show that while using a simplified model that calculates the seller's opportunity cost using public information may be an acceptable approximation for the NBS mechanism, it produces substantially different results than the private information case when STD mechanism is used by a strong seller. Consequently, care must be taken when simplifying the model in this way.

Appendix A

Appendix A: Selecting a Model for Twitter-Based Customer Service Quality Metrics

A.1 Transitions

Table A.1: Transition Table - Model Selection Data

		FinalSentiment			
		Negative	Neutral	Positive	Total
Initial Sentiment	Negative	33	88	45	166
	Neutral	31	69	48	148
	Positive	3	11	6	20
	Total	67	168	99	334

Table A.2: Transition Table - Test Data

		FinalSentiment			
		Negative	Neutral	Positive	Total
Initial Sentiment	Negative	44	83	54	181
	Neutral	27	88	59	174
	Positive	3	7	7	17
	Total	74	178	120	372

A.2 Descriptives

TABLE A.3: Full Data Set - Descriptives

	N	Minimum	Maximum	Mean	Std. Dev.
Ratio of company to total messages	706	.01	.71	.3321	.09899
Company response time	706	.00	6.99	1.0651	1.43519
Customer response time	706	.00	14.00	.4699	1.24833
Number of company related messages in network	706	.00	147.00	5.0722	11.81532
Ratio of positive to total network messages	706	.00	1.00	.1115	.22997
Ratio of positive to total network messages squared	706	.00	1.00	.0652	.20284
Ratio of negative to total network messages	706	.00	1.00	.0897	.19468
LN(Date of initial message)	706	.00	5.71	4.5326	.97564

A.3 Variables Considered for Model Inclusion

A.3.1 Customer Attribute - Initial Mention

This variable is 1 if the customer's initial message included "@company" or "@companysupport," and 0 otherwise.

A.3.2 Customer Attribute - Weekend

This variable is 1 if the customer's initial message occured during a weekend and 0 otherwise.

A.3.3 Customer Attribute - Business Hours

This variable is 1 if the customer's initial message occured during business hours and 0 otherwise.

A.3.4 Customer Attribute - Initial Sentiment

This variable indicates the sentiment (positive, negative or neutral) of the customer's initial message.

A.3.5 Customer Attribute - Number of customer messages

This variable measures the total number of messages sent by the customer in a case.

A.3.6 Service Attribute - Number of company messages

This variable measures the total number of messages sent by the company's company support team in a case.

A.3.7 Service Attribute - Company response time

This variable is the average time between a customer message and the company's response in a case.

A.3.8 Service/Customer Attribute - Average Time

This variable is calculated by dividing elapsed time by the total number of messages in a case.

A.3.9 Service/Customer Attribute - Ratio of company to total messages

This variable divides the number of company messages by the total number of messages in a case (that is, the number of company messages plus the number of customer messages in a case).

A.3.10 Service/Customer Attribute - Ratio of customer to company messages

This variable divides the number of customer messages by the number of company messages.

A.3.11 Service/Customer Attribute - Number of messages

This variable measures the total number of messages sent by the company and the customer in a case.

A.3.12 Service Attribute - Priority

This variable is the priority (high-medium-low) assigned to a case by the company.

A.3.13 Service Attribute - Elapsed Time

This variable measures the time elapsed between the customer's first message and the company's final message in a case.

A.3.14 Service Attribute - Initial Response Time

This variable measures the time elapsed between the customer's first message and the company's first message in a case.

A.3.15 Service Attribute - Date of initial message

This variable indicates the date of the customer's first message, where "1" indicates the first day in the data set (February 22). This value is subsequently incremented (e.g. Februrary 25 is "4").

A.3.16 Customer Attribute - Customer response time

This variable is the average time between a company message and the customer's response in a case.

A.3.17 Customer Attribute - Number of followers

This variable is the number of Twitter followers a customer had at the time of the server interaction.

A.3.18 Customer Attribute - Number of friends

This variable is the number of Twitter friends a customer had at the time of inquiry (after the server interaction).

A.3.19 Customer Network Attributes - Number of positive network messages from friend-followers

This variable is the number of messages involving the company or its products with positive sentiment sent in the week prior to the customer's first message by those friends who are also followers.

A.3.20 Customer Network Attributes - Number of neutral network messages from friend-followers

This variable is the number of messages involving the company or its products with neutral sentiment sent during the case, as well as in the week prior to the customer's first message by those friends who are also followers.

A.3.21 Customer Network Attributes - Number of negative network messages from friend-followers

This variable is the number of messages involving the company or its products with negative sentiment sent during the case, as well as in the week prior to the customer's first message by those friends who are also followers.

A.3.22 Customer Network Attributes - Number of positive network messages from friend-but-not-followers

This variable is the number of messages involving the company or its products with positive sentiment sent during the case, as well as in the week prior to the customer's first message by those friends who are not also followers.

A.3.23 Customer Network Attributes - Number of neutral network messages from friend-but-not-followers

This variable is the number of messages involving the company or its products with neutral sentiment sent during the case, as well as in the week prior to the customer's first message by those friends who are not also followers.

A.3.24 Customer Network Attributes - Number of negative network messages from friend-but-not-followers

This variable is the number of messages involving the company or its products with negative sentiment sent during the case, as well as in the week prior to the customer's first message by those friends who are not also followers.

A.3.25 Customer Network Attributes - Number of positive network messages

This variable is the total number of messages involving the company or its products with positive sentiment sent during the case, as well as in the week prior to the customer's first message by the customer's friends.

A.3.26 Customer Network Attributes - Number of neutral network messages

This variable is the total number of messages involving the company or its products with neutral sentiment sent during the case, as well as in the week prior to the customer's first message by the customer's friends.

A.3.27 Customer Network Attributes - Number of negative network messages

This variable is the total number of messages involving the company or its products with negative sentiment sent during the case, as well as in the week prior to the customer's first message by the customer's friends.

A.3.28 Network Attribute - Number of company related messages in network

This variable is the total number of positive, negative, neutral and indeterminate messages involving the company or its products sent during the case, as well as in the week prior to the customer's first message, by the customer's friends.

A.3.29 Network Attribute - Ratio of positive to total network messages

This variable is the number of positive messages involving the company or its products sent during the case, as well as in the week prior to the customer's first message, by the customer's friends, divided by the number of company related messages in network variable. In the event that the number of company related messages in network variable had a value of 0, this variable was also coded as 0.

A.3.30 Network Attribute - Ratio of negative to total network messages

This variable is the number of negative messages involving the company or its products sent during the case, as well as in the week prior to the customer's first message, by the customer's friends, divided by the number of company related messages in network variable. In the event that the number of company related messages in network variable had a value of 0, this variable was also coded as 0.

A.3.31 Customer Network Attribute - Number of promotional messages

This variable is the number of messages sent during the case, as well as in the week prior to the customer's first message through one of the company's non-support Twitter accounts.

A.3.32 Variable notes

Some of these variables are redundant. Additionally, quadratic, square root and natural log transformations of these variables were also considered, where appropriate. In the case where the variable may have a value of 0 (number of friends, number of followers, number of a certain type of network message, etc.), the natural log of the value of the variable plus one was taken.

A.4 Additional Results

Using the chosen model on the entire data set, we find the results given in table A.4. It is important to note that this test uses some data that was used to chose the variables in the model, so significance is overstated and unreliable. Overall model significance is 0.001. The test of parallel lines resulted in a significance of .776, indicating the the proportional odds assumption is not rejected, so ordinal regression remains appropriate. In contrast to the test data only results, the parameters for the ratio of company to total messages and the ratio of positive to total network messages are significant, so we can now address hypotheses 1 and 2. The parameter for the natural log of the date of the initial message is also significant, as it was before, so the comments in 2.5.1 still apply. As before, the service time variables are insignificant, so the comments in section 2.5.2 still apply.

Table A.4: Chosen model - full data set

	Parameter Value	Std. Error	Sig.
$\alpha_{neutral}$	-2.202	.427	.000
$\alpha_{positive}$	.052	.418	.900
β Ratio of company to total messages	1.673	.728	.022
β Company response time	.039	.051	.444
β Customer response time	-.091	.058	.116
β Number of company related messages in network	.000	.007	.986
β Ratio of positive to total network messages	1.944	.988	.049
β Ratio of positive to total network messages squared	-1.908	1.110	.086
β Ratio of negative to total network messages	-.281	.375	.454
β LN(Date of initial message)	-.308	.077	.000

A.4.1 Evaluation of Hypothesis 1

We find that the ratio of company to total messages has a parameter value of 1.673 with significance of 0.022 when the model was applied to the full data set. This result confirms our hypothesis that a higher ratio of company to total case messages increases the probability of a more positive case resolution (holding the date, company response time, customer response time, number of network messages and the percentage of positive and negative network messages constant). As discussed before, this may indicate that a higher number of messages provided for the same amount of customer information provided implies higher service level which results in higher customer satisfaction.

A.4.2 Evaluation of Hypothesis 2

We find that the ratio of positive to total network messages has a parameter value of 1.944 with significance of 0.049 when the model was applied to the full data set. However, the quadratic term is not significant. While this result does not confirm the quadratic nature of our hypothesis, it does correspond with the idea that a higher ratio of positive to total network messages in the period between one week prior to the customer's first message and the customer's last message increases the probability of a more positive case resolution (holding the date, ratio of company to total messages, company response time, customer response time, number of network messages and the percentage of negative network messages constant). As discussed before, this may related to the findings of Ma et al [48] who found that positive sentiment expression in a customer's network led to more positive sentiment expression if the customer was already in a positive state, and more negative sentiment expression if the customer was already in a negative state.

A.5 Extra References

'ordinal' by Rune Haubo B Christensen

'glmulti' by Vincent Calcagno

'vgam' by Thomas Yee

'reshape2' by Hadley Wickham

Wikipedia "AIC"

http://www.ats.ucla.edu/stat/spss/output/ologit.htm

http://www.ats.ucla.edu/stat/spss/dae/ologit.htm

http://www.ats.ucla.edu/stat/r/dae/ologit.htm

http://research.cs.tamu.edu/prism/lectures/iss/iss_l13.pdf

http://www.stanford.edu/ hastie/Papers/ESLII.pdf

Appendix B

Appendix B: Necessary Condition for Finite Delay Moments for FIFO GI/GI/K Queues with Integral Load

B.1 Proof of Lemma 3.4

Lemma. Let $Y_i = S_i - T$, where T is deterministic and equal to E[S], and S has $P(X > u) \sim u^{-\alpha}$ with $1 < \alpha < 2$. Then,

$$P(\frac{\Sigma Y_i}{n^{\frac{1}{\alpha}}} > \epsilon) \sim U_\alpha(\epsilon) \tag{B.1}$$

where U is the stable distribution determined (including centering) by the characteristic function

$$\psi(\zeta) = |\zeta|^\alpha C \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)}[cos\frac{\pi\alpha}{2} \mp i(p-q)sin\frac{\pi\alpha}{2}]. \tag{B.2}$$

Proof.

Definition B.1. Feller [24], page 172

"F belongs to the domain of attraction of U iff there exist constants $a_n > 0$ and b_n such that the distribution of $a_n^{-1} S_n - nb_n$ tends to U."

Let φ be the characteristic function of F, and ω be the characteristic function of U. a_n and b_n are scaling parameters.

In other words, Lemma 3.4 states that our netput process Y belongs to a domain of attraction, allowing us to characterize the behavior of the accumulation of work in a system with a load exactly equal to the number of servers available.

Feller [24] provides the two lemmas needed to prove Lemma 3.4: the first (Lemma B.2) presents the requirements needed for a distribution to belong to a domain of attraction, the second (Lemma B.3) presents the necessary scaling parameters a_n and b_n to ensure the domain of attraction to be that described in equation B.2. After stating these lemmas, we show that the the sum of service times S will belong to the domain of attraction of U when scaling parameters $a_n = n^{\frac{1}{\alpha}}$ and $b_n = T = E[S]$ are used.

Lemma B.2. *Theorem 2 from Feller [24] section XVII.5*

"(a) In order that a distribution F belong to some domain of attraction it is necessary that the truncated moment function μ varies regularly with an exponent 2-α $(0 < \alpha \leq 2)$.

(b) If $\alpha = 2$, this condition is also sufficient provided F is not concentrated at one point.

(c) If $\mu(x) \sim x^{2-\alpha}L(x), x \to \infty$ holds with $0 < \alpha \leq 2$ then F belongs to some domain of attraction iff the tails are balanced so that as $x \to \infty$

$$\frac{1 - F(x)}{1 - F(x) + F(-x)} \to p \tag{B.3}$$

$$\frac{F(-x)}{1 - F(x) + F(-x)} \to q \tag{B.4}$$

where L varies slowly."

Lemma B.3. *Theorem 3 from Feller [24] section XVII.5*

"Let U be the stable distribution determined (including centering) by the characteristic function

$$\psi(\zeta) = |\zeta|^{\alpha} C \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)}[cos\frac{\pi\alpha}{2} \mp i(p-q)sin\frac{\pi\alpha}{2}] \tag{B.5}$$

if $\alpha \neq 1$ or

$$\psi(\zeta) = -|\zeta| \cdot C[\frac{1}{2}\pi \pm i(p-q)log|\zeta|] \tag{B.6}$$

if $\alpha = 1$.

Let the distribution F satisfy the conditions of [Lemma B.2], and let a_n satisfy

$$\frac{n}{a_n^2}\mu(a_n) \to C$$

.

(i) If $0 < \alpha < 1$ *then* $\varphi^n(\zeta/a_n) \to \omega(\zeta) = e^{\psi(\zeta)}$.

(ii) If $1 < \alpha \leq 2$ *and* $\mu(\infty) = \infty$ *the same is true provided* F *is centered to zero expectation.*

(iii) If $\alpha = 1$ *then*

$$(\varphi(\zeta/a_n)e^{-ib_n\zeta})^n \to \omega(\zeta) = e^{\psi(\zeta)},$$

where

$$b_n = \int_{-\infty}^{+\infty} sin\frac{x}{a_n}F\{\mathrm{d}x\}.''$$

To summarize the application of Lemmas B.2 and B.3, Lemma 3.4 will be true if the following four conditions are met:

1) the truncated moment function of S varies regularly with an exponent $2 - \alpha$, with $0 < \alpha \leq 2$

2) the tails of S are balanced so that as $x \to \infty$

$$\frac{1 - F(x)}{1 - F(x) + F(-x)} \to p$$

$$\frac{F(-x)}{1 - F(x) + F(-x)} \to q$$

3) a_n satisfies

$$\frac{n}{a_n^2}\mu(a_n) \to C$$

4) S is centered to 0 expectation.

We will demonstrate point-by-point that these conditions are met.

1) the truncated moment function of S varies regularly with an exponent $2 - \alpha$, with $0 < \alpha \leq 2$:

The truncated moment function is defined as $\mu(x) = \int_{-x}^{x} y^2 f(y)dy$. For S with $P(X > u) \sim u^{-\alpha}$ and $1 < \alpha < 2$, we find $\mu(x) = \int_0^x y^2 \frac{\alpha}{y^{\alpha+1}}dy \sim x^{2-\alpha}$, so this condition is fulfilled.

2) the tails of S are balanced so that as $x \to \infty$

$$\frac{1 - F(x)}{1 - F(x) + F(-x)} \to p$$

$$\frac{F(-x)}{1-F(x)+F(-x)} \to q$$

We have S with $P(X > u) \sim u^{-\alpha}$, with $F(-x) = 0$:

$$\frac{1-F(x)}{1-F(x)+F(-x)} = \frac{x^{-\alpha}}{x^{-\alpha}+0} \to 1$$

$$\frac{0}{x^{-\alpha}+0} \to 0$$

,

so this condition is fulfilled.

3) a_n satisfies

$$\frac{n}{a_n^2}\mu(a_n) \to C$$

Equation B.1 implies that $a_n = n^{\frac{1}{\alpha}}$. With $\mu(a_n) \sim a_n^{2-\alpha}$, we see

$$\frac{n}{a_n^2}\mu(a_n) \sim \frac{n}{(n^{\frac{1}{\alpha}})^2}(n^{\frac{1}{\alpha}})^{2-\alpha} = 1 \to C$$

4) S is centered to 0 in expectation, which is fulfilled by using $b_n = T = E[S]$ □

B.2 Asymptotic convergence of bounds

Here, we consider the limit of the ratio of the previous bounds to the new bounds for the GI/GI/K case as the number of servers K approaches infinity.

Lemma B.4. *The newly established necessary conditions for finite mean delay with integral load for FIFO GI/GI/K queues, with service times both belonging to class* $\mathcal{L}^{\alpha+1}$ *with* $dF(x) \sim x^{-\alpha}$ *approach the previously established lower bounds of* $\frac{K-R+1+r}{K-R+1}$ *as the number of servers K approaches infinity.*

Proof.

$$\begin{aligned}
\lim_{K\to\infty} \frac{\frac{K-R+1+r}{K-R+1}}{\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{r}{K-R}}} &= \lim_{K\to\infty} \frac{\frac{K-R+1+r}{K-R+1}}{\frac{1}{2}+\sqrt{\frac{K-R+4r}{4(K-R)}}} \\
&= \lim_{K\to\infty} \frac{\frac{K-R+1+r}{K-R+1}}{\frac{1}{2}+\frac{\sqrt{K-R+4r}}{\sqrt{4(K-R)}}} \\
&= \lim_{K\to\infty} \frac{\frac{K-R+1+r}{K-R+1}}{\frac{1}{2}+\frac{\sqrt{K-R+4r}}{2\sqrt{K-R}}} \\
&= \lim_{K\to\infty} \frac{2(\frac{K-R+1+r}{K-R+1})}{1+\frac{\sqrt{K-R+4r}}{\sqrt{K-R}}} \\
&= \frac{\lim_{K\to\infty} 2(\frac{K-R+1+r}{K-R+1})}{\lim_{K\to\infty} 1+\sqrt{\lim_{K\to\infty}\frac{K-R+4r}{K-R}}} \\
&= \frac{\lim_{K\to\infty} 2(\frac{1}{1})}{\lim_{K\to\infty} 1+\sqrt{\lim_{K\to\infty}\frac{1}{1}}} = \frac{2}{2} \\
&= 1
\end{aligned}$$

L'Hôpital's Rule is used for the penultimate step. As the number of servers approaches infinity, the ratio of the value of the new bounds to the value of the old bounds approaches 1. □

B.3 Workloads are greater than $C'x^{\frac{1}{\alpha}}$

The workload at server R will be greater than $C'x^{\frac{1}{\alpha}}$, given that we have $C > \left(\frac{R^{R-1}-1}{\epsilon R^{R-2}(R-1)}\right)^{\alpha}(2^{R+1}RE[S])$.

Lemma B.5. *$\frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n} > C'x^{\frac{1}{\alpha}}$ for $R \geq 2$ where C' is a positive constant, if $C > \left(\frac{R^{R-1}-1}{\epsilon R^{R-2}(R-1)}\right)^{\alpha}(2^{R+1}RE[S])$.*

Proof. We will solve backwards to find the values of C that will result in $\frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n} > C'x^{\frac{1}{\alpha}}$.

$$\frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n} > C'x^{\frac{1}{\alpha}}$$
$$\left(\frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]})^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{1}{R^n}\right)x^{\frac{1}{\alpha}} > C'x^{\frac{1}{\alpha}}$$

So $\frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n} > C'x^{\frac{1}{\alpha}}$ if $\left(\frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]})^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{1}{R^n}\right) > 0$

$$\frac{1}{R}\epsilon(\frac{C/2^R}{2RE[S]})^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{1}{R^n} > 0$$

$$\frac{1}{R}\epsilon(\frac{C/2^R}{2RE[S]})^{\frac{1}{\alpha}} > \sum_{n=1}^{R-1}\frac{1}{R^n}$$

$$C^{\frac{1}{\alpha}}\frac{\epsilon}{R}(\frac{1}{2^R 2RE[S]})^{\frac{1}{\alpha}} > \sum_{n=1}^{R-1}\frac{1}{R^n}$$

Recognizing the summation term as a geometric series

$$C^{\frac{1}{\alpha}}\frac{\epsilon}{R}(\frac{1}{2^R 2RE[S]})^{\frac{1}{\alpha}} > \sum_{n=0}^{R-1}\frac{1}{R^n} - 1$$

$$C^{\frac{1}{\alpha}}\frac{\epsilon}{R}(\frac{1}{2^{R+1}RE[S]})^{\frac{1}{\alpha}} > \frac{1-(1/R)^R}{1-(1/R)} - 1$$

$$C^{\frac{1}{\alpha}}\frac{\epsilon}{R}(\frac{1}{2^{R+1}RE[S]})^{\frac{1}{\alpha}} > \frac{R^{R-1}-1}{R^{R-1}(R-1)}$$

$$C^{\frac{1}{\alpha}} > \frac{R}{\epsilon}(2^{R+1}RE[S])^{\frac{1}{\alpha}}\frac{R^{R-1}-1}{R^{R-1}(R-1)}$$

$$C^{\frac{1}{\alpha}} > \frac{R(R^{R-1}-1)(2^{R+1}RE[S])^{\frac{1}{\alpha}}}{\epsilon R^{R-1}(R-1)}$$

$$C > \left(\frac{R^{R-1}-1}{\epsilon R^{R-2}(R-1)}\right)^{\alpha}(2^{R+1}RE[S])$$

So $\frac{1}{R}\epsilon(\frac{\Upsilon}{2RE[S]}x)^{\frac{1}{\alpha}} - \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n} > C'x^{\frac{1}{\alpha}}$ when $C > \left(\frac{R^{R-1}-1}{\epsilon R^{R-2}(R-1)}\right)^{\alpha}(2^{R+1}RE[S])$. □

At time $t = \frac{\Upsilon}{2}x + \sum_{n=1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n}$, the workload at each server i in the group of servers 1 through R-1 will have a workload equal to $\frac{R-i}{i}\frac{x^{\frac{1}{\alpha}}}{R^{R-i}} - \sum_{n=R-i+1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n}$.

Lemma B.6. $\frac{R-i}{i}\frac{x^{\frac{1}{\alpha}}}{R^{R-i}} - \sum_{n=R-i+1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n} > C'x^{\frac{1}{\alpha}}$ *where C' is a positive constant, for $1 \le i \le R-1$ and $R \ge 2$.*

Proof.

$$\frac{R-i}{i}\frac{x^{\frac{1}{\alpha}}}{R^{R-i}} - \sum_{n=R-i+1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n} = x^{\frac{1}{\alpha}}[\frac{R-i}{i}\frac{1}{R^{R-i}} - \sum_{n=R-i+1}^{R-1}\frac{1}{R^n}]$$

$$= x^{\frac{1}{\alpha}}[\frac{R-i}{i}\frac{1}{R^{R-i}} - (\sum_{n=0}^{R-2}\frac{1}{R}(\frac{1}{R})^n - \sum_{n=0}^{R-i-1}\frac{1}{R}(\frac{1}{R})^n)]$$

If $[\frac{R-i}{i}\frac{1}{R^{R-i}} - (\sum_{n=0}^{R-2}\frac{1}{R}(\frac{1}{R})^n - \sum_{n=0}^{R-i-1}\frac{1}{R}(\frac{1}{R})^n)] > 0$, then $\frac{R-i}{i}\frac{x^{\frac{1}{\alpha}}}{R^{R-i}} - \sum_{n=R-i+1}^{R-1}\frac{x^{\frac{1}{\alpha}}}{R^n} > C'x^{\frac{1}{\alpha}}$. Now we can recognize the terms $\sum_{n=0}^{R-2}\frac{1}{R}(\frac{1}{R})^n$ and $\sum_{n=0}^{R-i-1}\frac{1}{R}(\frac{1}{R})^n$ as geometric series.

$$
\begin{aligned}
x^{\frac{1}{\alpha}}[\frac{R-i}{i}\frac{1}{R^{R-i}} - (\sum_{n=0}^{R-2}\frac{1}{R}(\frac{1}{R})^n - \sum_{n=0}^{R-i-1}\frac{1}{R}(\frac{1}{R})^n)]&\\
&= x^{\frac{1}{\alpha}}[\frac{R-i}{i}\frac{1}{R^{R-i}} - (\frac{1}{R}\frac{1-(\frac{1}{R})^{R-1}}{1-\frac{1}{R}} - \frac{1}{R}\frac{1-(\frac{1}{R})^{R-i}}{1-\frac{1}{R}})]\\
&= x^{\frac{1}{\alpha}}[\frac{R-i}{iR^{R-i}} - \frac{1-(\frac{1}{R})^{R-1}}{R-1} + \frac{1-(\frac{1}{R})^{R-i}}{R-1}]\\
&= x^{\frac{1}{\alpha}}[\frac{R-i}{iR^{R-i}} + \frac{1-(\frac{1}{R})^{R-i}-1+(\frac{1}{R})^{R-1}}{R-1}]\\
&= x^{\frac{1}{\alpha}}[\frac{R-i}{iR^{R-i}} + \frac{(\frac{1}{R})^{R-1}-(\frac{1}{R})^{R-i}}{R-1}]\\
&= x^{\frac{1}{\alpha}}[\frac{(R-i)(R-1)+(\frac{1}{R})^{R-1}iR^{R-i}-(\frac{1}{R})^{R-i}iR^{R-i}}{iR^{R-i}(R-1)}]\\
&= \frac{x^{\frac{1}{\alpha}}}{iR^{R-i}(R-1)}[(R-i)(R-1)+iR^{1-i}-i]
\end{aligned}
$$

The smallest possible value for $(R-i)(R-1)$ given $1 \leq i \leq R-1$ occurs when $i = R-1$. Similarly, the largest possible value for i occurs when $i = R-1$. We can incorporate this information into an inequality:

$$
\begin{aligned}
\frac{x^{\frac{1}{\alpha}}}{iR^{R-i}(R-1)}[(R-i)(R-1)+iR^{1-i}-i] &> \frac{x^{\frac{1}{\alpha}}}{iR^{R-i}(R-1)}[(R-1)+iR^{1-i}-(R-1)]\\
&= \frac{x^{\frac{1}{\alpha}}}{iR^{R-i}(R-1)}[iR^{1-i}]\\
&> C'x^{\frac{1}{\alpha}}
\end{aligned}
$$

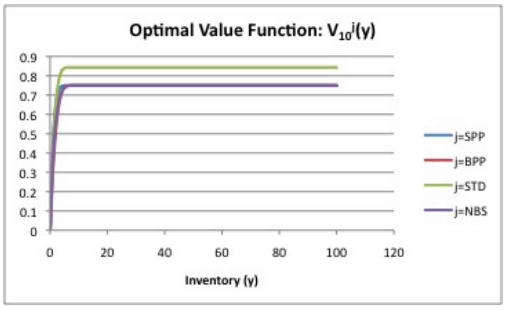
because $\frac{1}{iR^{R-i}(R-1)} > 0$ and $iR^{1-i} > 0$, the entire expression is greater than $C'x^{\frac{1}{\alpha}}$. □

Appendix C

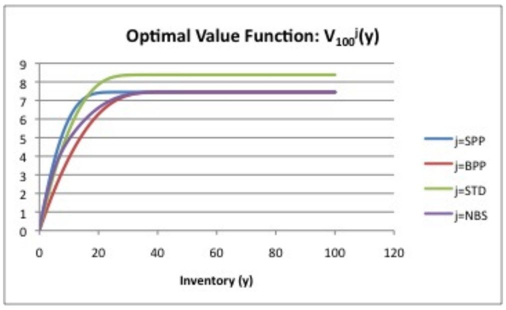
Appendix: Revenue Management with Bargaining and a Finite Horizon - Additional Numerical Results

In this Appendix, we summarize numerical results for parameter value combinations other than $\lambda = 0.3, r = 0.05$. Results for $\lambda = 0.6, r = 0.05$ are displayed in Section C.1 as Figures C.1 to C.6. Results for $\lambda = 0.9, r = 0.05$ are displayed in Section C.2 as Figures C.7 to C.12. Results for $\lambda = 0.3, r = 0.10$ are displayed in Section C.3 as Figures C.13 to C.18. Results for $\lambda = 0.6, r = 0.10$ are displayed in Section C.4 as Figures C.19 to C.24. Finally, results for $\lambda = 0.9, r = 0.10$ are displayed in Section C.5 as Figures C.25 to C.30.

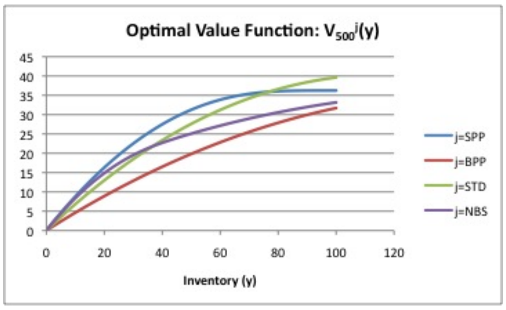
C.1 Results for $\lambda = 0.6, r = 0.05$ (Figures C.1 to C.6)



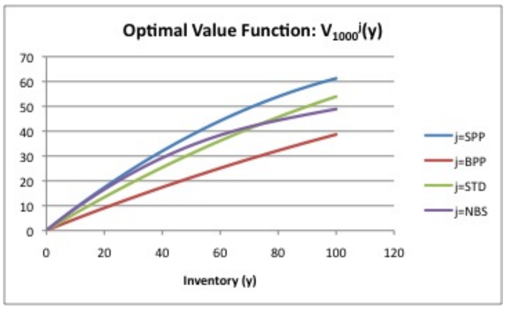
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.1: Optimal value function ratio $(V_t^k(y)/V_t^j(y))$ for different times-to-go, with $r = 0.05$ and $\lambda = 0.60$

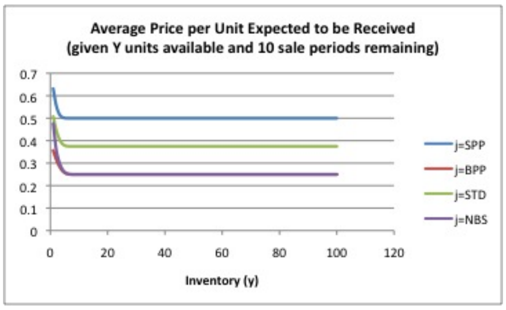
10 periods to go

100 periods to go

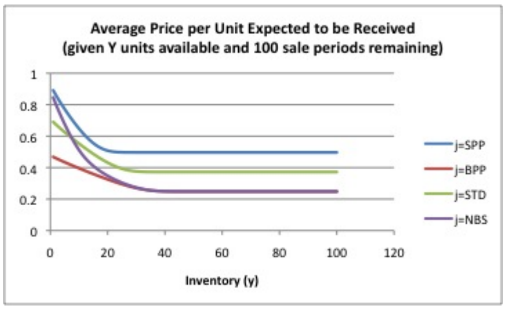
500 periods to go

1000 periods to go

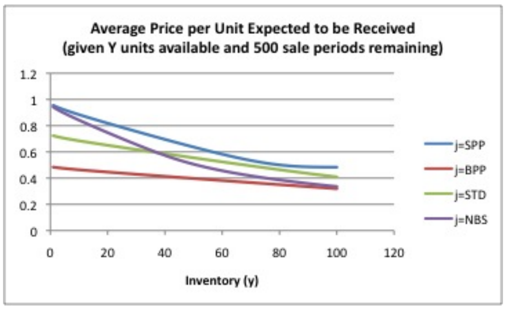
FIGURE C.2: Expected quantity sold under the four mechanisms for different times-to-go, with $r = 0.05$ and $\lambda = 0.60$



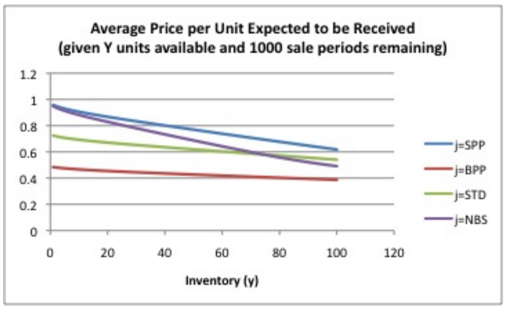
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.3: Average price per unit expected to be received under the four mechanisms for different times-to-go, with $r = 0.05$ and $\lambda = 0.60$

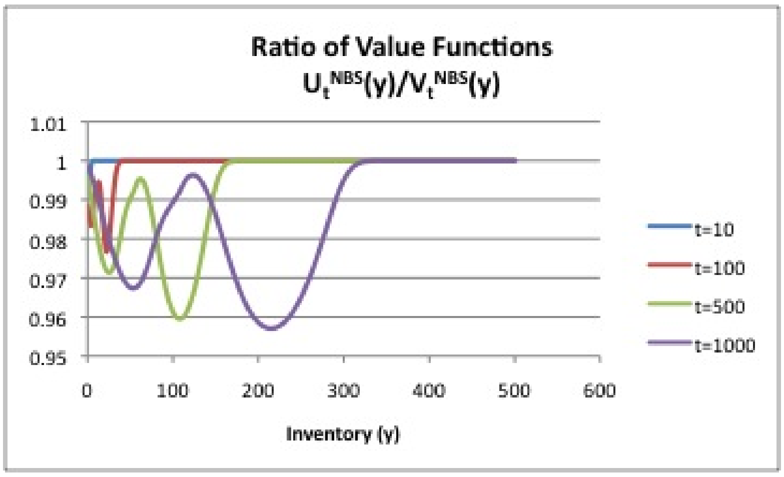


Figure C.4: Ratio of approximate optimal value function to optimal value function under the NBS mechanism, with $r = 0.05$ and $\lambda = 0.60$

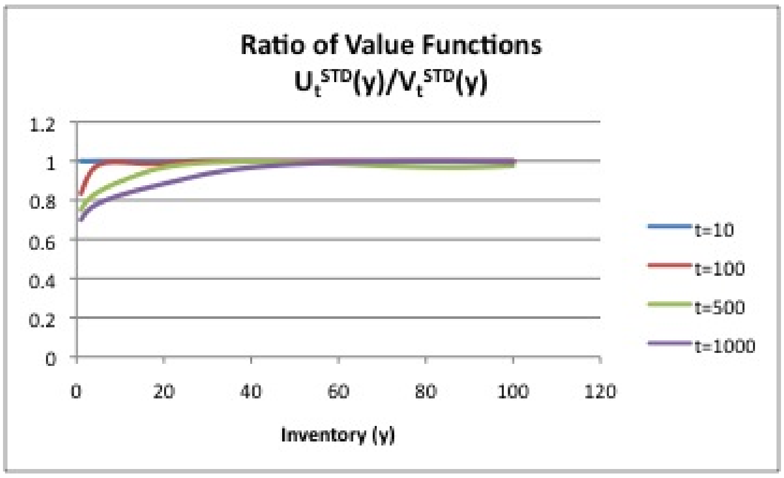
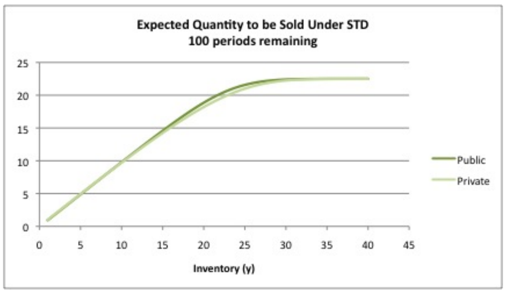
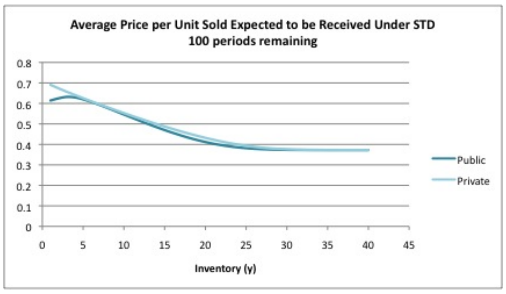


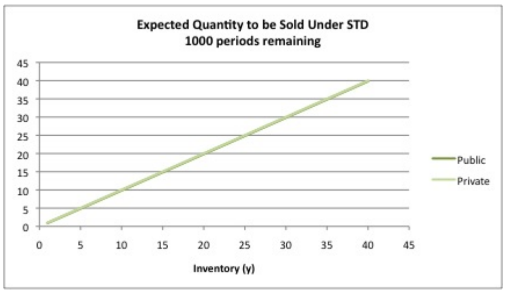
Figure C.5: Ratio of approximate optimal value function to optimal value function under the STD mechanism, with $r = 0.05$ and $\lambda = 0.60$



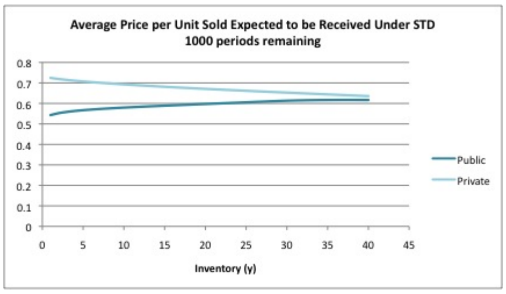
Expected quantity to be sold, 100 periods-to-go



Average price per unit expected to be received, 100 periods-to-go



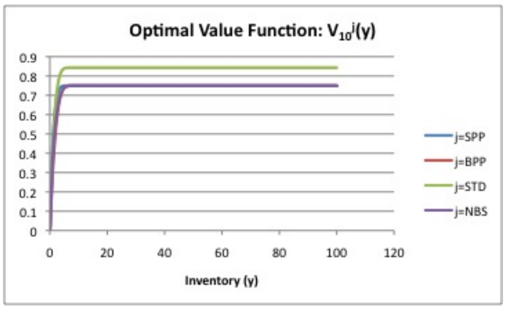
Expected quantity to be sold, 1000 periods-to-go



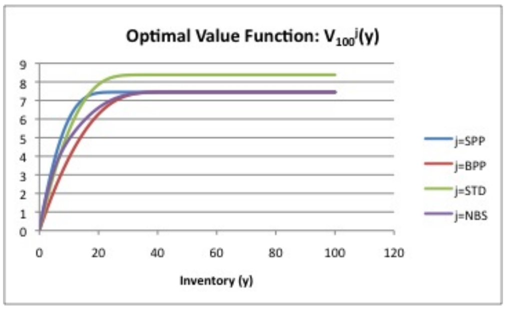
Average price per unit expected to be received, 1000 periods-to-go

FIGURE C.6: Expected quantities sold and average price per unit expected to be received under Model (4.2) specified for the STD mechanism and Model (4.10), for different times-to-go, with $r = 0.05$ and $\lambda = 0.60$

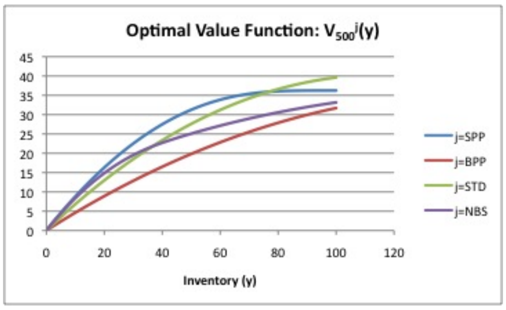
C.2 Results for $\lambda = 0.9, r = 0.05$ (Figures C.7 to C.12)



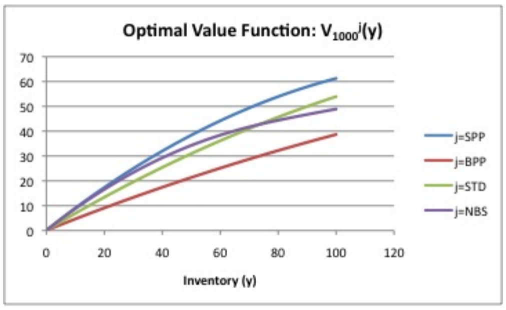
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.7: Optimal value function ratio $(V_t^k(y)/V_t^j(y))$ for different times-to-go, with $r = 0.05$ and $\lambda = 0.90$

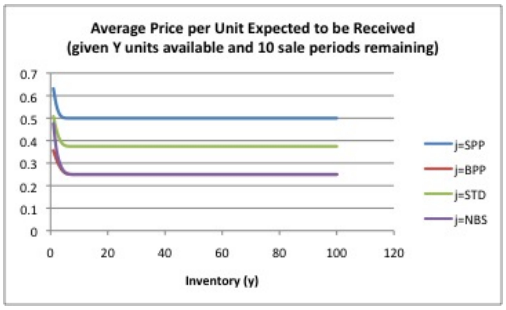
10 periods to go

100 periods to go

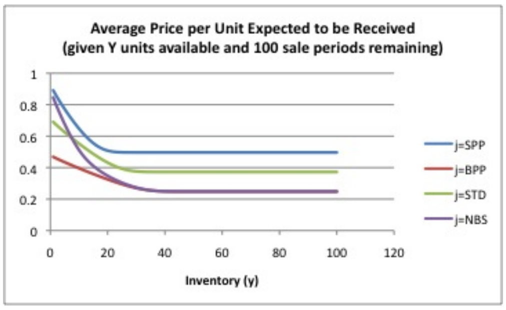
500 periods to go

1000 periods to go

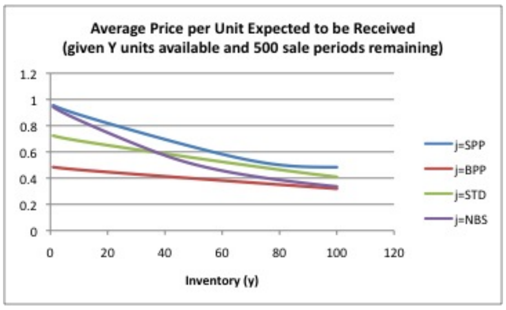
FIGURE C.8: Expected quantity sold under the four mechanisms for different times-to-go, with $r = 0.05$ and $\lambda = 0.90$



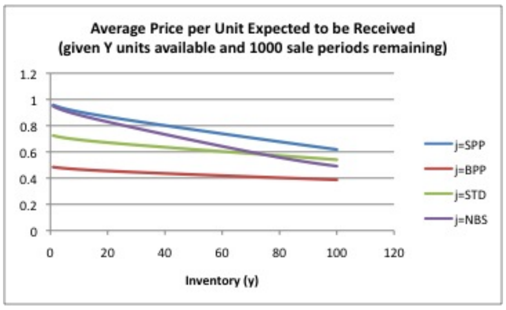
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.9: Average price per unit expected to be received under the four mechanisms for different times-to-go, with $r = 0.05$ and $\lambda = 0.90$

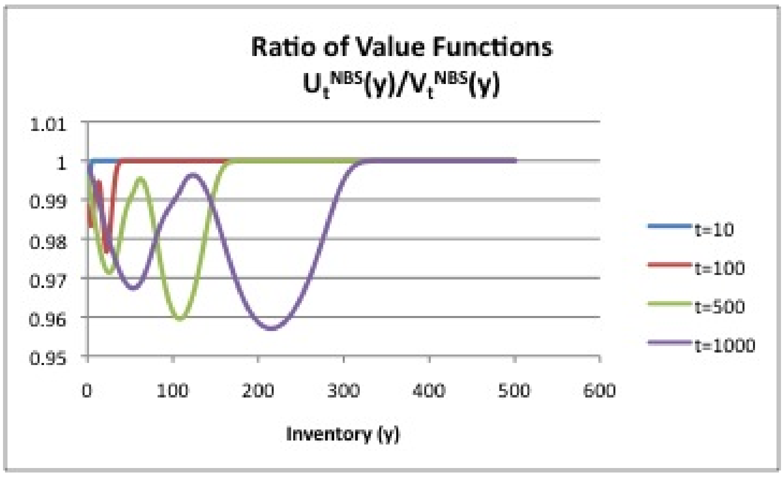


FIGURE C.10: Ratio of approximate optimal value function to optimal value function under the NBS mechanism, with $r = 0.05$ and $\lambda = 0.90$

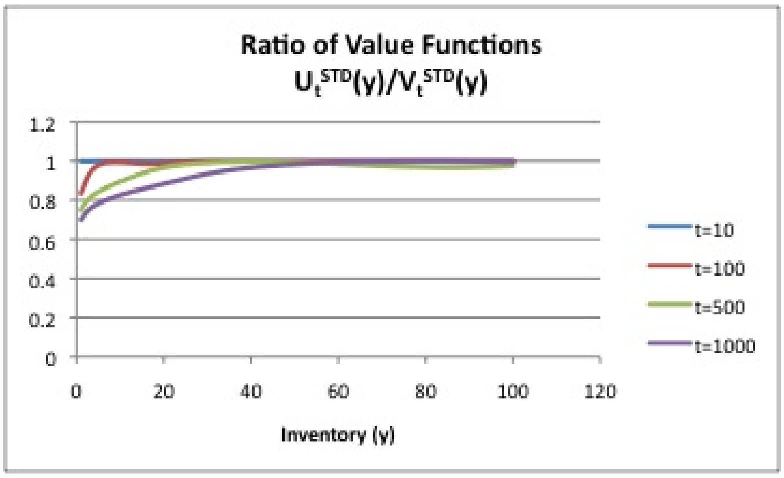
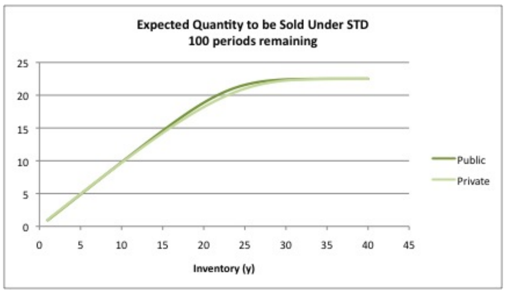
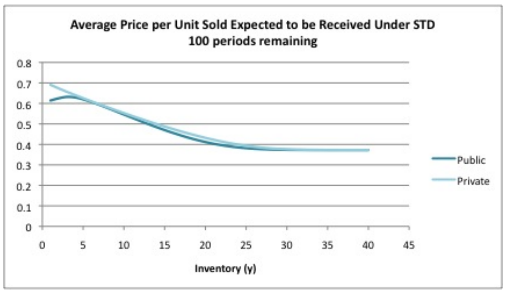


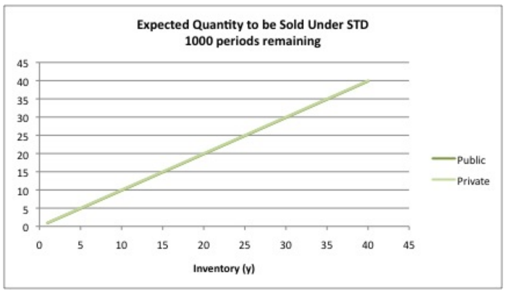
FIGURE C.11: Ratio of approximate optimal value function to optimal value function under the STD mechanism, with $r = 0.05$ and $\lambda = 0.90$



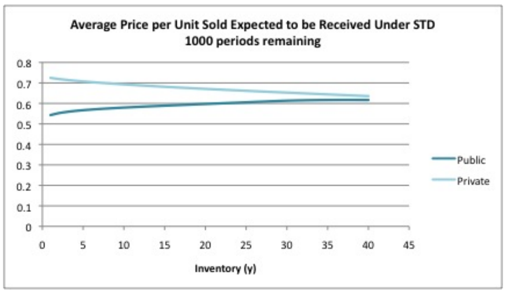
Expected quantity to be sold, 100 periods-to-go



Average price per unit expected to be received, 100 periods-to-go



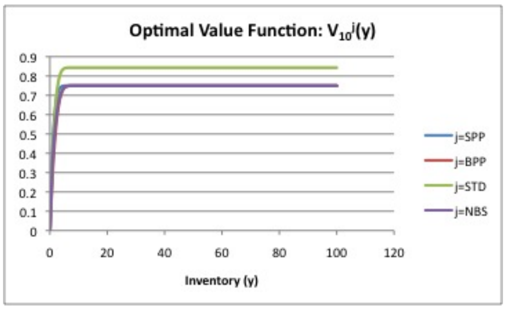
Expected quantity to be sold, 1000 periods-to-go



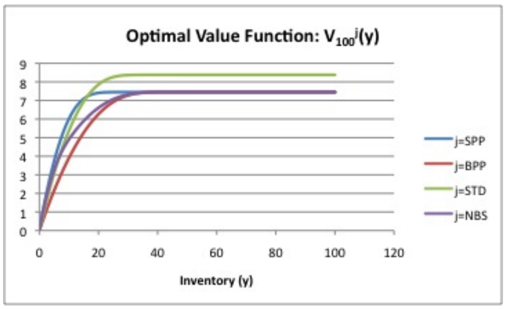
Average price per unit expected to be received, 1000 periods-to-go

FIGURE C.12: Expected quantities sold and average price per unit expected to be received under Model (4.2) specified for the STD mechanism and Model (4.10), for different times-to-go, with $r = 0.05$ and $\lambda = 0.90$

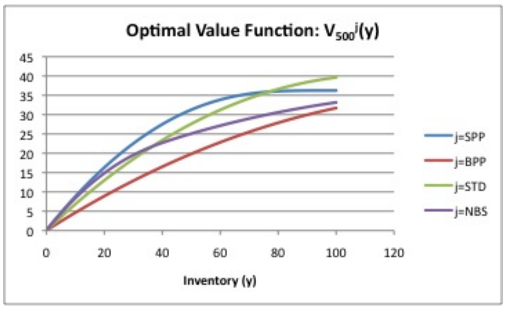
C.3 Results for $\lambda = 0.3, r = 0.10$ (Figures C.13 to C.18)



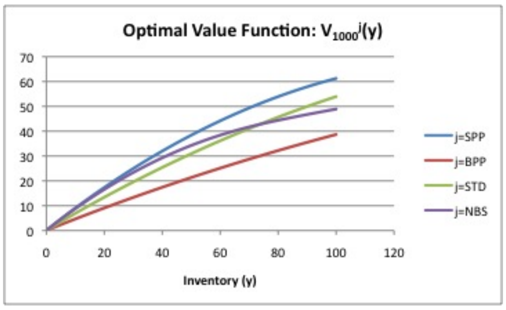
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.13: Optimal value function ratio $(V_t^k(y)/V_t^j(y))$ for different times-to-go, with $r = 0.10$ and $\lambda = 0.30$

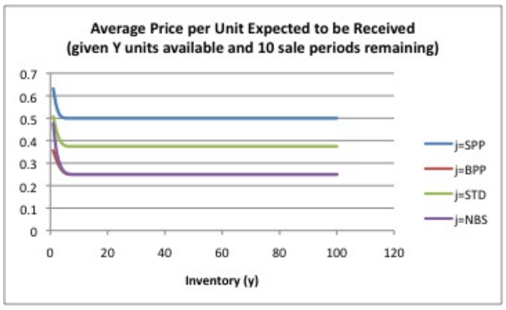
10 periods to go

100 periods to go

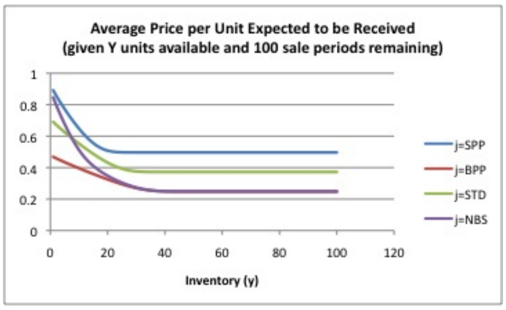
500 periods to go

1000 periods to go

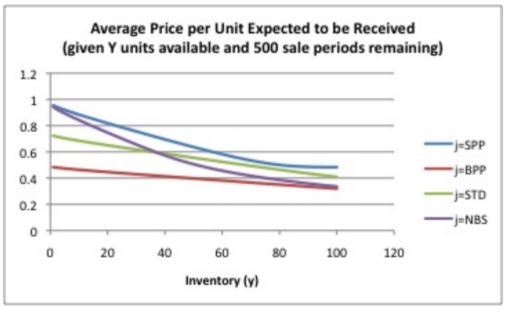
FIGURE C.14: Expected quantity sold under the four mechanisms for different times-to-go, with $r = 0.10$ and $\lambda = 0.30$



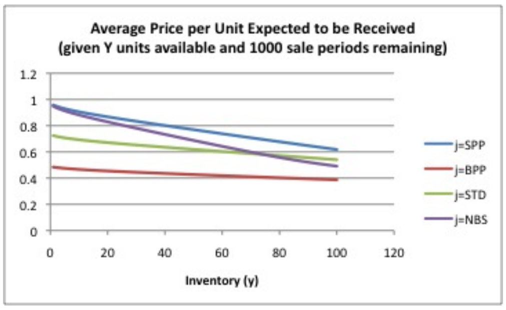
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.15: Average price per unit expected to be received under the four mechanisms for different times-to-go, with $r = 0.10$ and $\lambda = 0.30$

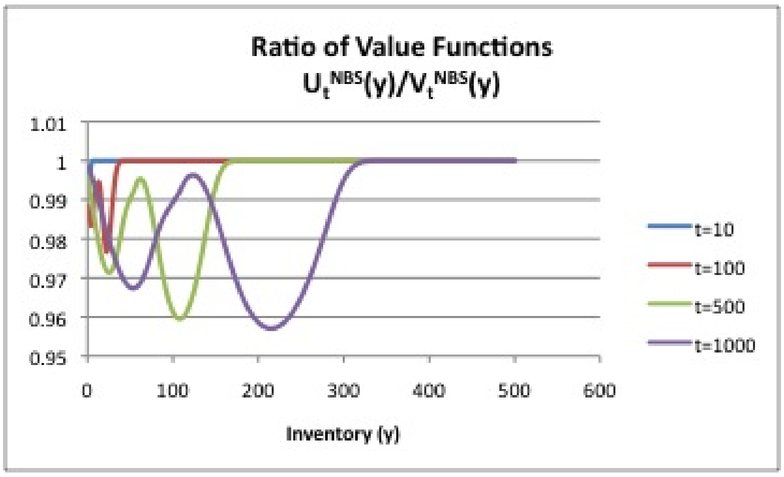


FIGURE C.16: Ratio of approximate optimal value function to optimal value function under the NBS mechanism, with $r = 0.10$ and $\lambda = 0.30$

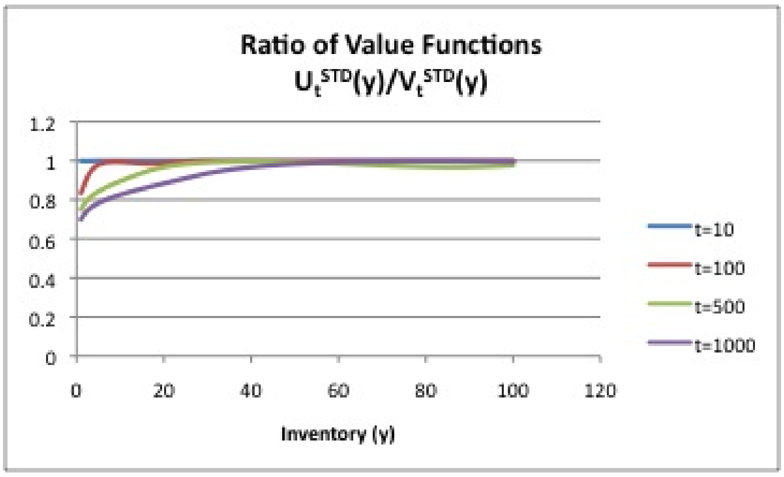
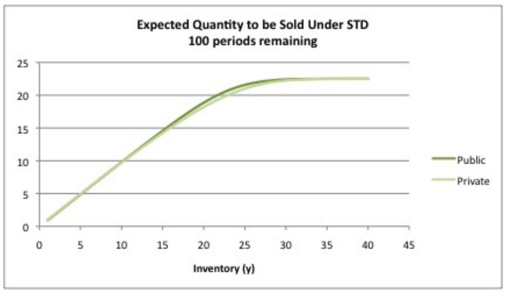
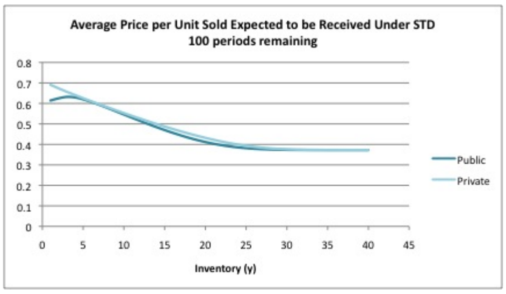


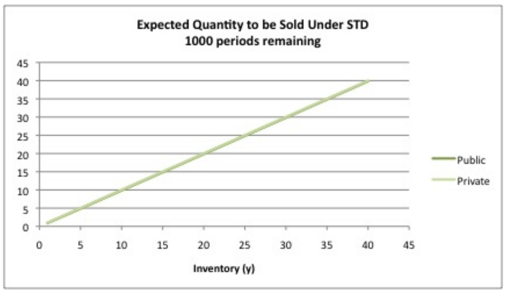
FIGURE C.17: Ratio of approximate optimal value function to optimal value function under the STD mechanism, with $r = 0.10$ and $\lambda = 0.30$



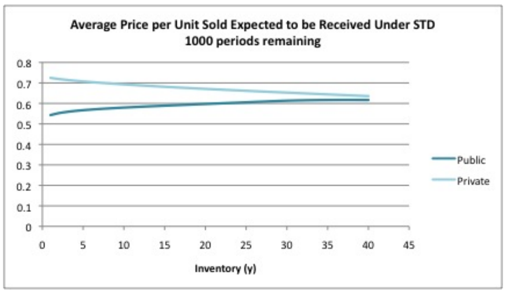
Expected quantity to be sold, 100 periods-to-go



Average price per unit expected to be received, 100 periods-to-go



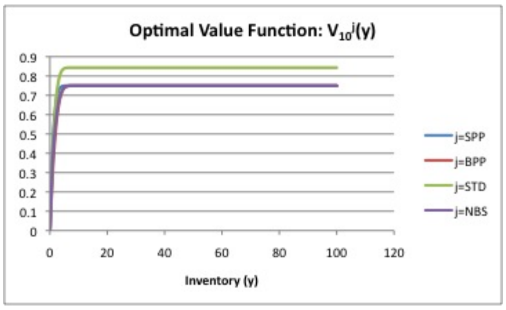
Expected quantity to be sold, 1000 periods-to-go



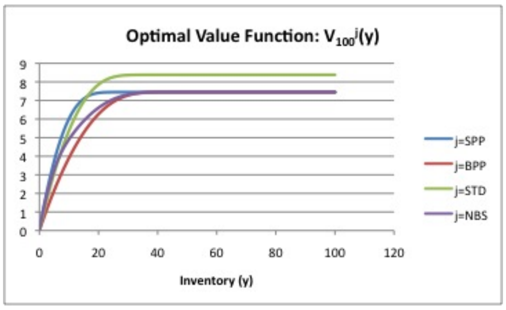
Average price per unit expected to be received, 1000 periods-to-go

FIGURE C.18: Expected quantities sold and average price per unit expected to be received under Model (4.2) specified for the STD mechanism and Model (4.10), for different times-to-go, with $r = 0.10$ and $\lambda = 0.30$

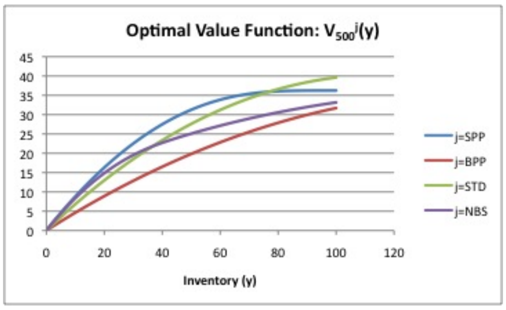
C.4 Results for $\lambda = 0.6, r = 0.10$ (Figures C.19 to C.24)



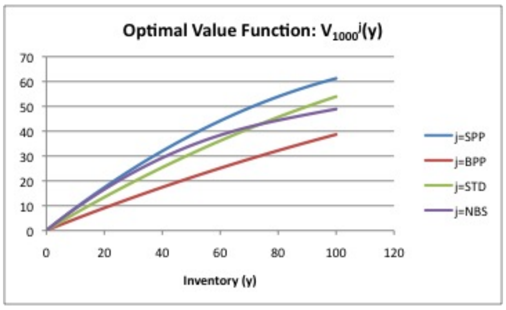
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.19: Optimal value function ratio $(V_t^k(y)/V_t^j(y))$ for different times-to-go, with $r = 0.10$ and $\lambda = 0.60$

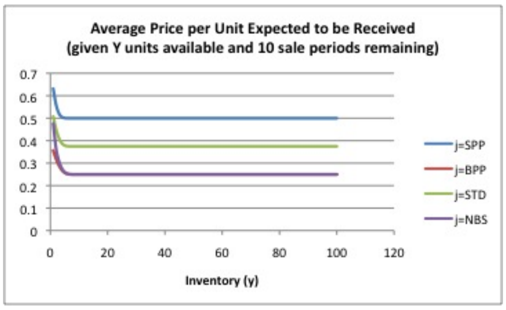
10 periods to go

100 periods to go

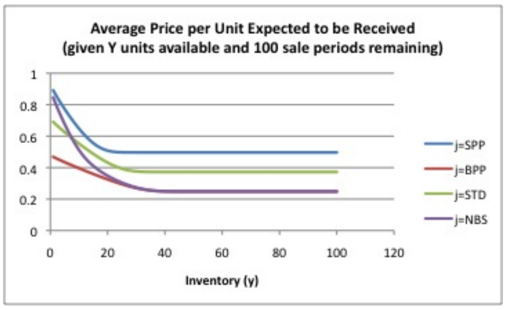
500 periods to go

1000 periods to go

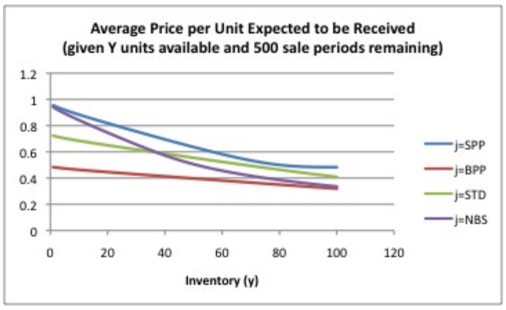
FIGURE C.20: Expected quantity sold under the four mechanisms for different times-to-go, with $r = 0.10$ and $\lambda = 0.60$



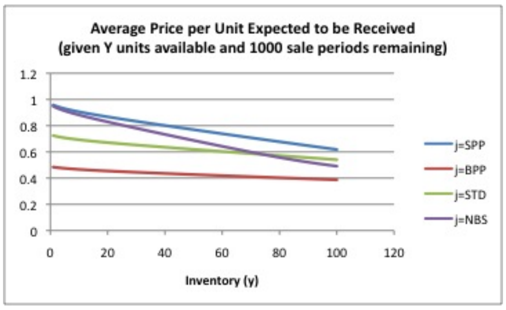
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.21: Average price per unit expected to be received under the four mechanisms for different times-to-go, with $r = 0.10$ and $\lambda = 0.60$

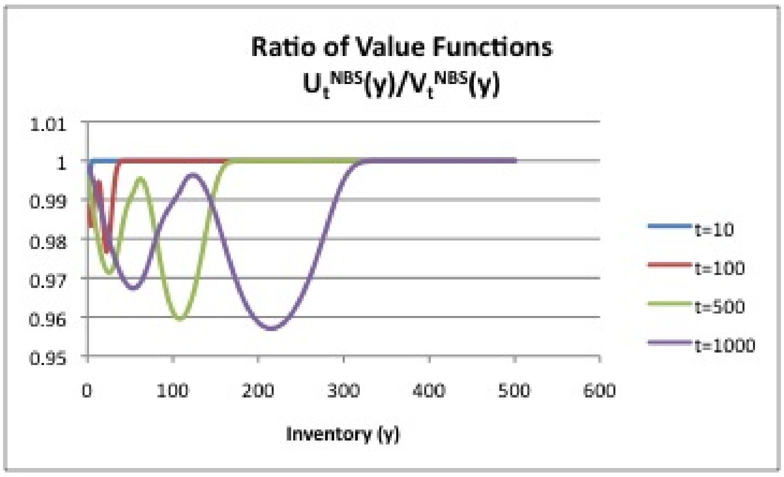


FIGURE C.22: Ratio of approximate optimal value function to optimal value function under the NBS mechanism, with $r = 0.10$ and $\lambda = 0.60$

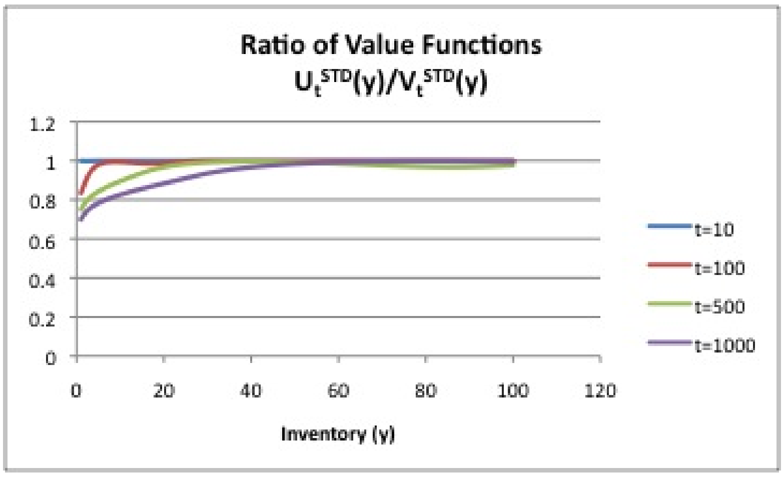
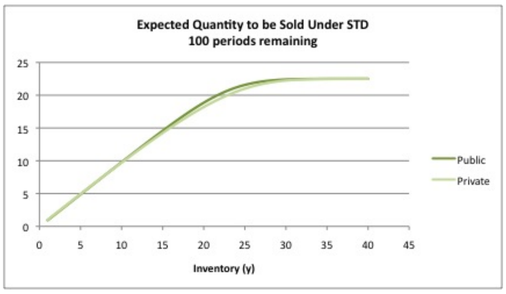
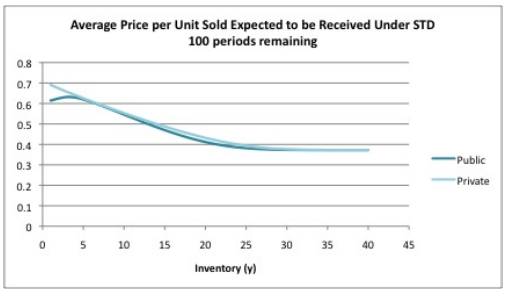


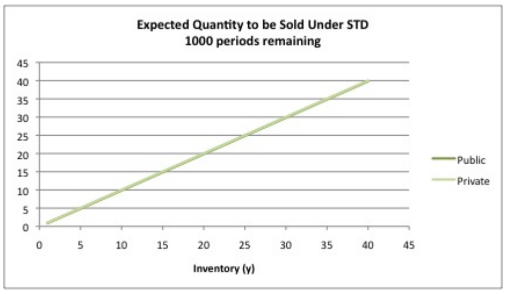
FIGURE C.23: Ratio of approximate optimal value function to optimal value function under the STD mechanism, with $r = 0.10$ and $\lambda = 0.60$



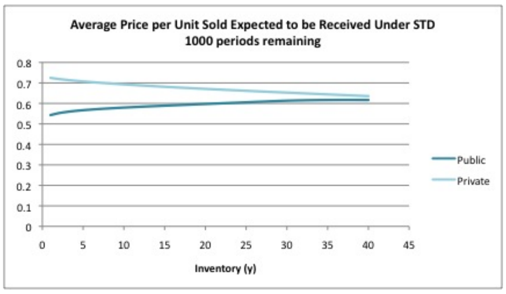
Expected quantity to be sold, 100 periods-to-go



Average price per unit expected to be received, 100 periods-to-go



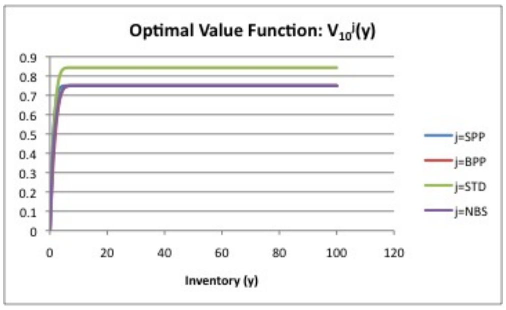
Expected quantity to be sold, 1000 periods-to-go



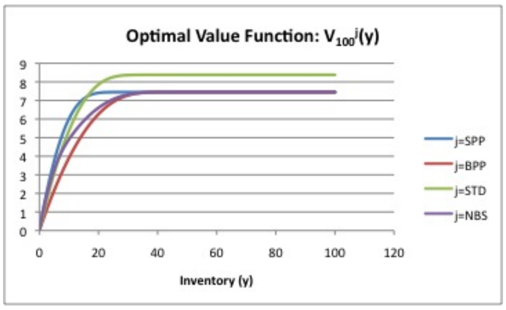
Average price per unit expected to be received, 1000 periods-to-go

FIGURE C.24: Expected quantities sold and average price per unit expected to be received under Model (4.2) specified for the STD mechanism and Model (4.10), for different times-to-go, with $r = 0.10$ and $\lambda = 0.60$

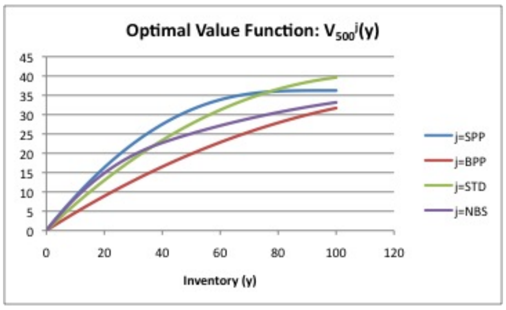
C.5 Results for $\lambda = 0.9, r = 0.10$ (Figures C.25 to C.30)



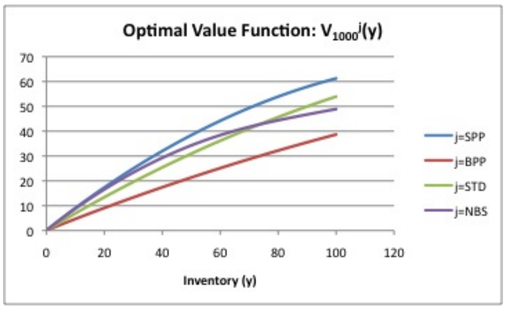
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.25: Optimal value function ratio $(V_t^k(y)/V_t^j(y))$ for different times-to-go, with $r = 0.10$ and $\lambda = 0.90$

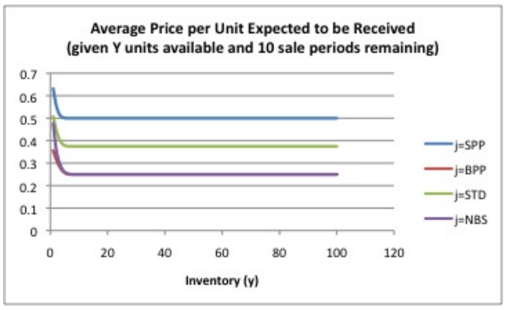
10 periods to go

100 periods to go

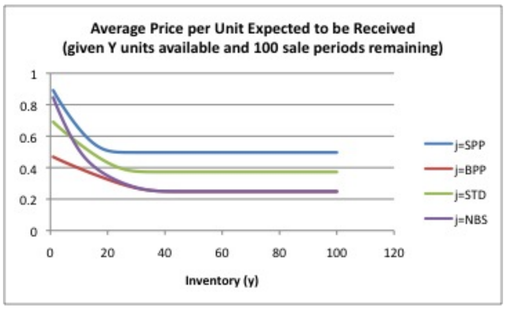
500 periods to go

1000 periods to go

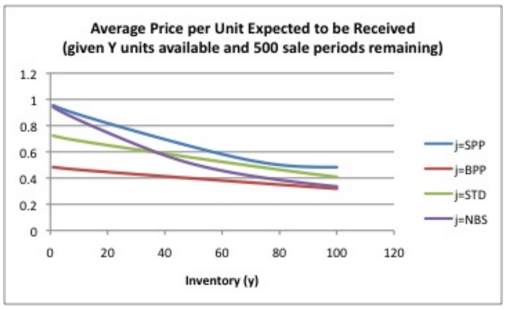
FIGURE C.26: Expected quantity sold under the four mechanisms for different times-to-go, with $r = 0.10$ and $\lambda = 0.90$



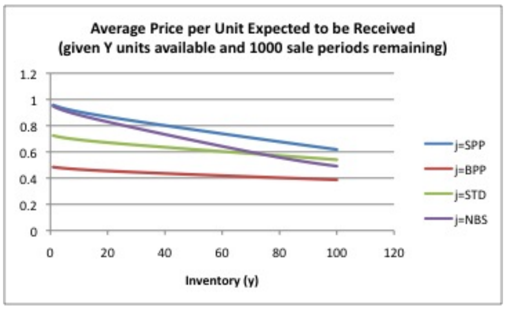
10 periods to go



100 periods to go



500 periods to go



1000 periods to go

FIGURE C.27: Average price per unit expected to be received under the four mechanisms for different times-to-go, with $r = 0.10$ and $\lambda = 0.90$

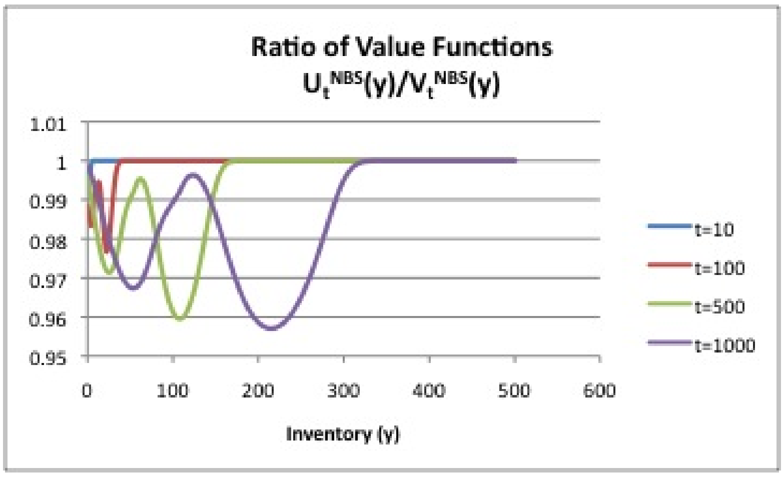


FIGURE C.28: Ratio of approximate optimal value function to optimal value function under the NBS mechanism, with $r = 0.10$ and $\lambda = 0.90$

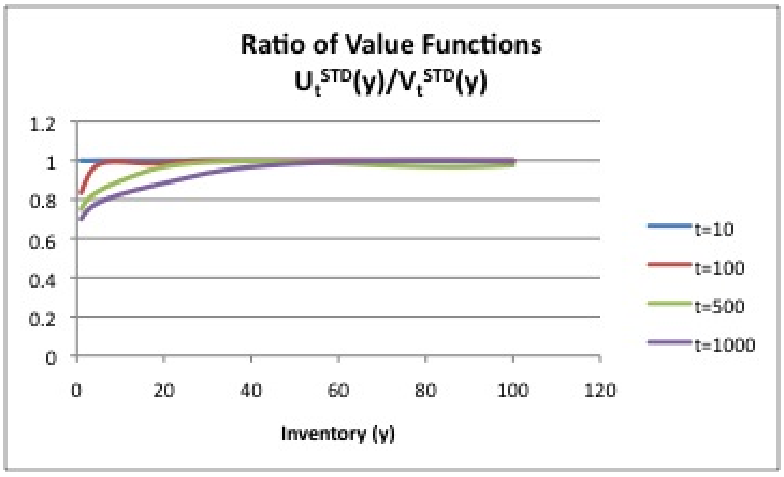
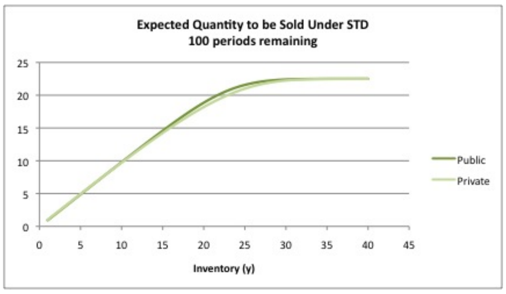
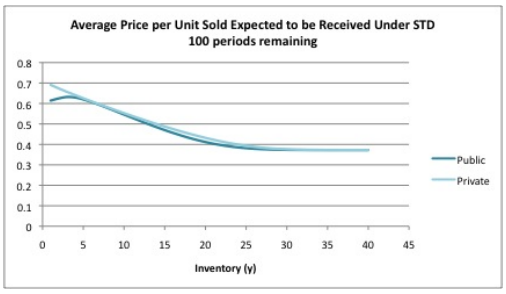


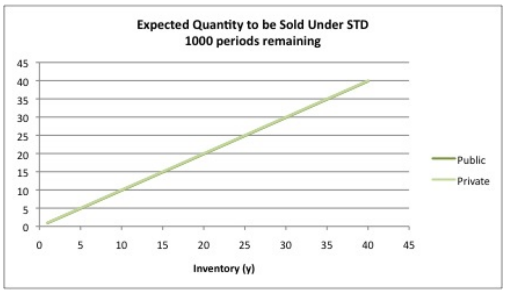
FIGURE C.29: Ratio of approximate optimal value function to optimal value function under the STD mechanism, with $r = 0.10$ and $\lambda = 0.90$



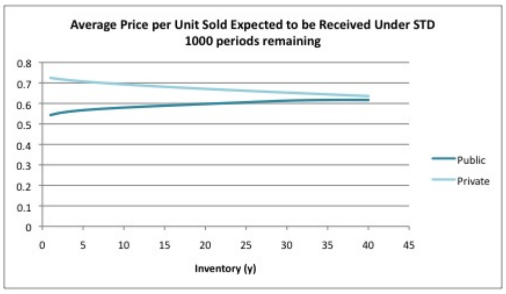
Expected quantity to be sold, 100 periods-to-go



Average price per unit expected to be received, 100 periods-to-go



Expected quantity to be sold, 1000 periods-to-go



Average price per unit expected to be received, 1000 periods-to-go

FIGURE C.30: Expected quantities sold and average price per unit expected to be received under Model (4.2) specified for the STD mechanism and Model (4.10), for different times-to-go, with $r = 0.10$ and $\lambda = 0.90$

References

[1] *SPSS Version 19 Manual*.

[2] C. Adamson. Evolving complaint procedures. *Managing Service Quarterly*, 2(2):439–444, 1993.

[3] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[4] W. Andreassen. What drives customer loyalty with complaint resolution? *Journal of Service Research*, 1(4):324–332, 1999.

[5] N. Ayvaz-Cavdaroglu, S. Kachani, and C. Maglaras. Revenue maximization via sequential bilateral negotiations. Columbia University Center for Pricing and Revenue Management Working Paper Series: No. 2013-2, April 4, 2013. Available at: `http://www7.gsb.columbia.edu/cprm/sites/default/files/files/2013_2_rev_max_bilateral_neg.pdf`.

[6] S. Bach and S. Kim. Online consumer complaint behaviors: The dynamics of service failures, consumers' word of mouth, and organization-consumer relationships. *International Journal of Strategic Comminication*, 6:59–76, 2012.

[7] L. Berry and A. Parasuraman. Listening to the customer - the concept of a service-quality information system. *Sloan Management Review*, pages 65–76, Spring 1997.

[8] A. Bhandari and N. Secomandi. Revenue management with bargaining. *Oper. Res.*, 59(2):498–506, 2011.

[9] J. Blodgett and R. Anderson. A bayesian network model of the consumer complaint process. *Journal of Service Research*, 2(4):321–338, 2000.

[10] V. Bosch and F. Enriquez. Tqm and qfd: exploiting a customer complaint management system. *International Journal of Quality & Reliability Management*, 22(1):30–37, 2005.

[11] A. B. Casado, J. L. Nicolau, and F. Mas Ruiz. The negative effects of failed service recoveries. Working papers. serie ec, Instituto Valenciano de Investigaciones Económicas, S.A. (Ivie), 2008.

[12] J. Cavanaugh. 171:290 model selection - lecture vi: The bayesian information criterion. Available at: http://myweb.uiowa.edu/cavaaugh/ms_lec_6_ho.pdf, 2009.

[13] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010.

[14] H. C. Y. Chan and E. W. T. Ngai. What makes customer discontent with service providers? an empirical analysis of complaint handling in information and communication technology services. *Journal of Business Ethics*, 91:73–110, 2010.

[15] K. Chatterjee and W. Samuelson. Bargaining under incomplete information. *Oper. Res.*, 31(5):835–851, 1983.

[16] P. Chelminski and R. Coulter. An examination of consumer advocacy and complaining behavior in the context of service failure. *Journal of Services Marketing*, 25(5):361–370, 2012.

[17] Y. Cho and J. Fjermestad. *Electronic Customer Relationship Management*, chapter 3: Using Electronic Customer Relationship Management to Maximize/Minimize Customer Satisfaction/Dissatisfaction, pages 34–50. M.E. Sharpe, 2006.

[18] Y. Cho, I. Im, R. Hiltz, and J. Fjermestad. An analysis of online customer complaints: Implications for web complaint management, 2002.

[19] G. Das Varma and N. Vettas. Optimal dynamic pricing with inventories. *Econom. Lett.*, 72(3):335–340, 2001.

[20] M. Davidow. Organizational responses to customer complaints: What works and what doesn't. *Journal of Service Research*, 5(3):225–250, 2003.

[21] K. de Ruyter and A. Brack. European legal developments in product safety and liability: The role of customer complaint management as a defensive marketing tool. *Intern. J. of Research in Marketing*, 10:153–164, 1993.

[22] A. Durrande-Moreau. Waiting for service: ten years of empirical research. *International Journal of Service Industry Management*, 10(2):171–194, 1999.

[23] A. Faed. Handling e-complaints in customer complaint management system using fmea as a qualitative system. In *2010 6th International Conference on Advanced Information Management and Service (IMS)*, pages 205–209, Perth, Australia, November 30 - December 2 2010.

[24] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. 2*. Wiley, 1971.

[25] C. Fornell and B. Wernerfelt. Defensive marketing strategy by customer complaint management: A theoretical analysis. *Journal of Marketing Research*, 24(4):337–346, 1987.

[26] C. Fornell and B. Wernerfelt. A model for customer complaint management. *Marketing Science*, 7(3):287–298, 1988.

[27] S. Foss. Some open problems related to stability. Invited Talk at Erlang Centennial Conference, Copenhagen, 1-3 April, 2009, April 2009. arXiv:0909.0462v1.

[28] S. Foss and D. Korshunov. Heavy tails in multi-server queues. *Queueing Systems*, 52:31–48, 2006.

[29] C. Froehle and A. Roth. New measurement scales for evaluating perceptions of the technology-mediated customer service experience. *Journal of Operations Management*, 22:1–21, 2004.

[30] A. Fundin and B. Bergman. Exploring the customer feedback process. *Measuring Business Excellence*, 7(2):55–65, 2003.

[31] D. Fuscaldo. How to effectively handle customer complaints online, April 8 2011. Available at: http://smallbusiness.foxbusiness.com/technology-web/2011/04/08/effectively-handle-customer-complaints-online/.

[32] J. Gallien. Dynamic mechanism design for online commerce. *Oper. Res.*, 54(2):291–310, 2006.

[33] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. Available at: http://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf.

[34] J. Griffin. Double edged sword of communicating via social media. *Keeping good companies*, pages 659–663, December 2011.

[35] C. Gulas and J. Larsen. Silence is not golden: Firm response and nonresponse to consumer correspondence. *Services Marketing Quarterly*, 33:261–275, 2012.

[36] L. Harrison-Walker. E-complaining: a content analysis of an internet complaint forum. *Journal of Services Marketing*, 15(5):397–412, 2001.

[37] B. D. R. Huberman and F. Wu. Social networks that matter: Twitter under the microscope. *First Monday*, 14(1-5), 2009.

[38] C. M. Hurvich and C. L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297–307.

[39] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188, 2009.

[40] A. Java, T. Finin, X. Song, and B. Tseng. Why we twitter: Understanding microblogging usage and communities. In *Joint 9th WEBKDD and 1st SNA-KDD Workshop '07*, San Jose, CA USA, August 12 2007.

[41] A. Kaplan and M. Haenlein. The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, 54:105–113, 2011.

[42] J. Kiefer and J. Wolfowitz. On the theory of queues with many servers. *Trans. Amer. Math. Soc.*, 78:1–18, 1955.

[43] J. Kiefer and J. Wolfowitz. On the characteristics of the general queueing process with applications to random walk. *Ann. Math. Statist.*, 27:147–161, 1956.

[44] C.-W. Kuo, H.-S. Ahn, and G. Aydin. Dynamic pricing of limited inventories when customers negotiate. *Oper. Res.*, 59(4):882–897, 2011.

[45] S. Lee and Z. Lee. An experimental study of online complaint management in the online feedback forum. *Journal of Organizational Computing and Electronic Commerce*, 16(1):65–85, 2006.

[46] Y. Lee and S. Song. An empirical investigation of electronic word-of-mouth: Informational motive and corporate response strategy. *Computer in Human Behavior*, 26:1073–1080, 2010.

[47] Y. Lu, K. Jerath, and P. Singh. The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. *Management Science*. Forthcoming.

[48] L. Ma, S. Kekre, and B. Sun. Squeaky wheel gets the grease - an empirical analysis of customer voice and firm intervention on twitter. 2012.

[49] D. Maister. The psychology of waiting lines. In M. S. Czepiel, J. and C. Suprenant, editors, *The Service Encounter*.

[50] A. Mattila and D. Mount. The impact of selected customer characteristics and response time on e-complaint satisfaction and return intent. *Hospitality Management*, 22:135–145, 2003.

[51] L. McQuilken, R. Breth, and R. Shaw. Satisfaction, complaining behaviour and repurchase: an empirical study of a subscription service, December 2001.

[52] J. Murphy, R. Schegg, D. Olaru, and C. Hofacker. *Information and Communication Technologies in Tourism 2007*, chapter Exploring Email Service Quality (EMSQ) Factors, pages 425–434. Springer Vienna, 2007. Editors: Sigala, M., L. Mich and J. Murphy.

[53] R. Myerson. Two-person bargaining problems with incomplete information. *Econometrica*, 52(2):461–487, 1984.

[54] R. Myerson. Analysis of two bargaining problems with incomplete information. In A. Roth, editor, *Game-Theoretic Models of Bargaining*, pages 115–147. Cambridge University Press, 1985.

[55] R. Myerson and M. Satterthwaite. Efficient mechanisms for bilateral trading. *J. Econom. Theory*, 29(2):265–281, 1983.

[56] J. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

[57] Q. H. Nguyen and C. Y. Robert. New efficient estimators in rare event simulation with heavy tails. *Journal of Computational and Applied Mathematics*, 261(0):39 – 47, 2014.

[58] U. of California Los Angeles Institute for Digital Research and Education. Annotated spss output ordered logistic regression, Accessed 2014-05-08. Available at: `http://www.ats.ucla.edu/stat/spss/output/ologit.htm`.

[59] E. Omey and S. Van Gulck. Domains of attraction of the real random vector (x,x^2) and applications. HUB Research Paper 2008/02, January 2008.

[60] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, pages 79–86, 2002.

[61] V. Petrov. *Limit Theorems of Probability Theory.* Clarendon Press, 1995.

[62] J. Riley and R. Zeckhauser. Optimal selling strategies: When to haggle, when to hold firm. *Quart. J. Econom.*, 98(2):267–289, 1983.

[63] S. Roth, H. Woratschek, and S. Pastowski. Negotiating prices for customized services. *J. Service Res.*, 8(4):316–329, 2006.

[64] A. Scheller-Wolf. Further delay moment results for fifo multiserver queues. *Queueing Systems*, 34:387–400, 2000.

[65] A. Scheller-Wolf. Necessary and sufficient conditions for delay moments in fifo multiserver queues with an application comparing s slow servers with one fast one. *Oper. Res.*, 51(5):748–758, 2003.

[66] A. Scheller-Wolf and K. Sigman. Delay moments for fifo gi/gi/c queues. *Queueing Systems*, 25:77–95, 1997.

[67] A. Scheller-Wolf and R. Vesilo. Structural interpretation and derivation of necessary and sufficient conditions for delay moments in fifo multiserver queues. *Queueing Systems*, 54:221–232, 2006.

[68] A. Scheller-Wolf and R. Vesilo. Sink or swim together: necessary and sufficient conditions for finite moments of workload components in fifo multiserver queues. *Queueing Systems*, 67:47–61, 2011.

[69] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[70] B. Stauss and W. Seidel. *Complaint Management: The Heart of CRM.* Thomson, 2004.

[71] S. Taylor. Waiting for service: The relationship between delays and evaluations of service. *The Journal of Marketing*, 58(2):56–69, 1994.

[72] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.

[73] M. Thelwall, K. Buckley, G. Paltoglou, D. Ca, and A. Kappasi. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.

[74] R. Wang. Bargaining versus posted-price selling. *Eur. Econom. Rev.*, 39(9):1747–1764, 1995.

[75] W. Whitt. The impact of a heavy-tailed service-time distribution upon the m/gi/s waiting-time distribution. *Queueing Systems*, 36:71–87, 2000.

[76] W. Willinger, M. Taqqu, W. Leland, and W. D. Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. *Statistical Science*, 10(1):67–85, 1995.

[77] R. Wolff. *Stochastic Modeling and the Theory of Queues.* Prentice-Hall, 1989.