

2013-12-17

Order Fitting Policies for Service Systems with Pricing and Inventory Decisions

Salvador Romo-Fragoso
University of Miami, salvador@sthal.net

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation

Romo-Fragoso, Salvador, "Order Fitting Policies for Service Systems with Pricing and Inventory Decisions" (2013). *Open Access Dissertations*. 1130.
https://scholarlyrepository.miami.edu/oa_dissertations/1130

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.

UNIVERSITY OF MIAMI

ORDER FITTING POLICIES FOR SERVICE SYSTEMS WITH PRICING AND
INVENTORY DECISIONS

By

Salvador Romo-Fragoso

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2013

©2013
Salvador Romo-Fragoso
All Rights Reserved

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

ORDER FITTING POLICIES FOR SERVICE SYSTEMS WITH PRICING AND
INVENTORY DECISIONS

Salvador Romo-Fragoso

Approved:

Murat Erkoç, Ph.D.
Associate Professor of
Industrial Engineering
CASCM Director

Hernan Awad, Ph.D.
Assistant Professor of
Management

Nazrul Shaikh, Ph.D.
Assistant Professor of
Industrial Engineering

M. Brian Blake, Ph.D.
Dean of the Graduate School

Shihab Asfour, Ph.D.
Professor and Chair,
Industrial Engineering

ROMO-FRAGOSO, SALVADOR
Order Fitting Policies for Service Systems with
Pricing and Inventory Decisions

(Ph.D., Industrial Engineering)
(December 2013)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Murat Erkok.
No. of pages in text. (135)

The present work is the foundation for our line of research, looking for policy solutions for: order acceptance, capacity management with interruptions, pricing, and trade-offs between different assets.

The first model represents an organization choosing order fitting and pricing policies for a service company selecting between deals and regular jobs with poisson arrivals. The prices for regular jobs are exogenous, whereas winning the deals depends on pricing bids. We derived and compared optimal pricing for two distinct order acceptance policies under a setting where the server had no queue. We investigated the impact of the system parameters on the optimal policies.

In the second part, we extended our analysis to a server that admits queues only for the regular jobs. We proposed optimal state-based pricing for the deals under this setting.

In the last part, we integrate inventory management into the order fitting policies, we studied optimal inventory positions for rotatables at a maintenance-repair-overhaul company. Our study sheds light to understand trade-offs between inventories, pricing and capacity management. This work originated utilizing real industry cases for each of the models proposed, providing market driven solutions for capacity management needs.

Keywords: *order acceptance policies, capacity management, MRO, markov models,*

Hilda my wife, Evangelina my mom, Salvador my dad, and my family for their support.

Acknowledgements

Multiple people have influenced my life, starting with my interest in engineering and science nurtured by those long dinner conversations with my mother and father. Their incredible ways to question everything, guidance to set the conclusion, and then wisdom to derive lessons learned, became a habit which grew in time as part of me.

Uncle Benjamin, who gave me the opportunity to develop skills by allowing me to manage his apple farm along side him. The abilities I gained from a real job at that young age proven invaluable for my undergraduate preparation. I must mention Dr. Jose Grapa for his priceless advices on how to apply my skills in the corporate environment. Then came Jose Luis Ortega, who helped to visualize the future, be thorough, and understand the implications especially for your people, so you do not regret your decision, while maintaining a lively pace on the operations. Also, to acknowledge Don Schmickrath for his mentorship with the understanding what makes corporate America tick in the global context and the opportunity to work in the supply chain field. I am grateful to Dr. Shihab Asfour, who encouraged me to take the next challenge in my life by pursuing my Ph.D. degree, with the quite a gratifying experience to put pragmatic knowledge acquired doing professional and management service into theories to solve the most pressing problems I faced in those years.

The work along Dr. Erkoc Murat, my tutor, the opportunity to be part of the Center for Advanced Supply Chain Management research center continued the right conditions to work solving real business problems. And last but not least, I thank Hilda my wife, whose wisdom and advice always comes to the rescue.

Miami, Florida 2013

Prologue

During my work at Hewlett Packard, I was hired to apply my engineering knowledge to specific problems related to design, development and manufacturing of memory boards. The opportunity to apply science to solve real life problems has been always a constant in my personal work.

When later I received the opportunity to continue my contributions as manager, it became a challenge to understand why the management team, even at the highest levels, usually trusted their “guts” to make decisions when confronted by business problems, often times involving non-linear and stochastic elements, where it is difficult for us humans to extrapolate results based on experience.

It is known without an insight provided by mathematics and statistical modeling, most decisions would be dead wrong. For this reason, when I got assigned to manage supply chain operations, it was a real incentive to use in this particular area observations and data, understand from them; businesses, processes and operations, and subsequently model them until an insight was uncovered and a particular policy to handle the problem was suggested as a solution. Eventually, by applying these principles, enormous value was created with direct impact on profit, customer satisfaction and employee gratification.

During these years I had the opportunity to understand the particular area of supply chain management. At the time, the area struggled with complex and difficult problems which frequently dealt with nervous signals and compounded business strategy decisions. Finding solutions in this area is my motivation to study policies for effective and efficient capacity management under such complex settings.

Table of Contents

List of Figures	viii
Chapter 1. Introduction	1
Chapter 2. Literature Review	10
Chapter 3. Order Admission and Optimal Pricing for Regular Jobs and Big Deals at a Service Company	19
<i>3.1 Overview</i>	<i>19</i>
<i>3.2 Literature Review</i>	<i>21</i>
<i>3.3 Model Basic Settings</i>	<i>23</i>
3.3.1 Deals-Only: Pricing with Single Job Type	24
3.3.2 Mixed Policy with Regular Jobs	30
<i>3.4 Policy Comparisons</i>	<i>40</i>
<i>3.5 Commentary and Insights on Order Acceptance</i>	<i>43</i>
Chapter 4. Order Acceptance with Queuing and State Dependent Pricing	46
<i>4.1 Overview</i>	<i>46</i>
<i>4.2 Literature Review</i>	<i>48</i>
<i>4.3 Model, Assumptions and Methods</i>	<i>50</i>
4.3.1 Notation	53
4.3.2 Model Description and Analysis	54
<i>4.4 Model Evaluation and Numerical Analysis</i>	<i>64</i>
<i>4.5 Discussion on the Order Acceptance with Differentiated Services Model</i>	<i>71</i>
Chapter 5. Order Acceptance For Service Systems with Exchange Inventory	73

<i>5.1 Overview</i>	73
<i>5.2 Literature Review</i>	80
<i>5.3 Basic Settings for the Models</i>	82
5.3.1 Pure Exchange Model	86
5.3.2 Exchange Model Numerical Examples.	95
5.3.3 The Mixed Policy for Exchange and Service	101
5.3.4 Mixed Model Numerical Examples	108
<i>5.4 Pure Exchange vs. Mixed Model Comparison</i>	116
<i>5.5 Discussion on the Order Acceptance with Walk-in Model</i>	122
Chapter 6. Comments and Conclusions	125
Chapter 7. Next Steps	127
Bibliography	132

List of Figures

Figure 1. The deals-only policy model	26
Figure 2. The mixed policy model.....	32
Figure 3. Join Impact of λ_s and v on optimal pricing	35
Figure 4. Traffic density isolines and the optimal price under the mixed policy	37
Figure 5. Service rate for the “deals” and the optimal capacity allocations.....	38
Figure 6. Service rate for the regular jobs and the optimal capacity allocations.....	39
Figure 7. Impact of regular-job arrival rate on price (Δ_p) and revenue (Δ_G) gaps.....	42
Figure 8. Impact of the deals service rate on the price and revenue gaps between policies.	42
Figure 9. The small and big job policy model	56
Figure 10. Expected queue as function of the small jobs service.....	64
Figure 11. Pricing by state (first 7 states only for clarity)	65
Figure 12 a and b. Pricing evolution with one customer as the big deal service level increases.....	66
Figure 13. Queue impact by switching small job to big job service rate.	68
Figure 14. Optimal derivatives vector evaluated on the optimal price.	68
Figure 15. Gain function derivative sensitivity when $n = \infty$ at 0, this figure example has $n = 14$	69
Figure 16. Price vs big deal service rate when there is 1 small customer $x=1$	70
Figure 17 a and b. Capacity vs service rates.....	71
Figure 18. The Pure Exchange and Mixed order policy models.....	85
Figure 19. Expected gain curves.	92
Figure 20. Threshold lines as functions of the number of servers	95
Figure 21. Optimal inventory as function of number of servers ($\mu = 4.5$ and $h = 25$).....	96
Figure 22. Arrival rate vs. optimal inventory ($c = 9, \mu = 4.5, h = 25$	98

Figure 23. Optimal inventory as function of number of servers ($\lambda = 20$ and $h = 25$)	99
Figure 24. Service rate vs. optimal inventory ($c = 9, \lambda = 25, h = 25$)	100
Figure 25. Expected gain curves under the mixed policy with optimal inventory levels of $S^{m*} = 0, S^{m*} \leq c,$ and $S^{m*} > c.$	107
Figure 26. Threshold curves as functions of the arrival rate ($r=100, h=55, \alpha=0.7, \mu=4.5, c=5$).....	109
Figure 27. Threshold curves and optimal inventory levels as functions of the arrival rate ($r=100, h=55,$ $\alpha=0.5,$	110
Figure 28. Impact of the arrival rate on the optimal gain ($r=100, h=55, \alpha=0.5, \mu=4.5, c=5$).....	112
Figure 29. Threshold curves as functions of the service rate ($r=100, h=70, \alpha=0.7, \lambda=20, c=4$).	113
Figure 30. Service rate and optimal exchange inventory ($r=100, h=70, \alpha=0.7, \lambda=20, c=4$).	114
Figure 31. Threshold curves as functions of number of servers ($r=100, h=40, \alpha=0.5, \lambda=20, \mu=4.5$).....	115
Figure 32. Threshold curves under same rate increases in λ and μ ($r=100, h=25, \alpha=0.5, c=5, \rho=2$).	116
Figure 33. Gains Mixed vs. Exchange Models evaluated on I^*	117
Figure 34. Trade-off between Pure and Mix Policy (λ, μ, h, c constant)	118
Figure 35. Effect of ΔG_{Mix^*} vs. G_{Pure^*} at high and low levels of discount.	119
Figure 36. ΔG_{Mix^*} vs. G_{Pure^*} for different arrival rates and high and low cost levels.....	120
Figure 37. Service rate vs alpha Mix Model and Pure Ex S^* and $G(S^*)$	121

Chapter 1. Introduction

This document reflects our investigation results on studies for order acceptance when services firms, facing random demand, are making pricing and inventory control strategic positioning decisions. The analysis focuses on finding policies that lead to efficient use of capacity. Capacity management continues to be a critical component on most of business and industries. Solid understanding of demand and assets—appropriate policies to support sales with the right inventory level for example—is critical for the cash flow of most all enterprises with delivery of goods and services to customers. Particular attention is given to service operations where target customer selection, pricing and—in many cases—inventory control are crucial factors for profitability.

Traditional research on pricing and demand management focuses on tangible products. In an increasingly competitive service economy, there is a substantial need to direct research in this *services* area towards pricing jobs and in particular projects. Since service is usually an intangible product, typically, it is a challenge to price and to manage its demand. Service providers carry out their pricing—for consulting, construction, maintenance, etc.—usually through bidding. This is especially the case for major service projects that require substantial resources and time commitments. Arrivals of such project requests are typically infrequent yet generate considerable revenues for service companies. As such they constitute big deals—as they are called by some in the industry—for service firms.

As mentined previously, bidding is the process most typically employed. However, there are cases where a customer unexpectedly needs a service. In this case, the

service provider must assess carefully the implications of admitting such customers for service.

Service providers which may have steady but small size business flow may attempt to bid for big deals that show up at times. For example, a roof contractor that works on individual residential houses may face a request for proposals for maintenance of all houses in a large residential community, or a printer producer who satisfies its regular and steady demand through retailers may choose to bid for a big-deal that asks for replenishment, installation, and maintenance of all printers operated by a government branch. In fact, some firms may choose to drop all their small regular businesses and concentrate on chasing big deals only. When such opportunities arise, a service provider should develop an optimal order acceptance policy which includes pricing and other business parameters to allocate its resources across regular jobs and big deals, while maximizing its gain.

The initial part of our study was primarily motivated by our research involvement with a major information technology (IT) firm that provided a variety of services including consultation for their customers. The firm maintains under its payroll teams of skilled experts so as to deliver such services. As such, the firm incurred high cost labor which it must utilize efficiently. Consequently, often times the company faces dilemmas in allocating its capacity across low margin yet steady flowing regular orders and “deals” usually generating high revenue yet displaying lumpy—that is sparse—arrivals.

Using Markov processes, in our first case developed in chapter 3, we derive optimal pricing decisions and expected revenues for two possible strategies, we call them; the pure and the mixed model strategies. In the former case, the firm adopts a

policy which only allows for chasing after big deals. This means, the firm does not accept ad-hoc small jobs. In the latter case, both types of jobs are admitted subject to available capacity. We present a detailed analysis for the impact of demand and service rates of both types of models and associated optimal policies and revenues. We compare both strategies to seek answers to some questions relevant for a service firm's strategic fit and selection of markets to serve.

We look to provide answers to questions, like for example: How do the optimal pricing policies differ under two demand management policies? What are the conditions under which the service company prefers one policy to the other? How do the business parameters impact the appeal of either policy for the service provider?

There are situations where a number of more frequent, less valuable activities are scheduled, and then the opportunity for a less frequent, more valuable project arises. The planner needs to understand the alternatives in such cases. Without proper tools planners may resort to naïve solutions: dismiss the customer; let go work at hand and take the more valuable job; perhaps pay a penalty; or increase reserved capacity.

The objective of our study is to develop and compare order acceptance policies for different business management settings. We consider distinctive strategy options, such as, one where the firm switches its allocation of capacity between regular jobs and deals while deciding the price it should bid for a big deal under varying business conditions. Our contribution should help decision makers to rationalize strategic questions, to understand the trade-offs among pricing, order selection and capacity management.

All of these solutions have consequences for the company and planners need tools to evaluate the variables in terms of the biggest benefit for both customer and company while maximizing profit. In all these areas, the problem in common is: What types of orders to accept, what price to bid and then how to allocate resources to maximize gains.

In the first part (Chapter 3) of our study, we consider a service firm which can handle one customer at a time and cannot admit customers when busy. As such, the firm goes after big deals when capacity is available, then we set a model where the firm can have a mix of big deal and regular jobs. Comparing two order acceptance policies, we observe that either one of the policies might generate higher prices and revenues depending on the system parameters. Typically, the preferable strategy depends on the price of standard (ad-hoc) orders and the price sensitivity of the deals.

Our conclusion is the demand rate of the deals and the service rate of the standard jobs are the most effective factors in policy selection decisions. Higher demand rates for the deals make the deals-only (pure) policy more appealing, whereas higher service rates of the standard orders appreciate the mixed policy.

Our results indicate that the service rate of the deals become relevant only when the regular orders' price is sufficiently low. In this case, higher service rate for the deals leads to the deals-only policy.

We also find out that the demand rate of the standard orders does not impact the firm's preference between the two strategy options. At the end, it all comes down to selecting the strategy that delivers more bang-for-the-buck for use of capacity.

Our analysis for pricing and demand management in the first part leads us to the subsequent stage (Chapter 4), where we relax the assumption of not having queues and analyze the optimal pricing decisions for the mixed policy. In this case, when the capacity is busy, standard jobs can be admitted to a single channel queue where they wait for service. However, due to their sizes (both in revenues and service requirements), big deals cannot be put on hold once the bidding is successful. As such, admitting the big deal, the firm cancels (or forwards to third parties) all waiting small jobs in return for penalties so that the firm can work on the big deal. Hence, the challenge in particular for the company is to assess what it takes to put aside standard jobs to bid for and deliver a big job.

This is our rationale for the setting, if any given time, a customer “walks” over an already established queue, think for a moment on those customers, who would likely have contract, already in queue for service. The manager will have to answer two questions: What would be the impact on the queue of the standard jobs (also referred to as the small jobs in comparison to big deals) in terms of the firm’s expected gain? Given a standard price for the service provided to standard jobs, how should the firm adjust and bid for the premium over a big job for this special case? Here, think of premium the incremental price bid for the big job or big deal. Admitting a big deal may have the consequence of disrupting the firm’s already established schedule and customers, which may incur substantial costs and penalties.

The bidding for the big deals involves pricing. Typically, higher bidding price diminishes the probability of winning the deal. The pricing in this case should depend on the system state, which can be defined as the number of standard jobs waiting to be

served. Pricing requires analyzing business conditions and then making a prediction if a given *premium* over the regular price could support the business to deliver a big job.

Major projects that require substantial resources and time commitment, could drain the business if the expected gain is not in line with the business delivered by the standard jobs. On the other hand, opportunities to make a substantial gain could be under evaluated making the business to commit without proper cost recovery and marginal gain. Arrivals of such big job requests are typically infrequent yet generate considerable revenues for service companies, as mentioned before. In this part of the model, our analysis extends to state-based dynamic pricing.

As in the previous case, we observed that service times are the most important factors in developing optimal strategies and policies with big-deals and regular jobs.

In the later part of this work (Chapter 5), we focus on integrating inventory decisions into the order fitting/selection problem. Many service settings involve use of inventories in addition to service capacity. Typically one of the two: service rate or the available inventory becomes the bottleneck for the systems. An obvious example for such systems is the maintenance, repair and overhaul (MRO) industry. MRO services are composed of many and often times by complex activities with specific human resources assignments and spare parts usages. The reader might think on a refurbish line in the aeronautical industry, the MRO firm typically handles all the work that needs to be done to bring the equipment up to par to specifications and regulations in the normal usage of a commercial airplane.

These industries, as in the previous cases, typically have standard customers with regular flow of demand arrivals and once now and then, unexpected arrivals of

emergency customers, who did not know beforehand that they were going to require these services. The service and the inventory management can be disrupted by the mix of service types in this case.

This part of our work is mainly motivated by our involvement with a major MRO company that provide MRO services for landing gear sets of commercial and military aircrafts. Services firms in this sector have a compliance program to make sure certain critical components do have a preventive maintenance, repair, and refurbishment schedule. This program ensures that the parts subject to fatigue are replaced or refurbished long before a catastrophic failure might occur(Regattieri, Gamberi, Gamberini, & Manzini, 2005) .

The MRO company can serve the customer by either overhauling the customer's original equipment (repair strategy) or by exchanging the incoming equipment with an in-stock ready-to-go equipment (exchange strategy). In the former case, the customer must wait until the service of its equipment is completed by the MRO process. This generally results with significant delays for the customer's operations (such as the operation of an aircraft) since the turnover times are usually long. In the latter case, the customer's equipment is immediately exchanged (or rented out) with another from the service provider's inventory which virtually eliminates the turnover times for the customer. The customer's original equipment—once serviced—is added to the MRO company's "exchange inventory" for a future exchange with another customer. In practice such inventory is also known as the rotatable inventory.

The exchange strategy based on rotatable inventory is usually appealing to customers since they eliminate the waiting time for the MRO operations by means of a

straightaway exchange. Typically, the downtime due to equipment overhaul is costly for the owners. Therefore, the equipment owners are willing to pay higher fees for quick turnovers, which may make the exchange option attractive also for the MRO service providers. On the other hand, in order to employ an “exchange” policy, the MRO service provider needs to continuously carry rotatable inventory of overhauled equipment, which may incur substantial costs.

We investigate the value of the exchange approach from the perspective of the MRO firm by incorporating the inventory decisions into the policy selection. We carry out our analysis for two policies: 1) pure exchange policy and 2) mixed policy. In the former case, the MRO adopts a policy on pure exchange where all arriving overhaul requests are responded by an exchange subject to availability of the finished rotatable inventory. If there is no exchange equipment in the MRO company’s inventory, the arriving jobs cannot be accepted. In the mixed policy, once the exchange inventory is depleted, the subsequent job arrivals are still accepted for service, where the customer needs to wait for the completion of her equipment’s overhaul. Under this policy, the service fees depend on whether the customer receives an immediate exchange or waits for the overall overhauling process.

In the final part of this dissertation (Chapter 7) we present some discussion and thoughts on next steps and future extensions for our models. Many real life situations unfortunately do not present themselves with known or established demand patterns. Especially, in the case of big deals, demand can be quite lumpy and historical data may involve many zero entries for demand across past periods. Our models work under parameters given, but when parameters are unknown or too sparse, the pattern could not

be well defined, if a pattern in the demand signal can be detected, and such pattern could be exploited, there might be still useful in terms of providing strategic managerial insights for the practitioners.

This document is organized as follows: Chapter 2 provides a survey of the relevant literature. The basic model for the order fitting and big deal pricing policies is presented in Chapter 3. Chapter 4 extends this model to a setting where the standard jobs are admitted to a queue but cancelled if a big deal is won, and an optimal pricing for accepting such deals is provided. As such, the pricing is to be determined based on a state. Then Chapter 5 analyzes and compares optimal rotatable inventory policies for two strategic order fitting policies: exchange only (pure) and mixed.

In Chapter 6 presents conclusions and final discussion regarding the results of our study.

In Chapter 7, we discuss future work to extend the current study. Finally, Chapter 8 provides the the references used in this body of work.

Chapter 2. Literature Review

There is a broad body of work studying the demand management and order admission problems for the manufacturing and service operations. An early seminal paper by (Lippman & Ross, 1971) considers “the street walker” problem where the main idea is to model the way a street vendor manages her customers to maximize her long run average gain. Customer arrivals follow a Poisson process. The model is set up in a way that the street vendor does not have any backlog and cannot admit customers when busy with another one. As she is ambulating down the street she is permitted to accept a customer if the expected reward is above a certain threshold. Otherwise, she rejects the customer. Contrary to our work, in their paper, the rewards from all customers are assumed to be exogenous and random. The authors propose their model as an analysis tool for a workshop evaluating whether or not an arrived job should be admitted.

Other early work in this area includes (Robinson, 1978), (Nishimura, 1982), and (Ghoneim & Stidham Jr., 1985). Robinson studies the problem from the standpoint of a semi-Markov chain with countable state space and unbounded cost. The proposed model considers two customer types arriving at different queues but attended by the same server. The orders are differentiated by arrival rates, cost of wait and the service time distributions. The author investigates optimal policies for the state-dependent allocation of server to the queues. The analysis reveals that when the service time distribution of the queue is exponential, the priority is pre-emptive if the length of the higher priority queue exceeds a critical value. Nishimura presents a model with constant delivery intervals under a discrete-time single facility system. The optimal admission decision is based on knowledge of the system's work backlog, the service times, and the reward should the job

be accepted. They study Markov decision processes that maximize the rewards earned by running the system over finite and infinite horizons. Ghoneim and Sthidham also consider two queues both controlled by accepting or rejecting arriving customers. They set up the objective to maximize the expected net profit where net benefit is determined by exogenous random rewards.

In a more recent work, (Örmeci & Burnetas, 2005) consider the problem of dynamic admission control in a multi-class Markovian loss system receiving random batches with exponential service rates and exogenous rewards. They show that the optimal admission policy is a sequential threshold policy with monotone thresholds. In a following paper, (Millhiser & Burnetas, 2013) look into the problem for the dynamic control of arrivals of multiple job classes in N-stage production systems with finite buffers and blocking. The authors show a monotonic decline on the net benefit when admitting jobs under system congestion. Their study identifies conditions for jobs to be admitted or rejected regardless of the state of the system. They also explore the blocking versus the admission policy effect.

Feng and Pang (Feng & Pang, 2010) consider the decision-making problem of dynamically scheduling the production of a single make-to stock (MTS) product in connection with the product's concurrent sales in a spot market and a long-term supply channel. The spot market is run by a business-to-business (B2B) online exchange, whereas the long-term channel is established by a structured contract. They investigate optimal policies for production planning and whether or not to accept incoming demand in the spot market simultaneously.

Similar to some of the work mentioned above, our paper considers a markov process based no-queue system with two types of orders, namely, regular jobs (akin to orders from spot market) and deals (analogous to contractual orders). However, our paper differs from the previously mentioned body of work as we incorporate pricing decisions for the deals in addition to admission decisions for the regular jobs. Clearly, pricing is a key strategic factor in demand management and controlling order arrivals, especially for the deals.

Cil *et. al.* (Çil, Örmeci, & Karaesmen, 2009) provides an excellent review of models that study the relations between system parameters and queuing control mechanisms. They present a general framework to investigate the policy implications of the system parameters by using event-based dynamic programming. One such parameter is the pricing. Miller and Buckman (Miller & Buckman, 1987) study a service department's optimal transfer pricing in the context of an M/M/s/s system where no queuing is allowed, the authors conclude that the optimal transfer price equals the expected value of opportunity costs. Johansen (Johansen, 1994) studies a pricing model for M/G/1 queuing system. The jobs are submitted by external customers or by other departments within the organization to which the system belongs. Requests for service of jobs are Poisson and the input control is exercised by announcing the price. In these two systems, the pricing is primarily modeled as internal transfers across departments and determined based on cost.

Gans and Savin (Gans & Savin, 2007) consider a joint dynamic admission control and pricing problem in a multi-server system. The authors consider a system where pre-contracted and walk-in customers arrive to rent an item where the inter-arrival times and

rental durations are assumed to follow exponential distributions. The rental prices are exogenous (preset by the contract) for the contracted customers whereas they are determined by the rental firm dynamically for walk-in customers based on the state of the system. Upon any arrival, if the rental item is not available, the customer is lost and if the arrival is a contracted customer the service company is penalized. The company also has the option of accepting or rejecting any arriving contracted customer in return for a preset penalty. The main difference of our work from theirs is that our analysis has a strategic perspective. We compare two distinct strategic business models: mixed model with two customer classes and pure model with single customer class. Under the dynamic decision making setting, (Gans & Savin, 2007) proves monotonicity for the decision variables and thresholds only when the service rates are identical for all customer types. For the differentiated service rates the authors could not prove monotonicity. However, at the strategic level that our paper considers, we do prove the monotonicity on the price thresholds for customer types with different service rates. In related work, Lewis *et. al.* (Lewis, Ayhan, & Foley, 2002) and Yoon and Lewis (Yoon & Lewis, 2004) also determine monotonicity properties of admission control and pricing problems under queuing settings that are different than ours.

Ziya *et al.* (Ziya, Ayhan, & Foley, 2008) consider static pricing with multiple classes and show how to reduce the pricing decisions to a single-dimension optimization problem. Their work assumes identical service time distributions for all customer classes. They observe that the optimal prices are decreasing in the capacity of the common server. Caro and Simchi-Levi (Simchi-levi & Caro, 2012) investigate optimal static prices that maximize the steady-state revenue for a multiple classes of customers under a Markov

process of arrivals and service. While they consider endogenous pricing for all customers, our model integrates regular orders with exogenous prices into handling of customers for which the firm needs to determine the prices. Their system accepts all arrivals as long as there is available capacity (server) whereas our decision model exclusively considers the admission policies for the regular jobs.

In our model we employ a pricing model that maps prices into winning probabilities. This is a reasonable and common approach in practice Bichler *et. al.* (M Bichler, J Kalagnanam, K Katirciglu, & A J King, 2002). The winning probability can be interpreted as the possibility that the customer does not have a better reservation price, which could be a standing, bid from competition. In this regard, the competition is implicitly incorporated into the model. In this context, losing an arriving job is akin to the job request being won by the competition. In our model, we employ a price mapping that is based on the exponential demand function. Exponential sensitivity is commonly assumed in the operations management literature. A discussion and review of such models and other pricing approaches are reviewed by Bitran (Bitran, 2003). Some examples using similar mappings in different contexts include Gallego and van Ryzin (Gallego & van Ryzin, 1994), You and Wu (You & Wu, 2007), Pachon *et. al.* (Pachon, Erkoc, & Iakovou, 2007), and Wen and Chen (Wen & Chen, 2010).

Currently in the literature substantial work is being done in the differentiated services space, especially on the last years as reviewed few paragraphs above while taking about or initial model and the “street walker” concept.

From the research body we uncover two research streams for capacity management, one that uses admissions controls to allocate fixed capacity, among

different classes of customers with the assumption that all prices are fixed, while the other proposes models where dynamic pricing is the tool to control acceptance see for example Örmeci *et al.* (Örmeci & Burnetas, 2005) and Gans *et al.* (Gans & Savin, 2007)

We leverage these concepts to our model advantage, in the understanding that small jobs will have a published price list, implying prices are relatively fix and subject to change only after long periods of time if market has evolve to a different equilibrium only.

Another piece of the model comes for the line of research in differentiated services, and the consideration for abandonment and loss probabilities (Millhiser & Burnetas, 2013) where the state definition has the consideration of service by customer class, in this model case the service for a small job delivered a much faster rate than the service for a big job. Taking that upon completion of service in queue optimal service rate does not change(Weber & Stidham, 1987) the differentiated service time analysis allows to have the state space expanded to account for the temporary differences between a small and big jobs.

Another important perspective to consider for order fitting strategies is the management of inventory. In the service industry, especially the rotatable inventory in the MRO industry, has become an important factor in leveraging profitability for the service firms. Most MRO operations involve providing service for heavy and expensive equipment. One example is the landing gear services. Landing gear for an aircraft might appear to be deceptively safe. Their robustness and massive sizes of some systems might suggest that this critical component has no vulnerabilities. In fact, improper maintenance might lead to increased overhaul costs and lower flight safety (Canada. Ministry of

Transport, 2009). Given the fact that the average loss of revenue of a landed airplane is large, more than \$29,000 dollars per day per plane (Newsmax, 04/02/2011).¹

Because the nature of the function of the landing gears it is one of the most prevalent areas, where failures in an aircraft are reported. There is a market for airplanes with non-scheduled certification for landing gear where repair would be needed.

On this area there are research lines devoted to planners with regular supply and demand patterns, the area has numerous methods and techniques they can use, queue theory and its standard applications, among many others. The question to be solved is, the results are not aligned with physical data what needs to be applied. For example, when the supply or the demand is lumpy under what conditions a capacity management model can be applied(Isken, Ward, & Littig, 2011).

In this area of order acceptance, there are a good number of works related, in particular, to the project scheduling in the area of Engineering, Procurement and Construction (Ishii, Takano, & Muraki, 2013). The question is then, how do you go about handling this signal lumpiness to minimize disruption and cycle time and, as a consequence, maximize gross margin and profit. Specifically cycle time, and its connotations for service level, customer satisfaction, labor cost reduction, and asset management, has a strong interest from multiple parties to gain insights to help companies and their planners to achieve stronger results (Wu, 2003)

Scheduling and queuing have been used over the years with strong success to solve problems in diverse fields, like transportation, networking, and telecommunications

¹ Although our focus of this study is on the economic implications of landing gear maintenance, the critical safety aspects should not be overlooked.

see for example Banks (Banks, 2010). There is then sufficient work done to help us to understand some basic premises, such as how best identify the objective function. Furthermore, from these studies, we also incorporate the idea of predictable variability, which suggests that we can compensate for factors such as seasonal trends, business cycles, and customer patterns. The system design can compensate with the appropriate capacity.

Other research suggest a way to eliminate decision variables, by using data from other sources that can then turn into an input for the model, thus reducing the dependency on the forecasting of the demands—lumpy demand for example. By understanding more upstream conditions, unpredictable variability can be reduced with a positive impact on the system outcomes. (Romeijnders, Teunter, & Van Jaarsveld, 2012)

Variability, then, is one key factor of cycle time and has a direct relationship with reserved capacity. When variability is substantial, introduction of scheduled capacity is in itself inadvisable from current queue theory insights, unless considerable reserve capacity is available. With the introduction of random unpredicted variability, the problem worsens.

Our research indicates, the approach to continue to look for trade-off between assets, and order acceptance policies, have strong potential specially on MRO (Maintenance and Repairs Organizations) Industries in general and the Aviation field in particular, by integration of scheduled and un-scheduled part consumption (Cohen & Wille, 2006)

There are other research areas that are important in the context of this document. One such area is the one investigating rotables. Rotables is the coined name for a

maintenance replacement modules and the associated practice of proactive maintenance, as opposite to repairables where components are replaced by repair in a more reactive way. The rotables are called generically also because they typically rotate through a closed loop supply chain (Arts, Flapper, & Vernooij, 2012). Rotables are very often used in industries such as Aeronautics, Trains, and Industrial Equipment for Manufacturing.

Chapter 3. Order Admission and Optimal Pricing for Regular Jobs and Big Deals at a Service Company

3.1 Overview

This first model then, will concentrate in providing order acceptance policies helping the manager to make decisions regarding whether to take small jobs when compared to a big job business, and the impact it will have on the business gain.

Traditional research on pricing and demand management focuses on tangible products, as we already pointed out. Since service is usually an intangible product, typically, it is a challenge to price and to manage its demand. Service providers carry out their consulting, construction, and maintenance pricing typically through bidding. This is especially the case for major service projects that require substantial resources and time commitments. Arrivals of such project requests are typically infrequent yet generate considerable revenues for service companies. As such they constitute big deals for such firms.

Service providers that may have steady but small size business flow may attempt to bid for big deals that show up at times. And as represented by the roof contractor, and the IT firm cases mentioned before, in fact, some firms may choose to drop all their small regular business and concentrate on chasing big deals only. When such opportunities arise, a service provider should develop an optimal pricing policy as well as the optimal business strategies to allocate its resources across regular jobs and big deals.

This chapter studies optimal pricing and demand management policies for a firm that faces deals with random arrivals. We consider two streams of order types: one is composed of recurring standard or regular jobs with pre-determined prices (exogenous

prices) and the other involves big deals that require pricing proposals (endogenous prices). In the former case, the prices are determined by the market conditions. Typically, these jobs arrive more frequently with relatively lower service times. They are analogous to commodity products. On the other hand the big jobs typically arrive less frequently and require longer service times. The service provider needs to bid for the service prices to win these deals as they usually are competitive. The probability to secure the big deals diminishes with the quoted price, and the reason for us to employ a function dependant on pricing to reflect this reality.

Remember the objective of our study is to develop and compare optimization models for different demand management settings. Specifically, we consider two distinct strategies: 1) a pure strategy in which the firm commits to bid for deals only and 2) mixed strategy where the firm switches its allocation of capacity between regular jobs and deals.

Also, it is important to keep in mind we are using markov processes, we derive optimal pricing decisions and expected revenues for both the pure and the mixed strategies. Again, we present a detailed analysis for the impact of demand and service rates of both types of orders on the optimal policies and revenues for both strategies. And also we compare both strategies to seek answers to some questions relevant for a service firm's strategic fit and selection of markets to serve:

1. How do the optimal pricing policies differ under two demand management policies?
2. What are the conditions under which the service company prefers one policy over the other?

3. How do the system parameters impact the appeal of either policy for the service provider?

Based on observations and the modeling of the problem, our conclusion is that the threshold values for these accept or reject / price parameters depend primarily on the demand rate of the deals and the service rate of the standard jobs and as we will demonstrate higher demand rates for the deals make the deals-only policy more appealing, whereas higher service rates of the regular orders appreciate the mixed policy, as might be intuitively found.

Further more we develop evidence that the service rate of the deals become relevant only when the regular orders' price is sufficiently low. In this case, higher service rate for the deals leads to the deals-only policy.

Also demand rate of the standard orders does not impact the firm's preference between the two strategy options.

3.2 Literature Review

Gans and Savin (Gans & Savin, 2007) consider a joint dynamic admission control and pricing problem in a multi-server system. The authors consider a system where pre-contracted and walk-in customers arrive to rent an item where the inter-arrival times and rental durations are assumed to follow exponential distributions.

The rental prices are exogenous (preset by the contract) for the contracted customers whereas they are determined by the rental firm dynamically for walk-in customers based on the state of the system. Upon any arrival, if the rental item is not available, the customer is lost and if the arrival is a contracted customer the service

company is penalized. The company also has the option of accepting or rejecting any arriving contracted customer in return for a preset penalty. The main difference of our work from theirs is that our analysis has a strategic perspective.

We compare two distinct strategic business models: mixed model with two customer classes and pure model with single customer class. Under the dynamic decision making setting, Gans and Savin (Gans & Savin, 2007) proves monotonicity for the decision variables and thresholds only when the service rates are identical for all customer types.

For the differentiated service rates the authors could not prove monotonicity. However, at the strategic level that our paper considers, we do prove the monotonicity on the price thresholds for customer types with different service rates. In related work, Lewis (Lewis et al., 2002) and Yoon and Lewis (Yoon & Lewis, 2004) also determine monotonicity properties of admission control and pricing problems under queuing settings that are different than ours.

Ziya (Ziya et al., 2008) consider static pricing with multiple classes and show how to reduce the pricing decisions to a single-dimension optimization problem. Their work assumes identical service time distributions for all customer classes. They observe that the optimal prices are decreasing in the capacity of the common server. Caro and Simchi-Levi (Simchi-levi & Caro, 2012) investigate optimal static prices that maximize the steady-state revenue for a multiple classes of customers under a markov process of arrivals and service. While they consider endogenous pricing for all customers, our model integrates regular orders with exogenous prices into handling of customers for which the firm needs to determine the prices. Their system accepts all arrivals as long as there is

available capacity (server) whereas our decision model exclusively considers the admission policies for the regular jobs.

In our model we employ a pricing model that maps prices into winning probabilities. This is a reasonable and common approach in practice Bichler *et. al.* (M Bichler et al., 2002). The winning probability can be interpreted as the possibility that the customer does not have a better reservation price which could be a standing bid from competition. In this regard, the competition is implicitly incorporated into the model. In this context, losing an arriving job is akin to the job request being won by the competition. In our model, we employ a price mapping that is based on the exponential demand function. Exponential sensitivity is commonly assumed in the operations management literature. A discussion and review of such models and other pricing approaches are reviewed by Bitran (Bitran, 2003). Some examples using similar mappings in different contexts include Gallego and van Ryzin (Gallego & van Ryzin, 1994), You and Wu (You & Wu, 2007), Pachon *et. al.* (Pachon et al., 2007), and Wen and Chen (Wen & Chen, 2010).

3.3 Model Basic Settings

To set up the initial model that will help us to start modeling our order acceptance policies, we consider a service provider that competes for jobs that have stochastic arrival rates and service requirements. The service provider works as a contractor who faces two types of jobs: 1) jobs that require pricing bids, which we refer to as the “deals” and 2) regular jobs that have exogenous market prices. Typically, the jobs in the first group are bigger in size in the sense that they require more resources and they are less frequent. The

service provider's chance of winning these deals depends on the quoted price. As such, the winning probability decreases in the service provider's price bid.

It is assumed that the service company has capacity to work on one job at a time. Since the demand is for service, it is assumed that an arriving demand request should be accepted straightaway otherwise it is lost. That is, the customers will not wait in a queue. Consequently, the service provider will be able to respond to demand if it has available capacity (*i.e.*, not working on any current job).

Based on this setting, our approach is built two-fold. In the first stage, we determine the optimal bidding prices for arriving *deals* under two different job acceptance settings. In the first setting, we consider only *deals* where the service company does not accept any regular job. In the second case, we study the pricing problem where the company accepts both types of jobs (mixed policy).

In the second stage of our analysis, we compare the optimal strategies under both job acceptance settings and identify conditions under which one policy outperforms the other. To focus on deriving managerial insights, we assume that job requirements within a given type are identical. We consider markov settings where job arrival rates follow Poisson processes and the service times are distributed exponentially. As mentioned above, we focus first on the *deals only* case, immediately followed by the *mixed policy* case.

3.3.1 Deals-Only: Pricing with Single Job Type

In the first part of our analysis, we assume the inter-arrival times and service durations are identical across all job requests. Both times follow exponential distributions with

means of $1/\lambda_d$ and $1/\mu_d$ respectively. The service company can win and execute one job/project at a time and as such, does not bid for other jobs while they are under a contract with one. Other times, when the company is not under any contract, it bids prices for arriving job requests. The company needs to decide on price to bid on arriving job requests. The winning probability of a job proposal is modeled as $e^{-\beta p}$, where β is a constant that denotes the price sensitivity of the arriving job requests and p is the company's price bid. Clearly, the winning probability decreases in p and as such the effective arrival rate for deal wins is $\lambda_d e^{-\beta p}$. It is straightforward to see that as the customer's price sensitivity β increases, the winning probability for the service provider diminishes at the same price level.

The resulting Markov chain is depicted in Figure 1. Basically, under this setting there are two states: 1) the company is idling (S_o) or 2) the company is busy working on a "won" project (S_D). Then, the transition rates from state S_o to S_D and from S_D to S_o are $\lambda_d e^{-\beta p}$ and μ respectively. Using limiting probabilities, the steady state probabilities for both states can be computed as follows as functions of price quote:

$$\pi_o(p) = \frac{\mu_d}{\mu_d + \lambda_d e^{-\beta p}} \quad (1)$$

$$\pi_1(p) = \frac{\lambda_d e^{-\beta p}}{\mu_d + \lambda_d e^{-\beta p}} \quad (2)$$

Where π_o and π_1 are the steady state probabilities of S_o and S_D respectively. Using the steady state probabilities, we can write the expected revenue of the company for any given price p , as follows:

$$G_d = p \lambda_d e^{-\beta p} \pi_o(p) = p \frac{\mu_d \lambda_d e^{-\beta p}}{\mu_d + \lambda_d e^{-\beta p}} \quad (3)$$

Since the company bids for jobs only when its capacity is available (that is under S_o), the effective arrival to the system is $\lambda_d e^{-\beta p} \pi_o(p)$. Consequently, the expected revenue is subject to this rate as laid out in the above function. We note that we ignore the cost of capacity maintenance in our computations (such as the labor cost) since they are assumed to be constant regardless of whether the company is busy or idle.

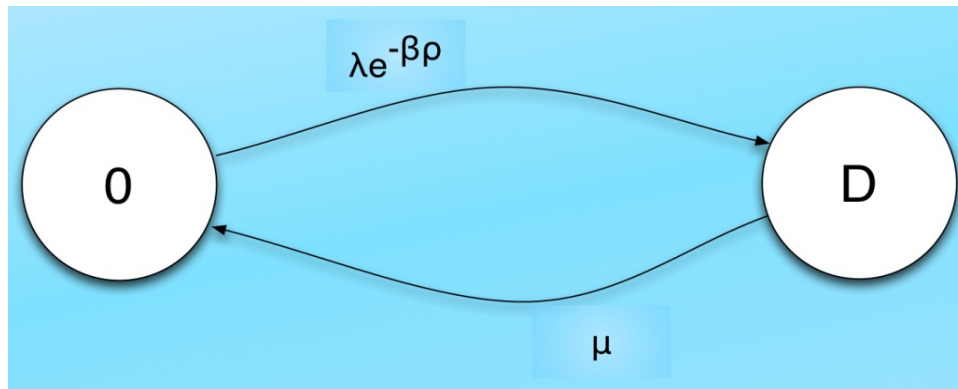


Figure 1. The deals-only policy model

The optimal price maximizes the expected revenue given in (3). Following Proposition provides the optimality condition for the optimal price, which is unique.

Proposition 1. *The expected revenue function given in (3) is unimodular in p with a unique maximizer. At optimality, the optimal price must satisfy the following equation:*

$$p_d^* = (\beta \pi_o(p_d^*))^{-1} \quad (4)$$

Proof. First, we need to show that G_d has a unique maximizer for $p \geq 0$. To achieve this, we define y such that $y = e^{-\beta p}$. With this transformation, we can rewrite the revenue function as follows:

$$G_d = -\frac{1}{\beta} \ln(y) \frac{\mu_d \lambda y}{\mu_d + \lambda y} \quad (5)$$

Then the first and second derivatives with respect to y are

$$\frac{\partial G_d}{\partial y} = -\frac{\mu_d \lambda_d}{\beta(\mu_d + \lambda_d y)} \left(1 + \ln(y) \frac{\mu_d}{\mu_d + \lambda_d y} \right) \quad (6)$$

$$\frac{\partial^2 G_d}{\partial y^2} = -\Delta' \left(1 + \ln(y) \frac{\mu_d}{\mu_d + \lambda_d y} \right) - \Delta \frac{\mu_d}{\mu_d + \lambda_d y} \left(\frac{1}{y} - \ln(y) \frac{\lambda_d}{\mu_d + \lambda_d y} \right) \quad (7)$$

Where Δ is the term in front of the parenthesis in (6). We note that at a stationary point, say y^* , the term in the parenthesis in (6) must be zero. As such, the second derivative evaluated at this point is

$$\frac{\partial^2 G_d(y^*)}{\partial y^2} = -\Delta \frac{\mu_d}{\mu_d + \lambda_d y^*} \left(\frac{1}{y^*} - \ln(y^*) \frac{\lambda_d}{\mu_d + \lambda_d y^*} \right) \quad (8)$$

Since $y^* > 0$ and $\ln(y^*) < 0$, the above term returns a strictly negative value implying that at any stationary point the expected revenue function must be concave. Consequently, we conclude that the function is unimodal with a unique stationary point that maximizes the function. By making the first derivative given in (6) equal to zero, we can easily observe that the price value that satisfies (4) corresponds to the unique stationary point and as such, it is the optimal price for the service company to bid. \diamond

Unfortunately, the optimality condition given in (4) does not provide a closed form solution for the optimal price; however, its numeric value can be computed easily via a line search. Under optimal price, the optimal revenue rate can be rewritten as

$$G_d^* = \frac{1}{\beta} \lambda_d e^{-\beta p_d^*}. \quad (9)$$

Using above results we can make insightful observations on the optimal pricing policy. First we investigate the sensitivity of the optimal price to system parameters.

Lemma 1. *The optimal price increases in arrival rate λ_d and, decreases in service rate μ_d and price sensitivity β .*

Proof. To determine the impact of λ_d , we need to compute the first derivative of optimal price, p^* . From (4) and using partial derivatives, we get

$$\frac{\partial p_d^*}{\partial \lambda_d} = \frac{1}{\beta} \left(-\frac{1}{\pi_0^2} \frac{\partial \pi_0}{\partial \lambda_d} - -\frac{1}{\pi_0^2} \frac{\partial \pi_0}{\partial p_d^*} \frac{\partial p_d^*}{\partial \lambda_d} \right) \quad (10)$$

Which can be rewritten as

$$\frac{\partial p_d^*}{\partial \lambda_d} = \frac{e^{-\beta p_d^*}}{\beta(\mu_d + e^{-\beta p_d^*})} \quad (11)$$

Clearly, the right hand side in the above equation is positive implying that the optimal price is increasing in λ_d . The proofs regarding the impact of μ_d and β can be obtained in similar method. \diamond

If the company experiences more frequent arrivals of job requests its steady state probabilities for working on a project increases. This provides the company with opportunity to increase its price as the risk of being idle diminishes with increased arrival frequency. On the other hand, higher service rates translate into lower prices due to the fact that they lead to higher turnovers. Higher turnovers help the company turn the job quicker and thus let them win more deals. Finally, as the price sensitivity increases, the probability of winning a job at the same pricing level decreases. To offset the drop in demand (winning probability) the company needs to lower its price.

By means of the Envelop Theorem it is straightforward to observe that the optimal revenue increases in arrival rates as expected. The same is also true for the service rates. Although higher service rates lead to lower bid prices, the steady state revenues for the company increase. This is due to the fact that the gain from increased

number of won deals outstrips the drop in pricing. We also observe that as the customers become more sensitive to prices the expected revenue goes down.

The joint impact of the arrival rate λ_d and service rate μ_d leads to different results for the optimal price and revenue. Let $\rho_d = \lambda_d/\mu_d$ represents the offered load for the system (*i.e.*, the traffic density). Clearly, if both values increase in the same rate, the offered load does not change. In this case we get $\pi_0^* = 1/(1 + \rho_d e^{-\beta p_d^*})$ and hence we observe that the steady state probability does not change as long as the offered load stays unchanged with the increase in both λ_d and μ_d . From (4), this implies the same thing for the optimal price. Consequently, we conclude that increase in λ_d and μ_d will not impact the optimal price as long as the rate of increase is identical for both parameters. In general, the optimal price decreases in the offered load. However, we cannot make the same conclusion regarding the optimal revenue. Since revenue increases in both the arrival and service rates, higher revenues will be enjoyed with increased rates even when the offered load stays unchanged. The relation between the optimal revenues and the offered load is not monotonic and depends on the factor that changes the offered load.

A quite interesting observation that we make is the fact that optimal winning probability does not change with price sensitivity.

Lemma 2. *The optimal winning probability, $e^{-\beta p_d^*}$, is independent of the price sensitivity, β .*

Proof. Taking the derivative of the optimal winning probability with respect to β , we get

$$\frac{\partial e^{-\beta p_d^*}}{\partial \beta} = -e^{-\beta p_d^*} \left(p_d^* + \beta \frac{\partial p_d^*}{\partial \beta} \right) \quad (12)$$

Using the approach employed in the proof of Lemma 1, we get

$$\frac{\partial p_d^*}{\partial \beta} = -p_d^*/\beta \quad (13)$$

Plugging the above result in (12), we can observe that the right hand side returns zero implying that the winning probability under optimal pricing does not depend on β .

◇

We note that the above result implies that the steady state probabilities with optimal pricing are also independent of the price sensitivity. That is under optimal policy, the frequency of being under a contract for the company is independent of how its customers are sensitive to pricing. This is due to the fact that on the long run the company would balance the increased price sensitivity by decreasing its price. This is an interesting result as it points that there is a unique arrival rate (*i.e.*, optimal effective arrival rate) that the company must generate through its pricing regardless of the type of its customers in terms of their sensitivity to price.

3.3.2 Mixed Policy with Regular Jobs

We extend the above model to study the optimal pricing policies when the company has a regular job flow in addition to the job requests as defined in the previous section. The company does not need to bid for regular jobs as they represent service requests from steady and/or existing customers. These are in general well-defined projects with relatively shorter processing times, more frequent arrivals and fixed exogenous prices. For the “deals”, we assume the similar setting described in the previous section. We investigate the optimal pricing bids for the deals under this policy and compare the mixed model and the “deals-only” strategy.

We let λ_r and μ_r denote arrival and service rates for regular jobs. As in the previous case, we define λ_d and μ_d as the arrival and service rates for the potential deals.

The revenue for each regular job is exogenous and denoted by v . As in the previous case, the company can receive or bid for jobs only when it is idle, that is when it is not under any contract. Clearly, in this case the company can be in one of the following three states at any time: 1) idle (S_o), 2) busy with a regular job (S_R), or 3) busy serving a won deal (S_D). The resulting Markov chain is depicted in Figure 2.

Solving limiting probability equations we get

$$\pi_o(p) = \frac{\mu_d \mu_r}{\mu_d(\lambda_r + \mu_r) + \mu_r \lambda_d e^{-\beta p}} \quad (14)$$

$$\pi_r(p) = \frac{\mu_d \lambda_r}{\mu_d(\lambda_r + \mu_r) + \mu_r \lambda_d e^{-\beta p}} \quad (15)$$

$$\pi_d(p) = \frac{\mu_r \lambda_d e^{-\beta p}}{\mu_d(\lambda_r + \mu_r) + \mu_r \lambda_d e^{-\beta p}} \quad (16)$$

Where $\pi_o(p)$, $\pi_d(p)$ and $\pi_r(p)$ are the steady state probabilities of S_o , S_D , and S_R respectively. Using the steady state probabilities, we can write the expected revenue of the company for this setting as follows:

$$G_m = \mu_d \mu_r \frac{p \lambda_d e^{-\beta p} + v \lambda_r}{\mu_d(\lambda_r + \mu_r) + \mu_r \lambda_d e^{-\beta p}} \quad (17)$$

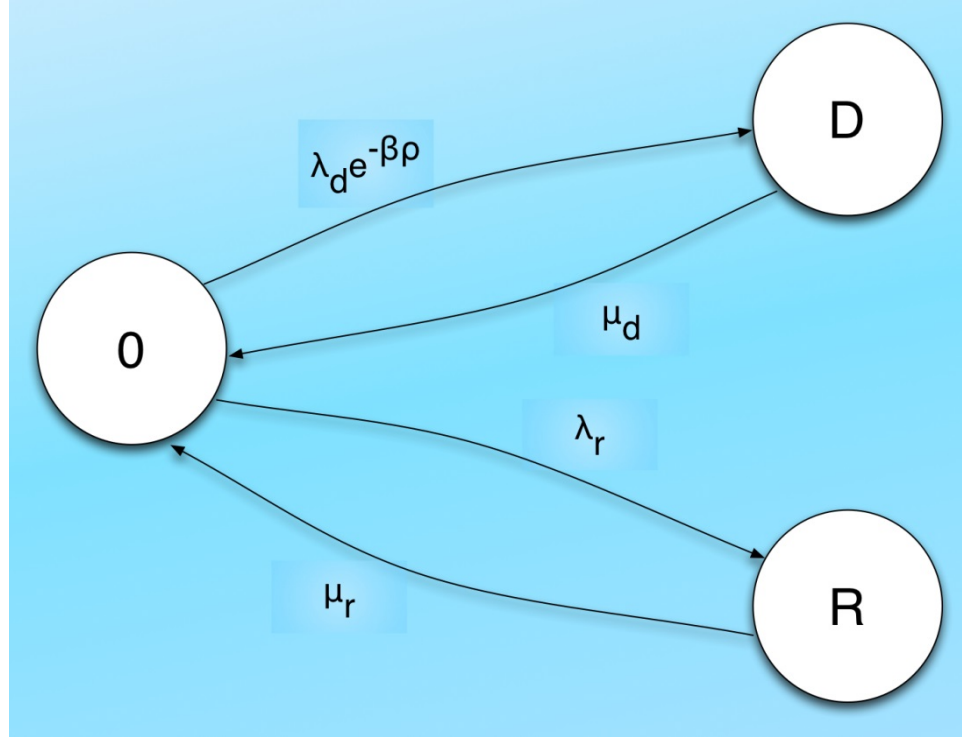


Figure 2. The mixed policy model

Similar to the previous case, we prove that the above function is unimodular in p with a unique maximizer. Following Proposition provides the optimality condition for the optimal price, which is unique.

Proposition 2. *The expected revenue function given in (17) is unimodular in p with a unique maximizer. At optimality, the optimal price must satisfy the following equation:*

$$p_m^* = \frac{\mu_d + \beta v \mu_r \pi_r}{\beta \mu_d (1 - \pi_d)} \quad (18)$$

Proof. First, we need to show that G_m has a unique maximizer for $p \geq 0$. Similar to the proof of Proposition 1, we employ y where $y = e^{-\beta p}$. With this transformation, we can rewrite the revenue function as follows:

$$G_m = -\frac{1}{\beta} \ln(y) \mu_d \pi_d + v \mu_r \pi_r \quad (19)$$

Then the first and second derivatives with respect to y are

$$\frac{\partial G_m}{\partial y} = -\frac{\pi_d}{\beta y} (\mu_d + \ln(y)\mu_d(1 - \pi_d) + \beta v \mu_r \pi_r) \quad (20)$$

$$\frac{\partial^2 G_m}{\partial y^2} = -\frac{(1-\pi_d)}{y} \frac{\partial G_m}{\partial y} - \frac{\mu_d \pi_d^2}{\beta y^2} \quad (21)$$

Since at a stationary point $\frac{\partial G_m}{\partial p} = 0$, from (20) and (21), the second derivative evaluated at that point will be $-\mu_d \pi_d^2 / \beta e^{-2\beta p}$. Consequently, any stationary must be a local maximizer implying that there in fact is a unique stationary point that maximizes G_m . By making the first derivative given in (20) equal to zero and solving it for p , we can easily observe that the price value that satisfies (18) corresponds to the unique stationary point and as such, it is the optimal price for the service company to bid under the mixed policy.

◇

As in the previous case, we observe that the first order optimality condition does not result with a closed form definition for the optimal price. However, the optimal value can easily be found with a simple line search. Following equation gives the optimal average revenue:

$$G_m^* = \frac{\mu_r (\lambda_d e^{-\beta p_m^*} + \beta v \lambda_r)}{\beta (\mu_r + \lambda_r)} \quad (22)$$

Where p^* is defined by (18). Next we discuss the influence of system parameters on the pricing decision and the optimal revenue.

Lemma 3. *The optimal price for deals increases in price of regular jobs, v , arrival rate of deals λ_d and, service rate for regular jobs, μ_s . On the other hand, the optimal price decreases in price sensitivity β and service rate for deals, μ_d .*

Proof. The proof can be carried out for each parameter similar to Lemma 1. Similar to the previous case, the optimal price is increasing in arrival rate of the deals λ_d , and decreasing in service rate of the deals μ_d , and price sensitivity β . In addition, we observe that the bidding price for deals increases in ν , implying that as the price/marginal revenue from regular jobs increases the optimal price bid for deals also increases. The intuition is that the increased marginal revenue from regular jobs decreases the risk for loss of opportunity and hence provides the company with leverage for asking higher prices from the arriving deals.

The optimal bidding price increases in the service rate of regular jobs μ_r . Basically, this indicates that as the service for regular job becomes faster, the company will have the opportunity of increasing its revenues from higher turnover on these jobs. As such relative need for winning a deal is lesser and consequently the bidding prices can be increased for the deals. On the other hand, as for arrival rates of regular jobs, the impact may go both ways depending on the revenue margin of the regular jobs. Typically, when the regular jobs are priced sufficiently high, the higher arrival rates for regular jobs provide leverage for bidding higher prices for the deals. The increased revenue from regular jobs compensates the decreased winning probabilities due to price increase. On the other hand, higher arrival rates with low ν will lead to diminished efficiency for the service capacity. To rectify this, the optimal price for the deals is lowered and thus, the share of deals on capacity usage is increased. The joint impact of λ_d and ν is illustrated by Figure 3.

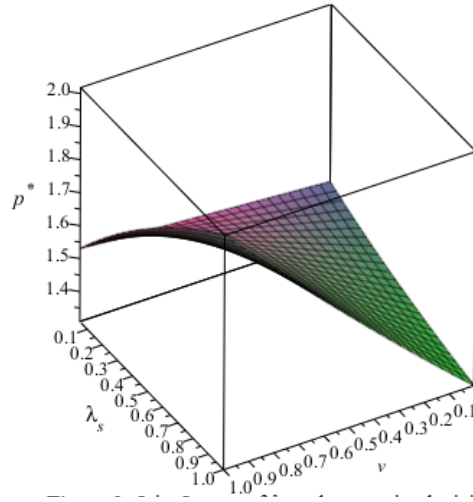


Figure 3. Join Impact of λ_s and v on optimal pricing

With the mixed policy, we observe that Lemma 2 holds no more. That is, in this case, the optimal winning probability does depend on the price sensitivity.

Lemma 4. *The optimal winning probability under mixed policy, $e^{-\beta p_m^*}$, strictly decreases with the price sensitivity, β .*

Proof. Using (18) and taking the derivative of the optimal winning probability with respect to β , we get

$$\frac{\partial e^{-\beta p_m^*}}{\partial \beta} = -e^{-\beta p_m^*} \left(p_m^* + \beta \frac{\partial p_m^*}{\partial \beta} \right) \quad (23)$$

Using the approach employed in the proof of Lemma 1, we get

$$\frac{\partial p_m^*}{\partial \beta} = -1/\beta^2 \quad (24)$$

Plugging the above result in (23), we get

$$\frac{\partial e^{-\beta p_m^*}}{\partial \beta} = -e^{-\beta p_m^*} \left(p_m^* - \frac{1}{\beta} \right) \quad (25)$$

It is straightforward to observe in (18) that the term in the above parenthesis must be positive implying that the derivative returns a negative value. Hence the proof is completed.

◊

The above result indicates that the decrease in price will occur at a lower rate leading to a decrease in win probabilities or in other words, effective “win rates”. Therefore we deduce that the pricing response of the service company to increased price sensitivity is less aggressive compared to the “deals only” case because of the fact that under the mixed policy the company can enjoy revenue from regular jobs in addition to the deals.

Lemma 3 implies that the optimal bidding price increases in the offered load (ρ_d) of the deals. We also observe that for a given price, the long run capacity utilization (*i.e.*, steady state probabilities) is a direct function of ρ_d . That is, the steady state probabilities do not change with the arrival and service rates of the deals as long as ρ_d is unchanged.

However, in contrast to the deals-only case, our analysis reveals that the varying arrival and service rates will lead to different pricing even under same ρ_d . As such, there is no direct mapping between the optimal price and the offered load.

Lemma 5. *Under any given fixed offered load (ρ_d) for deals, the optimal price decreases as the arrival and service rates increase.*

Proof. We first note that in order to keep the offered load unchanged the increase in the arrival rate and the increase in the service rate must be in same proportion. In this case, since the steady state probabilities stay unchanged, the first derivative given in (20)

decreases as both arrival and service rates increase. Since the profit function is unimodular with a concave peak, the optimality condition is satisfied at a decreased price.

◇

This is an interesting result in that it is optimal for the service provider to lower its prices with jointly increasing arrival and service rates even when the offered load does not change. Figure 4 illustrates the impact of offered load on the optimal pricing. Each curve represents an isoline that represents the points of equal traffic densities. It is clear from the figure that as the arrival and service rates increase the price goes down.

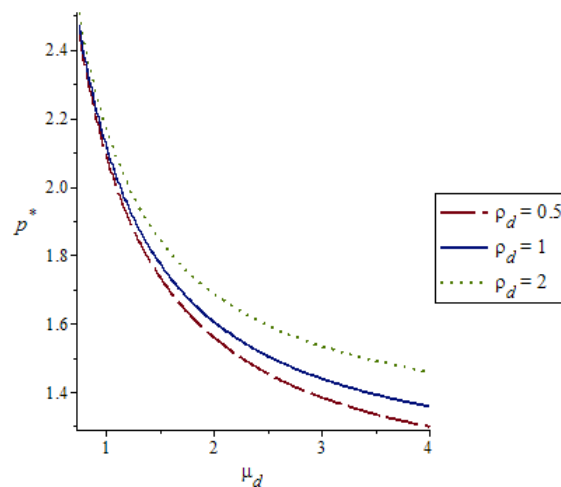


Figure 4. Traffic density isolines and the optimal price under the mixed policy

At the same time, the figure shows that as the offered load increases so does the optimal price. As pointed out earlier, under a given price the allocation (utilization) of capacity on deals would not change as long as the offered load is unchanged.

Clearly, increased arrival and service rates present revenue opportunities for the company through higher turnover. In order to take advantage of the increased turnover, the company needs to lower its price so that more wins are realized and hence deals' share of capacity increases.

It is intuitive and easy to deduce that as the arrival rate for deals increases, the allocation of capacity to deals (*i.e.*, proportion of time the capacity is utilized for deals) will also increase. Similar conclusion is straightforward for the regular jobs. On the other hand, the impact of service rates is not straightforward. Our numerical analysis reveals that the relations between the capacity usage and service rates are not strictly monotonic. Figure 5 demonstrates the effect of the service rate of the deals on the usage of capacity.

When the service rate is too low it is not efficient for the service company to allocate its capacity on the deals unless it charges a high price to the customer. High prices lead to

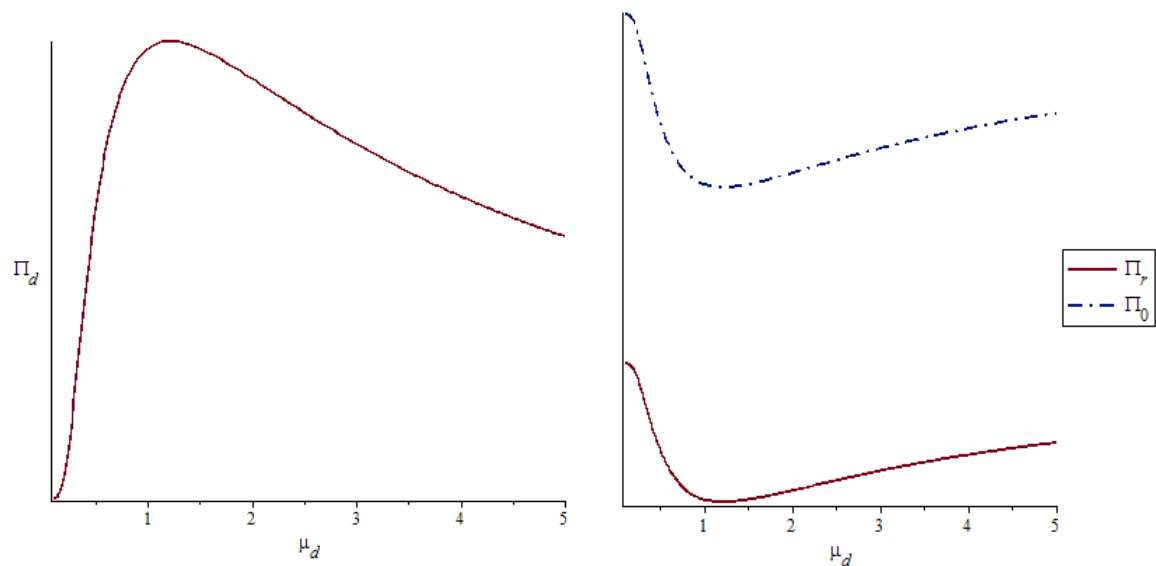


Figure 5. Service rate for the “deals” and the optimal capacity allocations

lower win probability and hence diminished need for capacity use for the deals. The company would rather be profitable by using its capacity on regular jobs with reasonable service rates. As the deals’ service rate increases, the optimal price goes down which boosts the win probability. As a result the share of capacity increases. After a certain

level, the increase in the service rate surpasses the increase in effective arrival rates resulting with reduction in the deals' effective traffic density. Consequently, the server can process more jobs faster and the proportion of capacity needed for the deals decreases.

The effect of the service rate for regular jobs is illustrated in Figure 6. When the service rate is low, each regular job keeps the system busy for a longer time on the average. Since all regular jobs are admitted, more time spent on regular jobs lead to less time on the deals. As the service time increases regular jobs are processed faster and as such, the deals will get more capacity usage opportunity. On the other hand, as explained in Lemma 3, higher regular job service rates result in higher price bids for the deals.

After a certain level, the high service levels justify more use of capacity processing regular jobs due to increased revenue return while the optimal prices and hence, the effective arrival rates for the deals decrease.

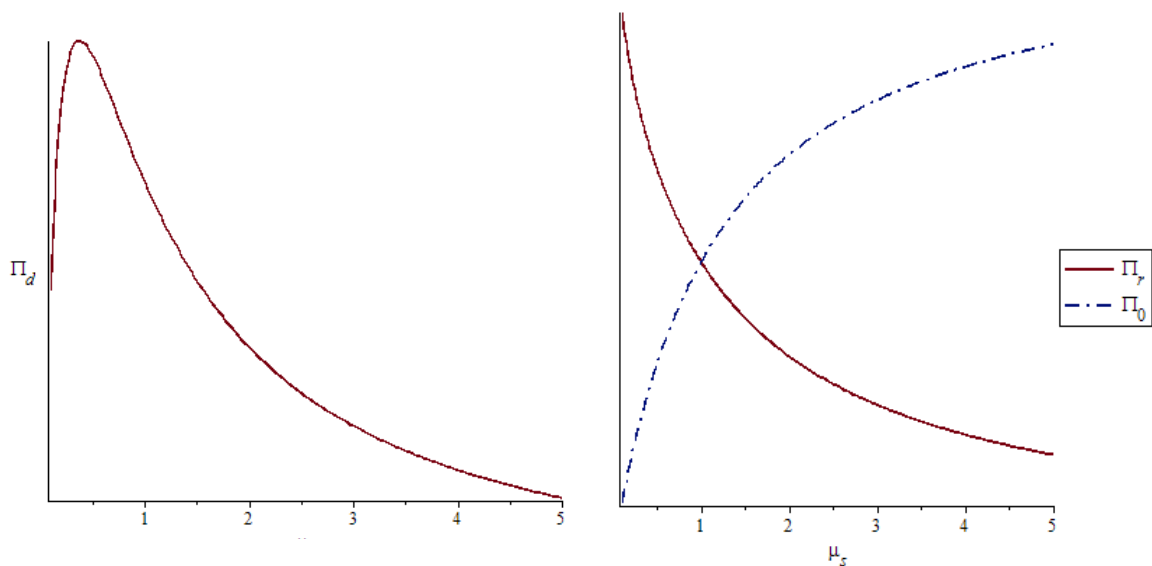


Figure 6. Service rate for the regular jobs and the optimal capacity allocations

3.4 Policy Comparisons

In this section we compare the single job system dedicated to “deals” only and the mixed policy that accepts both deal-type jobs and “regular” jobs. Specifically, we seek for answers for the following questions: 1) How do the optimal pricing policies differ across two policies? 2) Under what conditions the service company prefers one policy over the other? 3) How do the system parameters impact the appeal of either policy for the service provider?

Our first observation is due to the effect of exogenous prices of the regular jobs on the optimal pricing differences.

Proposition 3. *There exists a unique threshold for v , say v_t , such that $p_d^* > p_m^*$ and $G_d^* > G_m^*$ if and only if $v < v_t$. Otherwise, $p_d^* \leq p_m^*$ and $G_d^* \leq G_m^*$. Moreover,*

$$v_t = \frac{\lambda_d e^{-\beta p_d^*}}{\beta \mu_r} = \frac{G_d^*}{\mu_r} \quad (26)$$

Proof. The proof is straightforward from solving $\Delta_p = p_d^* - p_m^*$ and $\Delta_G = G_d^* - G_m^*$ for v . Based on above result, the bidding prices in the deal-only case is higher than the mixed setting if and only if the marginal revenue for regular jobs (v) is below a unique threshold. When the revenue from regular jobs is low, the company is better off by spending more of its time rather on the deals. This compels the company further pull its bidding price down to increase the effective rate of arrivals of the deals under the mixed setting. In the mixed policy, capacity is shared between the deals and the regular jobs. When marginal gain from regular jobs is low, the company adopts lower prices for the deals to stimulate the effective arrivals of the deals and as such allocate more capacity on the deals.

It is interesting to observe that same threshold also applies for the optimal revenue comparison. When v is below the threshold, the company is better off with the deals-only setting. In this case, accepting regular jobs degrades the company profits by allocating low-revenue jobs to capacity which otherwise could be utilized for more profitable deals.

We can deduce from earlier analysis discussed in the previous sections and (26) that the threshold increases in the arrival rate of the deals (λ_d) and the service rate of the deals (μ_d). This implies that as demand base for the deals grows, the deals-only policy becomes more appealing for the company. Same effect is also valid for increasing service rates in which case revenue per unit capacity increases making the deals more profitable for the company. On the other hand, the threshold decreases in price sensitivity (β) and the service rate of the regular jobs (μ_r). As expected, higher price sensitivity results in diminished margins and profitability for the deals making them less appealing for the company. Higher service rate for regular jobs means more revenue per capacity unit allocated to the regular jobs. As such, the faster the company processes the regular jobs the more it will lean towards the mixed policy.

Interestingly, we observe that the arrival rate of regular jobs does not influence the threshold and hence the company's policy preference. Recall that the price of the regular jobs is exogenous and hence is unchanged under changing arrival rates. The marginal revenue contribution of the regular jobs is determined only by price (v) and the service rate (μ_r). Higher service rate implies quicker returns for the regular jobs, which generates more bang-for-the-buck for the company. Consequently, the company's optimal policy selection between the deals-only and mixed settings is independent of λ_r .

However, we note that this does not mean that the sizes of the price/revenue gaps between the two

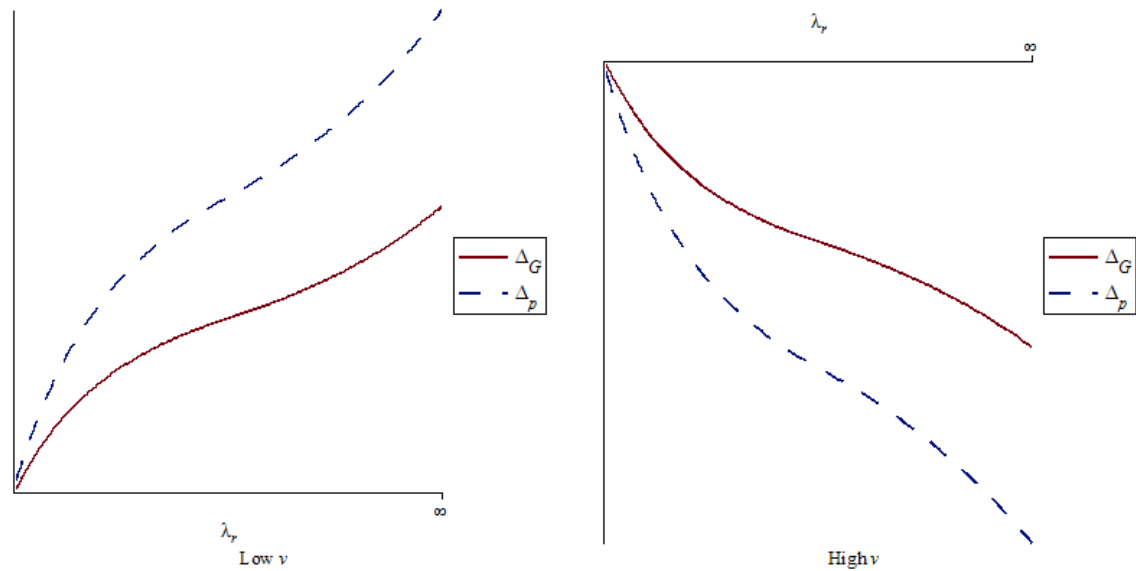


Figure 7. Impact of regular-job arrival rate on price (Δ_p) and revenue (Δ_G) gaps.

policies are independent. When the price of the regular jobs is below the threshold, increased demand volume enhances the positive price/revenue gap between the deals-only and mixed policies. Above the threshold, the positive gap grows with λ_r . Both situations are illustrated in Figure 7.

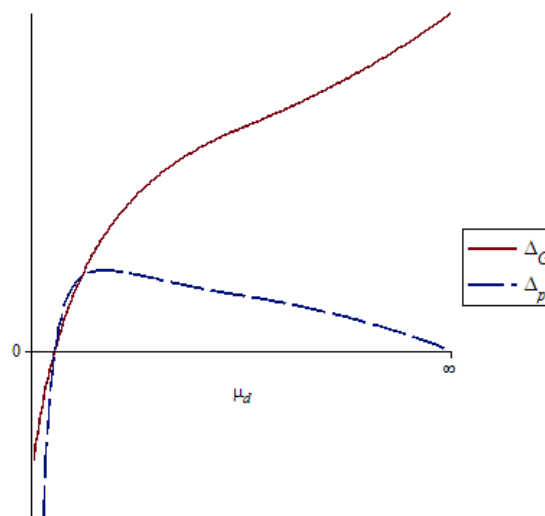


Figure 8. Impact of the deals service rate on the price and revenue gaps between policies.

As implied by the threshold expression given in (26), the difference in revenues strictly increases in μ_d . On the other hand, impact of μ_d on the price difference is somewhat ambiguous as depicted in Figure 8.

It is easy to observe from (26) that the mixed policy is preferable for any values of μ_d , if $v \geq \lambda_d/(\beta\mu_r)$. When v is sufficiently small, the price difference increases in μ_d up to a certain level in the positive half space. Then it starts to close as μ_d grows.

3.5 Commentary and Insights on Order Acceptance

In this Chapter, we investigate the optimal strategic fit for a service company and its pricing policies. We consider two streams of order types; one composed of recurring standard or regular jobs whose prices are pre-determined in the market (*exogenous prices*) and the other involving “deals” that require pricing proposals (*endogenous prices*) from the firm. In the latter case, the probability to secure the order diminishes with the quoted price. The service firm faces two strategic options: 1) allocate capacity to only the deals or 2) admit both order types.

We analyze both deals-only and mix policies separately before comparing them for profitability. We employed Markov Chain based models to derive the optimal pricing decisions for both policies and determine the optimal mean revenues. In the deals-only policy, we show that the optimal revenue increases in demand and service rates while decreasing in price sensitivity of the orders. Interestingly, we observe that the optimal winning probability of the deals is independent of the price sensitivity on the long run under the deals-only setting. This is not true for the mixed policy case where the winning

probability strictly decreases in price sensitivity. We present a detailed analysis for the impact of demand and service rates on the optimal pricing and revenues for both settings.

We compare both policies to seek answers to some questions that are relevant for a service firm's strategic fit and selection of markets to serve. We investigate 1) how the optimal pricing policies differ across two policies; 2) under what conditions the service company prefers one policy over the other; and 3) how the system parameters impact the appeal of either policy for the service provider. We observe that the differences between two strategies in terms of pricing and average gain are in accord. Under any given set of systems parameters, one of the strategies leads to both higher prices and revenues. Typically, the preferable strategy depends on the exogenous price of standard orders and the price sensitivity of the deals.

Typically, the preferable policy depends on the exogenous price of standard orders and the price sensitivity of the deals. Our conclusion is that the threshold values for these two parameters depend primarily on the demand rate of the deals and the service rate of the standard jobs. Higher demand rates for the deals make the deals-only policy more appealing, whereas higher service rates of the regular orders appreciate the mixed policy. Our results indicate that the service rate of the deals become relevant only when the regular orders' price is sufficiently low. In this case, higher service rate for the deals leads to the deals-only policy. Lastly and interestingly, we find out that the demand rate of the standard orders does not impact the firm's preference between the two strategy options. At the end, it all boils down to selecting the strategy that delivers more bang-for-the-buck, that is, more cost effective use of capacity.

We intend to extend our analysis to two relatively more general settings. In the first extension, we consider the allocation of multiple servers (work teams) across order types. Under varying cost and demand conditions we plan to compare cases with dedicated servers and cases where all servers work on all type of orders. For the second extension, we consider the case where the firm admits small jobs into a capacitated queue. In this case, if the company wins a deal it has the option to begin working on the deal immediately regardless of the state of the queue. Clearly, in this case the optimal bidding price is state dependent in terms of whether the bidding is taken place when the company is either idle or performing a regular job with others in the line.

Chapter 4. Order Acceptance with Queuing and State Dependent Pricing

4.1 Overview

In this chapter, we relax the no-queuing assumption of the previous chapter in two ways. First, we extend the space to accommodate to the fact there could be small jobs accumulation (queuing) except when the company is engaged in a big deal. If a big deal is admitted for service the small jobs queue is dispersed by cancelling the small jobs or forwarding them to an outsourced service, in which case a penalty is incurred. Second, because queuing is allowed, price decisions are now state dependent. The price to be bid for a big deal depends on the number of small jobs that the company is currently engaged.

The system will have many states due to accumulation of the queue in opposition to the previous model which embodied only three possible states: idle, busy with regular job and busy with a big deal. The rest of the settings remains pretty much similar to that of the previous model. The regular jobs and big deals are distinguished by their arrival and service rates. In addition, while the price for regular jobs is exogenous, the big deals are won as a function of price bids, whom are endogenous, that is set by the decision maker.

In this setting, regular jobs will be accepted when another regular job is still being processed. The rational being when we have small jobs, typically, it is expected to add the jobs to a waiting line because their service durations are reasonably short. However when a big deal arrives, since the durations are longer short and as such, the capacity requirement, is expected to be much bigger, there is no longer—realistically—the option

to process small jobs nor the option of having them waiting until the big deal is completed. Moreover to compound the problem, mostly when a big deal is won in a typical industry application, all the resources will be taken to service this new request and as consequence if there was a small job in the queue it has to be dismissed.

These considerations are required to extend our space, or possibilities if you like, to encompass the fact that we can have multiple number of small jobs waiting while we process them. In our observations from practice, it is typical the queue do not go beyond 6 to 7 jobs waiting, however we will not impose limits to this number to allow application of the model to other problems with similar nature.

Another consideration is that because the fact the queue has to be dismissed, then implicitly there is a sort of penalty to be paid. It could be as simple as having to return the payment to the customer, or even worse, take additional penalties as they could be stipulated by contracts.

Good will loses happen also in this context, and this penalty cost must consider them, in our particular case, since the firm under analysis was able to subcontract the work or lease rotatable inventory, the penalty translated more than in loss of good will from the customer, in loss of revenue, for the customer the transaction to subcontract or lease was not visible, as a design of the MRO company to maintain the customer relationship smooth in cases like this, of course one of the reasons why the premium for the big deal price on the long run must support this additional cost.

The model in our case, for the big deal portion, considers dynamic pricing, in the form of a premium, to be presented to the customer. If the price is accepted by the customer, the provider will increase its expectation to obtain some gain even with

penalties it had to pay to those customers whose service will not be fulfilled—or as in our case in the form of subcontracted or leasing cost if the inventory must be made available. On the other hand, if the big deal customer balks at the bid, provider will still have the benefit of the gains coming from the small regular jobs.

4.2 Literature Review

Planners with regular supply and demand patterns have numerous methods and techniques they can use, for example, covering even cases where the results are not going to align with physical data. (Isken et al., 2011). If you consider the fact you will provide the big deal as a disruption to the regular jobs being either under process or in the queue. The question is then, how do you go about handling this signal lumpiness to minimize disruption and cycle time and, as a consequence, maximize gross margin and profit.

Specifically cycle time, and its connotations for service level, customer satisfaction, labor cost reduction, and asset management, has a strong interest from multiple parties to gain insights to help companies and their planners to achieve stronger results. (Wu, 2003)

Scheduling and queuing have been used over the years with strong success to solve problems in diverse fields, such as transportation, networking, and telecommunications (Banks, 2010). There is then sufficient work done to help us to understand some basic premises, such as how best identify the objective function.

From this body of work we utilize the concepts of reserved capacity, (Allsop, 1972) defined as capacity above demand, which might be required to honor the booked

commitments, as the idea is that if we have information about variability we can then make some decisions regarding capacity and reserved capacity. Also, by working with their internal processes, the planner can reduce cycle time variability by categorizing activities, products, and customers, etc. Gans *et al.* (Gans & Savin, 2007) propose ways that reduce cycle time variability, eliminating unpredictable variability and gaining efficiency in the allocated resources. Such body of research is represented by models to manage and schedule projects at minimum cost. Ishii (Ishii et al., 2013) is a very recent example in this field.

This line of research looks for optimal results by reformulating the model to encompass parameters to make the variability as predictable as possible. By taking this action if the model does take into consideration that causality, then it will behave as if that variability is predictable. For example in the recent flu epidemic in the US in the winter of 2012-2013, even though it is known that there is an increase in respiratory ailments in winter months, (Blaisdell et al., 2002) if the model does not have ways to acquire information about variability, it behaves as unpredictable, thus some regions within the US ran out of vaccines. In a similar way, the arrival variability and competition response is captured by the price sensitivity parameter in our pricing function. Variability, is one key factor of cycle time and has a direct relationship with reserved capacity.

For our own particular application when variability is substantial, introduction of scheduled capacity is in itself inadvisable from current queue theory insights, unless considerable reserve capacity is available. With the introduction of random unpredicted variability, the problem worsens. Our proposal is to provide what is

the minimum reserved capacity allowing a feasible solution that evolves dynamically and presents a flexible solution. Understanding reserved capacity, either by letting customer go due to the acceptance of a major job and reflecting that as a penalty or compensation for the loss of the small jobs if needed, reserved capacity for us can be expressed in different assets, in our case we will apply the concept further with our implementation in Chapter 5.

For the moment, we go back to our setting, a how much is worth a big deal, given a number of steady small jobs in queue or with a brisk arrival rate, for us to dedicate our entire resources to winning and taking the arriving big deal.

4.3 Model, Assumptions and Methods

A service provider that competes for jobs who have a deterministic schedule and constitute their regular business under a long term contract structure, these firms also have stochastic arrival rates and service requirements driven by incidents. We call them big deals or from non-contract customers, who walk-in with a sizeable revenue generating deal.

We relax the no-queue assumption. This assumption relaxation is motivated to account for different job requirements coming from the big jobs—or namely “walk-ins”. Then it is assumed that the company can win and execute one job/project at a time and as such, does not bid for other jobs while they are under a contract with one big job. Other times, when the company is not under any contract it bids prices for arriving job requests. The company needs to decide on the bidding prices for the big deals. This bidding price must take into consideration what the expected queue of small

jobs is and what if any penalty must be paid so when winning the big job the service provider will make money.

The service times are described as a whole number of days. The number is used to calculate the proportions of customers within buckets of different cycle times that are big jobs—which is one by definition—and those of the small jobs with no capacity reservation. Our model will use probability theory, under Markovian settings, to quantify variability of cycle times and from there translate into the capacity requirements. In the first part of our analysis, we assume that the inter-arrival times and the service durations follow exponential distributions with rates λ_{small} , λ_{big} for the demand, and μ_{small} , μ_{big} for the service time respectively. We continue employing this assumption.

We recall that when establishing the framework on the first model in the previous chapter, we indicate this business environment for service companies such as rentals companies considered by Gans and Savin(2007). Particular consideration for the model is to recognize that there are significantly different service times. Other complexities of real life operations in a given company will be set aside for the moment, as right now our interest is mostly to extend the basic model for small vs. big jobs.

So initially we will start with the unrealistic assumption that there is unlimited capacity in the shop so all demand can be met in small jobs, and all capacity is taken when a big job has been won. Finally, with respect customers arriving for small or big job we will assume that their cycle time is essentially different, independent and identically distributed.

Also, for the purposes of this model we will consider a flow of small jobs that come from regular (standard) customers who are willing to join a queue as needed.

Finally every day also—whole units of time—for that matter, some of these engagements are finished successfully leaving the system; we assume the system behaves holistically within a steady state determined probability distribution.

Later in Chapter 6, we will comment on how our results help us extend the model to a more realistic scenario such as capacity bounds and re-scheduling if possible when no capacity is available. In most of the reported work in the literature the models seem to be built in two ways, by leaving holes in the agenda or by adding extra capacity (Gallivan *et al.* (Gallivan, Utley, Treasure, & Valencia, 2002), keeping in mind the similarities on equipment, parts, failure modes, are opportunities to relax the independence assumption.

Continuing with our model definition, small jobs needs and characteristics are assumed to be homogeneous, which means that their processes take similar cycle times and probability distribution even when coming from different customers or sources. Cycle times are assumed to be independent of one another and independent of the number of units under process in the system, and the occupied capacity. Another assumption is that there is no re-scheduling and preemption in processing. Jobs that are already initiated must be completed. However, the queues can be dispersed in return for a penalty that is proportional to the size of the queue.

When a customer big job is won, the assumption is that no small jobs will be interrupted; only the next available slot will be allocated to the big job if price tender—price premium as defined—is accepted. Another assumption is that next job to be executed after the big deal will be the next job on the arrived into a new schedule program that is new small jobs queue will re-start.

4.3.1 Notation

In these early stages, our model is fairly unsophisticated, just enough to understand the dynamics between what we consider the key factors of: regular jobs, on the other hand the sparsed events big deals.

We need the following parameters and assumptions:

1. Small job event arrival rate
2. Big job—lumpy event—as arrival data
3. Server does not experience any break downs
4. Labor/skill is available
5. Theree is a penalty for cancelling each waiting small job, U
6. Revenue, r , for each small job is exogenous
7. Premium to be bid for the big deal is the decision variable denoted by \mathcal{P}_x where x is the state related to the number of small jobs waiting at the time of the big job arrival.

All demand, once taken, will then use the equipment and labor for a number of complete time units—no fractions at this time although if needed the time fraction is a natural extension—the demand is associated with a customer. So all variables have a value when a demand is present.

We describe π_i as the probability of the state space. Arrival rates are random independent variables with a Poisson distribution for the equipment taken by the customer for a number of units of time. Later build up on this variable could be done, to differentiate the types of equipment, for the moment it represents the same equipment.

For the moment let's call this equipment the server, and we will have one unit available for the model, this equipment can work with small jobs (s) and lumpy—random, non-scheduled—demand or big jobs (B).

The small jobs (s) have a different smaller service time than big jobs (B) that take longer to service as previously stated.

Arrival rates for the s jobs are much more frequent than arrival rates for jobs B. There could be multiple s jobs waiting, but no B jobs. Once a B job arrives, it has higher priority than the s job, no s jobs can wait in queue meanwhile the B job is completed.

4.3.2 Model Description and Analysis

Let's say each of the arrival events are described as a markov chain with service μ_s, μ_B and λ_s, λ_B rate for the arrivals of small jobs and big jobs respectively. The probability π_i denotes the steady state probability of the system by the existing number of customers' s and B at the system and as later will be shown is always a function of the regular price r .

As mentioned B customers have higher priority over regular customers and the system can only have one B customer at a given state/time. Any B customers that find the server busy walk away. The system could have as many regular customers as they arrive, who can also wait while other jobs are processed.

Figure 9 represents the model schematic of the state spaces. Matrix given in (27) has the state space variables for the to calculate the transition probabilities for the model.

$$Q = \begin{pmatrix} 0 & \lambda_B e^{-\mathcal{P}\beta r} & \lambda_s & 0 & 0 & \dots \\ \mu_B & 0 & 0 & 0 & 0 & \dots \\ 0 & \lambda_B e^{-\mathcal{P}\beta r} & 0 & \lambda_s & 0 & 0 \dots \\ \dots & \lambda_B e^{-\mathcal{P}\beta r} & \mu_s & 0 & \lambda_s & 0 \\ \dots & \lambda_B e^{-\mathcal{P}\beta r} & 0 & \mu_s & 0 & \lambda_s \dots \\ & \vdots & 0 & 0 & \mu_s & 0 \\ & & \vdots & 0 & 0 & \mu_s \\ & & & \vdots & 0 & \vdots \\ & & & & \vdots & \ddots \end{pmatrix} \quad (27)$$

In the matrix Q , we express the effective transition rates. Let's bring back now the pricing we defined for our first model in Chapter 3, namely the decision on price to bid on arriving job requests. The winning probability of a job proposal is still modeled as $e^{-\beta \mathcal{P}}$, where β is a constant that denotes the price sensitivity of the arriving job requests. Now when the company bids for big deals it uses a *premium* \mathcal{P} as in lieu of price, which, in this case depends on the state of the system.

Clearly, the winning probability decreases in r and \mathcal{P} , and as such the effective arrival rate for wins is $\lambda_B e^{-\mathcal{P}\beta}$ for big deals (B). We recall that as customer's price sensitivity β increases, the winning probability for the service provider diminishes at the same price level. For small jobs (s) even though there is also a correlation between price and customer arrival, the work continues in the understanding that r is exogenous to the company driven by the market, so the arrival rates λ_s already have some form of erosion, but from the model stand point it is already considered in the variable.

The state space will be $S_{i,j} = (i, j)$, with probability $P_{i,j}$ for the transition from state i to state j , then we have matrix Q defined (27), from the depicted model in figure 9 starting with an infinitesimal generator matrix Q of dimensions $n \times n$. Each node connection will be represented by row with the column index the number of the node to where the transaction connects following the diagram in Figure 9.

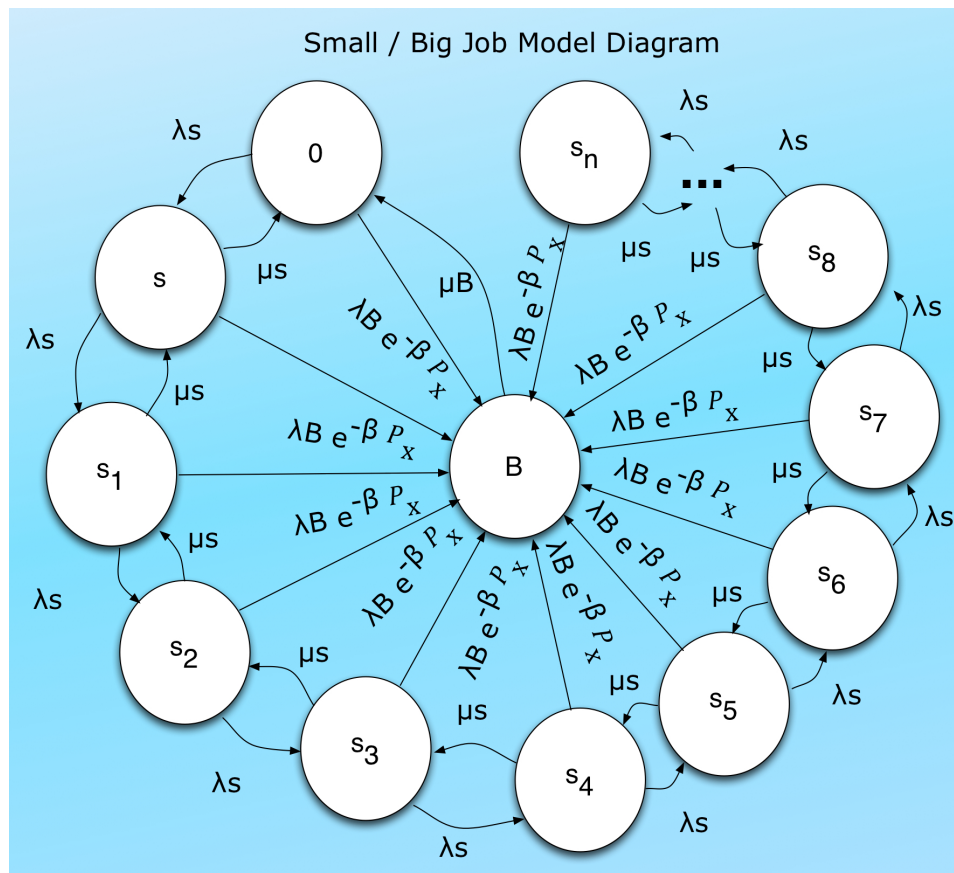


Figure 9. The small and big job policy model

The connections represented follow the markovian assumption the system steady state probabilities. These definitions help to understand the state space for the probability distribution of the random variable $X(t)$ over the state space S , since the markov variable is $\{X(t), t \in T\}$ with T in the Reals. Let's assume $X(t)$ is memory less, irreducible and time homogeneous, so when transition out of state i happens, with probability $P_{i,j}$ reaching state j , defining then for $i \neq j, i, j \in S$,

$$Pr(X(t + dt) = j | X(t) = i) = q_{i,j}dt + o(dt) \quad (28)$$

Now by applying the theorem for the Markov steady state probability distribution,

$$\lim_{t \rightarrow \infty} Pr(X(t) = x_k | X(0) = x_0) = \pi_k \quad (29)$$

and then $\{\pi_k, x_k \in S\}$ exists, thus k will represent both the number of transition probabilities and 2 regions within the space; *idle region* on State (i) where $i = 0$, no customers in the server, *s region* State (i), where $i = 2, 3, 4, \dots, n$ small customers only (s), *B region* State (i), where $i = 1$, with one Big deal under service and no big or small customers waiting, $n = i - 1$ also represents s small customers waiting in each state $i = 2, 3, \dots, s$, plus the small customers in the system in process or waiting, and finally $n = 1$, if a Big deal customer is on the system,. See figure 9 again for a diagram representing the model.

Since both s customers and B customers generate revenue r_s and \mathcal{P}_i where $i = 2, 3, \dots, s$, respectively, we define $\boldsymbol{\pi}_{i \text{ for } i = 1}$ for B and $\boldsymbol{\pi}_i$ for s customers. Continuing from our matrix Q definition, let's define the diagonal elements to ensure the sum of elements is zero, to represent the fact that only one Big job can be processed at the server at any given time.

$$q_{i,i} = -\sum_{j \in S} q_{i,j} \quad (30)$$

To maintain the global balance on the model equations we will equate the flows into a node to the flows out of that node.

$$\boldsymbol{\pi}_i \times \sum_{x_i \in S, j \neq i} q_{i,j} = \sum_{x_j \in S, j \neq i} (\boldsymbol{\pi}_j \times q_{j,i}) \quad (31)$$

and rearranging terms we have this matrix equation. Note the bold typography denotes matrices

$$\boldsymbol{\pi} \boldsymbol{Q} = \mathbf{0} \quad (32)$$

The overall probability must be 1 by definition; $\sum_{x_i \in S} \pi_i = 1$. If needed any textbook on Matrix Analysis Queuing will help the reader, see for example (Hillston, 2005), this brief summary was added for completeness.

Taking the equations stated above into the $Ax = b$ format defines the solution for the π_n probability unknowns. First we move \mathbf{Q} to the first position of the equation by transposing it, we will denote \mathbf{Q}^T its transpose matrix. We now normalize by replacing the last row with ones, and eliminate the redundant global balance equation; the matrix will be designated \mathbf{Q}_n^T , finally the solution vector denoted \mathbf{e}_n will be a column vector with all zeros and, in the last row of the column a one. This can be solved with a simple linear solution in vector π_n extracted from the solution of the linear equation set in the form of $Ax = b$ as stated few lines above.

$$\mathbf{Q}_n^T \pi = \mathbf{e}_n \quad (33)$$

With our transition probability in vector $\boldsymbol{\pi}_n$ —recall that probability $\boldsymbol{\pi}_n(r)$ is a function of the premium price, although some times not explicitly mentioned for simplicity on the notation—let's compute the steady state probabilities for *idle*, *s*, and *B*. with *Idle region* defined as probability π_0 .

Thus probability of *s region* can be then defined as,

$$\pi_s = \sum_{i=2}^n \pi_i \text{ for } i = 2, 3, 4, \dots, n, \quad (34)$$

while for the *B region* is

$$\pi_B = \pi_i \text{ for } i = 1 \quad (35)$$

Then, the busy probability will be $1 - \pi_0$ or,

$$1 - \pi_0 = \pi_s + \pi_B \quad (36)$$

and thus the probability to be able to accept a big deals will then be the sum of $\pi_0 + \pi_s$, that is,

$$1 - \pi_B = \pi_0 + \pi_s \quad (37)$$

while small jobs can be received in any other state, with exception when there is a big deal in the server. All of these probabilities are in fact functions of the premium price for big deals, with the following notation, $\pi_B(\mathcal{P}_x), \pi_s(\mathcal{P}_x)$ for $x = 1, 2, 3 \dots$ and $\pi_0(\mathcal{P}_x), \pi_1(\mathcal{P}_x), \pi_2(\mathcal{P}_x), \dots, \pi_n(\mathcal{P}_x), \dots, \pi_\infty(\mathcal{P}_x)$, for simplicity is that we are calling them π_B, π_s and $\pi_0, \pi_1, \pi_2 \dots \pi_\infty$.

With the definition of the winning probability, $\lambda_B e^{-\mathcal{P}\beta r}$ is the effective arrival rate for big deals (B). We recall again that as customer's price sensitivity β increases, the winning probability for the service provider diminishes at the same price level. For the regular jobs the price is denoted by r and is the company's regular price bid, as such the effective arrival rate for wins is λ_s for small jobs.

Following observations into the problem, we have r as price set by a market thus for the effects of this work a given parameter. For us the premium \mathcal{P} is an internal decision variable and the one the decision maker would like to obtain given certain factors and conditions in both the market price and the business dynamic expressed by factors such as winning deals arriving, the service rate for them, and the arriving regular customers and their corresponding service rate.

Finally a penalty U is paid each time a customer was on the queue and then is rejected once big job arrives and is accepted.

From (33) we now obtained the π_i distribution for all the state space, then

$$L_q = \sum_{i=1}^n i\pi_i \quad (38)$$

To find the price at which the decision maker wish to enter the market, with the acknowledge there could be penalties (U) for all those regular customers already in queue, when the big deal arrives. Let's then proceed to define our Gain function as follows:

$$G_{s\&B} = \sum_{x=0}^n (\mathcal{P}_x - xU)\lambda_B e^{-\beta \mathcal{P}_x} [\pi_x(\mathcal{P})] + r\lambda_s(1 - \pi_1(\mathcal{P})) \text{ for } x \neq 1. \quad (39)$$

Then the first derivative of the gain for the premium variable with respect of Gain is defined as the optimal value of the function, with a unique maximizer \mathcal{P}^* at:

$$G_{s\&B}^* \text{ when } \frac{dG}{d\mathcal{P}} = 0 \quad (40)$$

We now follow the process to make partial derivatives for each element of the eq. (39).

$$\begin{aligned} \frac{\partial}{\partial \mathcal{P}} G_{s\&B} &= \mathcal{P}_0 \lambda_B e^{-\beta \mathcal{P}_0} [\pi_0(\mathcal{P})] + \sum_{x=1}^n (\mathcal{P}_x - xU)\lambda_B e^{-\beta \mathcal{P}_x} [\pi_{x+1}(\mathcal{P})] \\ &+ r\lambda_s(1 - \pi_1) = 0. \text{ Where } x = 1..n \end{aligned} \quad (41)$$

Doing simultaneous equations of the following partial derivatives

$$\begin{aligned} \frac{\partial \mathcal{P}_x \lambda_B e^{-\beta \mathcal{P}_0} [\pi_y(\mathcal{P})]}{\partial \mathcal{P}_x} &= ((\mathcal{P}_x - xU)\lambda_B e^{-\beta \mathcal{P}_x}) [\pi_y(\mathcal{P})]' + [\pi_y(\mathcal{P})] ((\mathcal{P}_x - xU)\lambda_B e^{-\beta \mathcal{P}_x}') + \\ &((\mathcal{P}_x - xU) [\pi_y(\mathcal{P})]) \lambda_B e^{-\beta \mathcal{P}_x}' + (\lambda_B e^{-\beta \mathcal{P}_x} (\mathcal{P}_x - xU)) [\pi_y(\mathcal{P})]' + \\ &(\lambda_B e^{-\beta \mathcal{P}_x} [\pi_y(\mathcal{P})]) (\mathcal{P}_x - xU)' + ((\mathcal{P}_x - xU)\lambda_B e^{-\beta \mathcal{P}_x}) [\pi_y(\mathcal{P})]' \end{aligned} \quad (42)$$

for all $y \neq 0$, Then

$$\frac{\partial (r\lambda_s(1 - \pi_1(\mathcal{P})))}{\partial \mathcal{P}_x} = -r\lambda_s [\pi_1(\mathcal{P})]' \text{ for all } x \quad (43)$$

Proposition 4. *There are solutions for each of the equations on the partial derivatives indicated*

$$\frac{\partial}{\partial \mathcal{P}_x} G_{s\&B} = 0 \text{ where } x = 1,2,3, n \text{ has an optimal } \mathcal{P}_x^* \quad (44)$$

Such as $\mathcal{P}_x^* = \frac{-\log_e(\beta\alpha^x)}{\beta}$ with $|\beta\alpha^x| < 1$ and $\alpha = \beta \frac{\lambda_B}{\mu_B - \lambda_B e}$

Proof. The solution set will be obtained in the form. For all states

$$\mathcal{P}_x = \frac{\log(|\beta\alpha^x|)}{\beta} \text{ where } x = 0, 1, 2, \dots \quad (45)$$

And solution for state 1, the one from the definition of our state space contains the big deals, thus is $\mathcal{P}_1 = \mathcal{P}_B$, that is the premium when in state 1 (see model diagram in fig. 9) the value of pricing will be recommended. Although it has been defined no other big deals can be accepted once the big deal is in process, still to allow generalization of the model to other situations we will include this state, also will help us to confirm eq. (40).

Lets start our process by using the flow equations by node we know each equation has a solution in the form stated in eq. (33) and each probability in the node relates with $P_n = \alpha^n P_0$, so from the state = 0 we have

$$(\lambda_s + \lambda_B e^{-\beta P_0})\pi_0 = \mu_B \pi_1 + \mu_s \pi_2 \quad (46)$$

From the flow equation deploying the variables and rearranging we obtain

$$\mathcal{P}_0 = \left| \frac{-\log(|\beta\alpha^0|)}{\beta} \right| \text{ where } \alpha = 1 \text{ since } P_0 = \alpha^0 P_0 \quad (47)$$

Now from the state = 1 we also get the equation for $\mathcal{P}_1 = \mathcal{P}_B$ with $P_1 = \alpha^1 P_0$

$$\mu_B \pi_1 = \lambda_B e^{-\beta P_0} \pi_0 + \lambda_B e^{-\beta P_1} \pi_2 + \dots + \lambda_B e^{-\beta P_n} \pi_{n+1} \quad (48)$$

$$\frac{\mu_B}{\lambda_B} - e^{-\beta P_0} \alpha^{-1} = e^{-\beta P_1} \alpha + \dots + e^{-\beta P_n} \alpha^n \quad (49)$$

The general theorem for Dirichlet series definition, see Clark(2010) gets the appropriate solution to the multiplicative factor by applying the following identify

$$e^{-\beta P_1} \alpha + \dots + e^{-\beta P_n} \alpha^n = \sum_{n=1}^{\infty} e^{-\beta P_n} \alpha^n \text{ for } |\alpha| < 1 \text{ and } Re(\beta) \quad (50)$$

$$\lim_{k \rightarrow \infty} \frac{\log(|\alpha^k|)}{\beta \mathcal{P}_{k+1}} = 1 \quad (51)$$

As k goes to infinity in the long run, the function will take the value of one.

From eq. (50) then by applying Dirichlet theorem in eq. (46) we can now go for the solution

$$\frac{\mu_B}{\lambda_B} - e^{-\beta \mathcal{P}_0} \alpha^{-1} = e \quad (52)$$

By combining eq. (52) with eq. (47) we now can get the value of the multiplier α

$$\frac{\mu_B}{\lambda_B} - e^{-\beta \left| \frac{-\log(|\beta \alpha^0|)}{\beta} \right|} \alpha^{-1} = e \text{ with } \alpha = \beta \frac{\lambda_B}{\mu_B - \lambda_B e} \quad (53)$$

Since from comparison to eq. (45) by replacing values in eq. (53). We now have the value for \mathcal{P}_B for $|\beta \alpha^1| < 1$

◇

Lemma 6. The equations for nodes $n > 2$ are $\mathcal{P}_x^* = \frac{-\log_e(|\beta \alpha^{x+1}|)}{\beta}$ with $|\beta \alpha^{x+1}| \leq$

$1 \in \mathcal{R}$

Proof. From the flow equations for state = 2, please note from now on we will be using the index number of the variable \mathcal{P}_x as the number of customers in the queue to simplify notation, so from state = 2 on, $x = \text{state} - 1$, so when \mathcal{P}_1 , that would be the price for a state where there is 1 customer in the system, \mathcal{P}_2 , when there are two and so on.

So for state = 2, we then have \mathcal{P}_1 and the following from the flow equations

$$(\mu_s + \lambda_s + \lambda_B e^{-\beta \mathcal{P}_1}) \pi_2 = \lambda_s \pi_0 + \mu_s \pi_3 \quad (54)$$

$$\mathcal{P}_1 = \left| \frac{-\log \frac{\mu_s \alpha^3 - (\mu_s + \lambda_s) \alpha^2 + \lambda_s}{\lambda_B \alpha^2}}{\beta} \right| \quad (55)$$

$$\text{since previous state is } n = 1, \frac{\mu_s \alpha^3 - (\mu_s + \lambda_s) \alpha^2 + \lambda_s}{\lambda_B \alpha^2} = \beta \left(\beta \frac{\lambda_B}{\mu_B - \lambda_B e} \right)^2 \quad (56)$$

From here on nodes $n = 3 \dots n$ are all the same with the following analysis,

$$(\mu_s + \lambda_s + \lambda_B e^{-\beta \mathcal{P}_2})\pi_3 = \lambda_s \pi_2 + \mu_s \pi_4 \quad (57)$$

$$\mathcal{P}_2 = \left| \frac{-\log \frac{\mu_s \alpha^2 - (\mu_s + \lambda_s) \alpha + \lambda_s}{\lambda_B \alpha}}{\beta} \right| \quad (58)$$

Since previous in this case previous state is 3 then $x = 2$, $\mathcal{P}_{n-1=2}$

$$\frac{\mu_s \alpha^2 - (\mu_s + \lambda_s) \alpha + \lambda_s}{\lambda_B \alpha} = \beta \left(\beta \frac{\lambda_B}{\mu_B - \lambda_B e} \right)^3 \quad (59)$$

For node = 4, we then have $x = 3$, we take the flow equations too and obtain the solutions following the same rational for the indexes

$$\mathcal{P}_3 = \left| \frac{-\log \frac{\mu_s \alpha^2 + (\mu_s + \lambda_s) \alpha + \lambda_s}{\lambda_B \alpha}}{\beta} \right| \quad (60)$$

With the multiplying factor defined as

$$\frac{\mu_s \alpha^2 + (\mu_s + \lambda_s) \alpha + \lambda_s}{\lambda_B \alpha} = \beta \left(\beta \frac{\lambda_B}{\mu_B - \lambda_B e} \right)^4 \quad (61)$$

With Node $n = 4$ and those previous observations now we have the solution for all the remaining eq. for $x = 1, 2, 3, 4, \dots$

$$\mathcal{P}_x^* = \frac{-\log_e(\beta |\alpha^{x+1}|)}{\beta} \quad (62)$$

$$\alpha^{x+1} = \left(\beta \frac{\lambda_B}{\mu_B - \lambda_B e} \right)^{x+1} \quad (63)$$

Where α^{x+1} is defined by eq. (63) again with $x = 1, 2, 3, 4, \dots$

Replacing \mathcal{P}_x^* in eq. (44) $\frac{\partial}{\partial \mathcal{P}_x} G_{s\&B} = 0$ for $|\alpha| < 1$ and $Re(\beta)$, when $x \rightarrow \infty$, the proof is complete.

◇

4.4 Model Evaluation and Numerical Analysis

In our numerical analysis, we vary the parameters for all given variables U from 100 to 500, small jobs arrival rates 5 to 40, big jobs arrival rates 1 to 3, small jobs service rates from 10 to 40, and exogenous price r from 100 to 500. Big jobs service rate is kept constant at 1. With that we created a synthetic data set to evaluate the correspondent premiums and from those observations look for policies to apply to the decision-making problem.

Figure 10 helps us to confirm that the faster you process small jobs, the fewer number of customers will be affected once a big job shows up, which drives the overall tendency of the pricing. The case we are using, leads to a queue of about 10 customers, when the service capacity for small jobs is low, and will decrease as more service capacity for small jobs is added. This case also will have a low payment $r = 100$, while the penalty is 5 times the standard price.

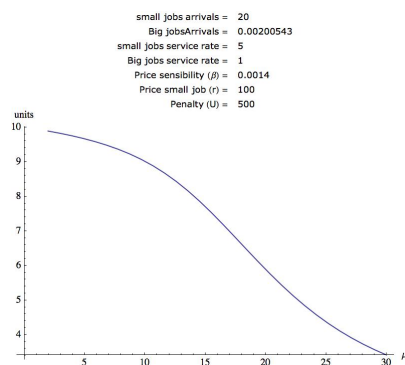


Figure 10. Expected queue as function of the small jobs service

The idea is to set the model in a extreme position, that yet could be typical in some businesses where the cancelation of a job might not be only return the money to the

customer, but imply expenses already made and the fulfillment of contractual obligations such as outsourcing or legal penalties.

It is worth to note on eq. 53 the factor for the pricing grows with the arrival rates on the big deals, regulated by the sensitivity and with the resolution of the service rate for big jobs limiting the overall delivery $\beta \frac{\lambda_B}{\mu_B - \lambda_B e}$ thus more, small business arrivals and service capacity only playing the role in the gain function when no big deal is around, which is limited by e from the $\lim_{k \rightarrow \infty} \frac{\log(|\alpha^k|)}{\beta P_{k+1}} = 1$

We will standardize our big deal prices by dividing them for the price when the server is idle, that is $\frac{P_x}{P_0}$ where $x = 1, 2, 3, 4, \dots$. The next observation is depicted on Figure 11 and does show a significant trait

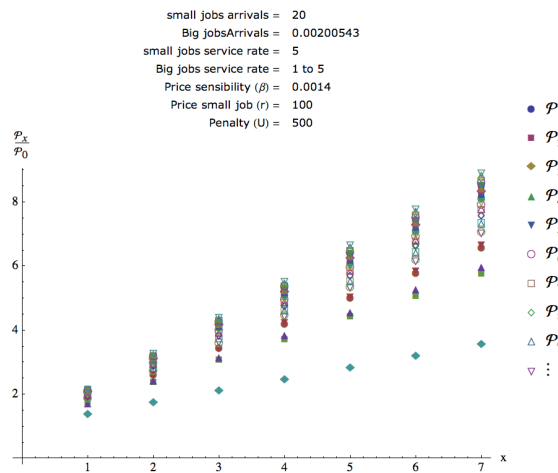


Figure 11. Pricing by state (first 7 states only for clarity)

First, on the state 0 when the plant is idle, the suggested premium, keeps up with the differential in queue size showed in the 10 as expected, the driver of this variation is the server for big deals going from low capacity in 1 to high capacity in 5, suggesting the price could go up and down just with the capacity availability, that is an important feature

as if there is available capacity you would like to account for that by offering an attractive price. Then after state 2—that is, one or more customers in queue—price has a power factor at play depending upon the arrival rates, and the service rate, for small jobs also, that will determine the size of the queue and in turn that will drive the amount and size of the penalties to contemplate within the pricing, in other words, premium will be at par after certain initial value with what is expected to be the penalty of the queue.

Even each of the prices depicted on the figure has a range to account for the variation on the big deal service capacity, still overall the tendency of the prices goes up as the number of customers in queue increases represented by x index and axes in the chart.

Each of the price points can be analyzed to understand more about their sensibility, since it is important to understand pricing in terms of available capacity. For example in Figure 12 we now set every variable constant except for μ_B , that is the big

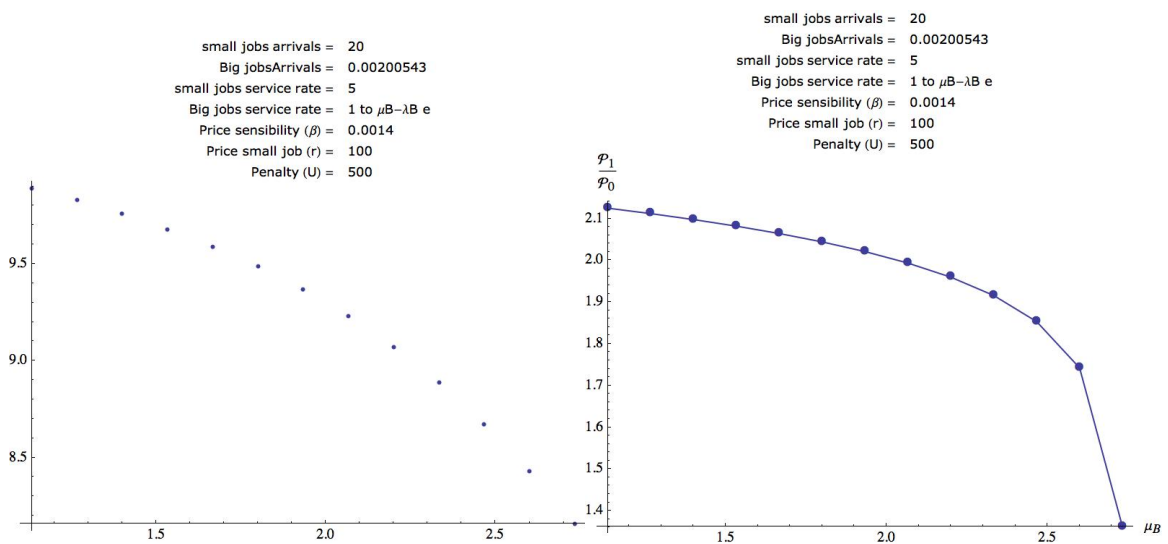


Figure 12 a and b. Pricing evolution with one customer as the big deal service level increases

deal service level, as it starts to go from 1 to 5—in relative terms this is a huge capacity for big deals, there is an interesting play, remember again, we set up the policy no more than one big deal. This in effect since the driver for this price is $\frac{\lambda_B}{\mu_B - \lambda_B e}$, so when capacity for the service level is low, yet more than the necessary for some big deals with probability $e^{-\beta p_1}$, even with that service levels, some deals get processed, allowing small jobs to be taken and delivered as well. Then the bidding price will be high enough as to account for the queue on small jobs.

Playing with the factors coming from the service rate μ_b , when service rate is equal to the arrival of the big deals λ_B times e , there is no more capacity left, also the queue is the lowest (see Figure 12b) and you get then into the optimal price for this state. but then in the long run, if capacity of big deal continues to increase, it will support an increase in the big deal prices as well. It is only when at capacity and in relation to the small job business dynamics the system can support lower the prices for big deals.

This insight suggest that the policy will do good for a certain established capacity, but once that capacity is full, premiums must account for the small job business dynamic, either by the brisky pace of the delivery or by the extension of the queue. Both cases will call for higher price quotes for the big deal pricing.

The structure of the premium for state 1 is quite interesting in that the capacity to service is increasing and in this state we have the expectation to have at least one small job in the queue. Since the server is busy, premium first goes down up to around 9. By inspection premium 1 defined by eq. 78, we notice that the actual driver is very close to

$1/\beta$, thus making in effect the qualifier for the big job arrivals close to 1. The model was running for $n = 10$.

We note that although more customers will be in queue on the higher states, the likelihood of those is remote. The need to set up the premium could be so high to overcome U penalties such that the probability of winning the big deal approaches to zero.

Figure 11 gives again the idea on how important is that first customer in place and the premium we can charge for the next big job. The gain curve follows of that first premium.

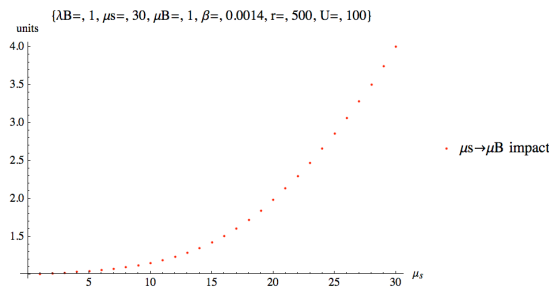


Figure 13. Queue impact by switching small job to big job service rate.



Figure 14. Optimal derivatives vector evaluated on the optimal price.

The model is meant to work under the assumption that once a customer comes with a big job any other job in queue will be dropped. Figure 13 shows the effect on what the difference between the faster small job service rate will be to the big job rate if the whole system is now processing at the lower rate. It does confirm the fact that the faster your

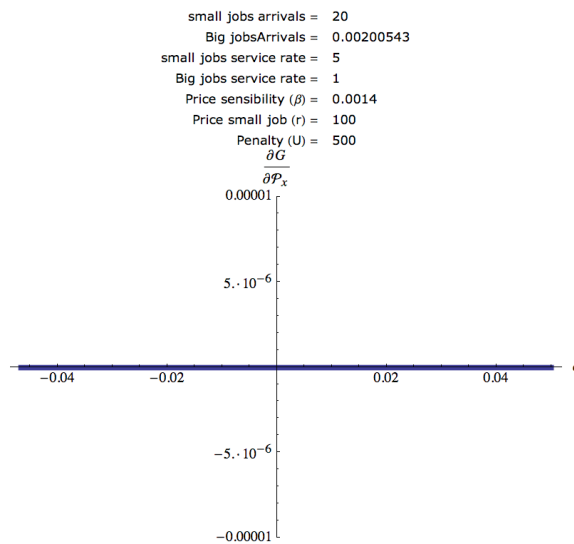


Figure 15. Gain function derivative sensitivity when $n = \infty$ at 0, this figure example has $n = 14$.

small job service, the more noticeable will be to customers once you accept the big job.

Figure 14 above shows the solution vector for each partial derivative of G when it is optimal. And it does show a strong fit for the established conditions in eq. (44)

$\frac{\partial}{\partial P_x} G_{S\&B} = 0$ for $|\alpha| < 1$ and $Re(\beta)$, when $x \rightarrow \infty$, while figure 15, shows a sensitivity

analysis for the vector when all the variables of eq. (44) were evaluated over the range +/-

5% on the setting of the values for optimal G.

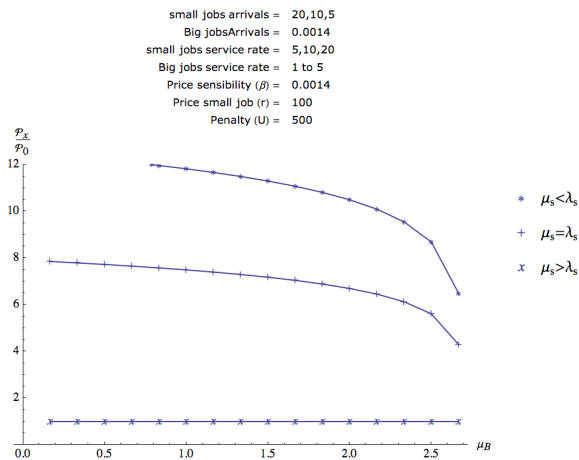


Figure 16. Price vs big deal service rate when there is 1 small customer $x=1$

When the capacity of the service rate for big deals varies, there are 3 major regions that we can observed, with the limit set on $\lim_{k \rightarrow \infty} \frac{\log(|\alpha^k|)}{\beta \mathcal{P}_{k+1}} = 1$, which in turn influences the denominator for the pricing for the multiplicative factor $\frac{\lambda_B}{\mu_B - \lambda_B e}$.

- i) When the small jobs service rate is much bigger than the small jobs arrival rate, queue for the small jobs does not built up and thus the need for increased big deal prices goes away, as depicted in Figure 16, so the pricing remains flat despite the increased capacity as no benefit is derived on processing faster the big deals..
- ii) When small job service rate is equal to their arrival rate, there there is a small build up, on the queue, and as the capacity increases, the fact the server can finish the big deals, and work on either generating $r \lambda_s$ revenue or preventing $x U$ penalties, makes the pricing come down until the point big deal service rate is equal to the limited capacity on the big deal arrival rate set also by

$$\frac{\lambda_B}{\mu_B - \lambda_B e}$$

- iii) When small jobs service rate smaller than their arrival rate, the queue increases driving up the prices further

The final observation is the capacity of the server, just to confirm what was commented on figure 13. The capacity gets constrained when $\mu_s \leq \lambda_s$, but the effect that $\mu_B \leq \lambda_B$ weights in terms of the overall capacity is negligible, as per the policy we have that once a big deal is taken, no other big deals, no other small jobs can be taken.

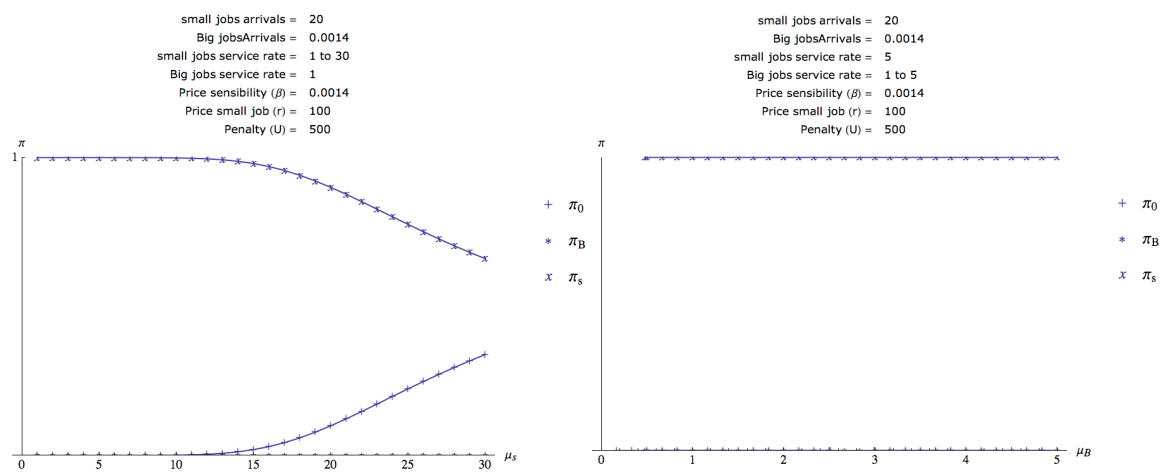


Figure 17 a and b. Capacity vs service rates

4.5 Discussion on the Order Acceptance with Differentiated Services Model

There are insights we like to highlight from the model.

- 1) Differentiated service, adds riches to the model to help guide the decision maker with business conditions on setting prices for big deals. The major driver for the price setting is the sensitivity β , that drives the initial consideration for big deal price, under the $e^{-\beta P_0}$ winning probability. In our model we employ a pricing model that maps prices into winning probabilities. This is a reasonable and common approach in practice, competition is already set up in the model from the interpretation as the possibility that the customer does not have a better

reservation price which could be a standing bid from competition. Losing an arriving job is akin to the job request being won by the competition. A discussion and review of such models and other pricing approaches are reviewed by Bitran and Caldentey (2003). Some examples using similar mappings in different contexts include Bichler *et. al.* (2002)., Gallego and van Rysin (1994), You and Wu (2007), Pachon *et. al.* (2007), and Wen and Chen (2010). Setting the initial price with parameter β , will then be both addressing the competition and the wining probability.

2) Capacity set up for small jobs will improve pricing offering for big jobs. $\mu_s \geq \lambda_s$

3) Big job service capacity will improve pricing until capacity equals the

$$\lim_{k \rightarrow \infty} \frac{\log(|\alpha^k|)}{\beta \mathcal{P}_{k+1}} = 1 \text{ or expressed in terms of service rate } \mu_B \leq e \lambda_B$$

4) Pricing will be optimal and minimum in the same limit for any given condition at

$$\lim_{k \rightarrow \infty} \frac{\log(|\alpha^k|)}{\beta \mathcal{P}_{k+1}} = 1, \quad \mu_B = e \lambda_B \text{ for } |\alpha^k| < 1 \text{ and } \beta \in \mathcal{R}, \text{ Thus } \mathcal{P}_k \text{ will be an}$$

ascending progression where k is equal to the number of big jobs waiting in queue.

Once the small vs. big job model is now complete, we move on into a model that will now propose a potential capacity setting with a trade off between two possibilities while a disruption is created by a “walk-in” customer., hold an asset from prompt response to a customer or use the resources to set up capacity to work and deliver to the customer scheduled .

Chapter 5. Order Acceptance For Service Systems with Exchange Inventory

5.1 Overview

The rest of this document will be referring to inspections, repairs, refurbishing for systems and sub-assembly collectively as: repairs. We note that in the specialized literature, there are several types of spare parts; those so called spare parts *repairable*, and those designated by the name of *rotables*.

The *rotables* are those modules that typically are proactively replaced from large systems such as ships, planes, trains, industrial production manufacturing, to avoid catastrophic failure of the components within the module. The repairs are then conducted at the module level either by maintenance prescriptive, or proactive maintenances, or by repair due to observation of the out of specification parameter of the spare part during inspection.

This document motivation, then continues to be, the formulation of models and policies needed to provide efficient turnaround to a critical *rotables*—in this particular example for aircrafts parts or modules when, for any reason, the component within the module, was not part of a scheduled repair.

In such conditions our modeling and contribution differs from the literature to recognize some business applications, where a *rotable* needs service or out of the proactive maintenance scheduled for repair, in other words a disruption of the programmed process.

This is especially the case for major service job of a landing gear that will require resources and time commitment. Arrivals of such requests are typically infrequent yet

generate considerable revenues for service MRO companies, this also is true in other service providers outside of those in the aeronautics industry such as train maintenance facilities, mass transportation, large industrial installations.

The definition of the landing gear is the undercarriage of an airplane. By landing gear we designate the structure that supports an aircraft on the ground and allows it to move—or taxi in the airlines lingo—takeoff—initiate flight—and land—finish flight. The components of the landing gear are typically: shock absorbing devices with oleo fluid levels, linkages, trusses and members, retracting and locking mechanisms, hydraulic lines, electrical systems, wheels, bearings, tires, brakes, and for some vehicles, other mechanisms, besides wheels are used, such as: skids, skis, floats or a combination of these and other elements.

At the operational level, most of the Aeronautical industry has extensive programs and schedules in place to ensure that every plane does have their components—for example landing gears—reviewed at the end of a prescribed period. The maintenance is usually done at the end, because if the review is done before, the timing for the next schedule resets, basically eliminating a number of useful days in the system. Ghobbar *et. al.* (Ghobbar & Friend, 2003) documented how significant a factor is the cost per day without operation of a plane. Thus the value of the program is to have the most useful hours of operation in the system, while avoiding days without operation.

Scheduling process for Certifications needs to be carried out according to federal regulations, all aerial and space vehicles are subject to a specific set of requirements to prove their *airworthiness* Title 14, part 39 (FAA, 2012) describes the conditions where an

aircraft is considered capable for flight; the process conducted to achieve this is called the certification process.

The certification process requires Service Providers to have trained technicians, facilities and resources, capable to perform inspections to any maintenance, preventive maintenance, rebuilding, and alteration done on any of the prescribed systems, components and parts deemed critical for the safe operation of an airplane.

The cited FAA regulation prescribes certain preventive maintenance that must be executed according to plan, after the aircraft has operated a predetermined number of hours. For example, Part 43 paragraph e, describes what each inspecting person is mandated to do during the annual or 100-hour inspection, while reviewing the components of the landing gear group. This involves scrutiny of diverse components such as: components' appearance, working order, insecure attachments, obvious or apparent defects, fluid levels, stress or wear, excessive wear, cracks, cuts, fatigue, distortion, proper operation mechanism verification, electrical systems checks.

The process for the preventive maintenance part is very well tuned for deterministic scheduling; it lends itself to be managed in this particular deterministic way.

Nevertheless, there could be complications that might disrupt the schedule and require the landing gear to be repaired, to keep certification.

These disruptions, coming from regular landing and departure operations originated from environmental factors, whose tendency is to result in non-linear behavior have the potential to complicate committed plans thus making them difficult to accommodate events using a deterministic scheduling program.

The issue is originated since this events are caused by several *complications* with stochastic sources.

The first potential complication is the result of material failures. Some of those mechanisms and components of the landing gear, for being subjected to the wear of materials, stress and fatigue—all this natural phenomena—related to the material composition, crystalline structures, and such micro components. A material dependent failure mechanism is known to be (Roth, Yanishevsky, & Beaudet, 1992) non-linear and stochastic in nature.

The second complication is related to landing gear deployment. Another stochastic source of landing gear certification needs is actual unpredictability of the operation of the aircraft, sometimes airplanes land without having deployed the landing gear. These phenomena are difficult to predict; still it is known to happen often, as many as 790 in 17,310 reported incidents (Aviation Safety Reporting System, 2012). All of these incidents require certification of the whole system or some of its components.

The third and fourth potential complications are derived from chapter 8 page 16 (Federal Aviation Administration, 2012), which describes two special inspections, which might trigger the need for a repair process for certification and airworthiness. Overweight landing, one of the two stochastic conditions that might require special inspections forces the reengagement of the landing gear certification process, in this condition a plane with passengers or cargo overload is flown and the landing gear subject to extreme operational conditions. Since this type of condition come in the course of daily operations, with multiple causes for the error, the behavior and presence of the issue is stochastic and non-linear as well.

Finally, the fourth and last complication is: the hard landing. The landing gear in this condition is subject to heightened stress due to the high speed and wrong angle / direction of the plane hitting the airport's tarmac while landing. Since those forces will depend of random factors, the behavior of this issue will be stochastic and non-linear too.

We focus on solving the need those repair stations—service providers or providers for simplicity—have to deliver prompt and cost effective service, on a non-scheduled landing gear certification. We provide models composed for the provider who performs these non-scheduled landing gear certifications, to make the most profitable decision to support the certification of these components.

The models will help to clarify the distinction between two approaches; the first maintaining a mixed an inventory and scheduled facility at minimum cost, minimum gain, and the second maintaining an inventory on hand which allows the provider to capture these non-scheduled customers—walk in's. The provider could in the second case command a premium price and thus realize significant gain, if proper consideration is done to the asset and investment cost, when they can provide immediate exchange.

We use markov chains modeling; first to propose what a Pure Exchange model will look like. By Pure Exchange we understand a facility that has a number of landing gears in stock and when one customer shows up, the landing gear is exchanged for another one, in this model there is not waiting, neither backlog is allowed, there are no landing gear inspection, rebuild and certification workstations available.

Second, we will entertain the idea of a Mixed Model where we have a number of workstations for landing gear repairs and we can provide some on hand stock. In this particular model constrains are relaxed on only exchange, and now, once the exchange

inventory is exhausted, make some workstations available. If there are landing gears or workstations available, landing gears can be accepted.

Think of this model as a way to generalize it to a case where you have stock on hand up to a point, but then once the inventory or the capacity are exhausted you can then have backlog. Capacity is represented by the number of workstations available for repair in conjunction with the stock available for exchange. This model we called it Mixed Model, with its results at hand we will have two data sets—pure and mixed models—available for analysis. Thus allowing the comparison in subchapter 5.5 for the models under different business conditions, this will allow us to get insights and provide answers to different business velocities—arrivals, reactions, and capacities.

Continuing in our model progression, in our previous case service providers that may have a steady but small size business flow and could attempt to receive a big job that shows up at times, the company have a steady business with scheduled repairs for small customers, if a big job requiring repair and the usage of some capacities shows up, the question is if the customer is to be taken.

In our aeronautical company example, we might think as what the proper price set to compensate the firm for providing this customer with an acceptance and with immediate accessible inventory. In fact same as in our previous analysis, the firm may choose to drop all repair capacity and offer only exchange if the price is right.

This model is the continuation of our progression to better understand order acceptance policies for big deals vs. regular jobs. In the previous chapters models help to understand pricing of big deals vs. small jobs, then relaxing models' assumptions, we reflected the fact that small jobs might be waiting in queue when a big job arrived and

policies were derived to evaluate business conditions and make a decision based on parameters such as; arrivals, services rates, penalties, regular job prices. As a result policies were provided to help the decision maker to get maximum gain by describing the situation to the model based on those parameters.

With this new model now the idea is to bring the acquired knowledge into yet another possible situation. In this occasion the decision maker needs to execute into a situation, where big deals will turn into cases where there is an interruption to an established schedule—also called walk-ins—and providing product for exchange in this conditions can command a premium as long as they are immediately available, and the decision maker wishes to know if the business conditions are such that if holding inventory at a certain premium will generate additional gains over the possibility to have their money invested in expanded capacity to handle more work.

For these opportunities we developed the model on subchapter 5.3 to look for the optimal premium, and the resource allocation policies under this light, it is a continuation on studies for optimal premiums for pricing, demand and order acceptance policies for a firm that faces deals under a markov chain setting. The objective of the study is to develop and compare optimization models for different demand management settings by pricing. Specifically, we consider two different strategies: 1) a pure strategy in which the firm commits to bid inventory exchanges only; 2) mixed strategy where the firm switches its allocation of capacity between inventory exchange and repair capacity.

This final question is addressed by the model mixing both concepts holding inventory and then once it is gone, continue to receive customers at a lower price in standard repair workstations.

We then, by the end of the chapter, we compare what are those conditions that make the Pure Exchange policy more successful than a Mixed policy, and by consequence when the opposite is true, and comment on the strategies insights with the support of different models developed as relaxations from the basic model.

5.2 Literature Review

Think off a facility—to illustrate the problem disruptions into the service queue—of an Aeronautical company working on landing gears certification, and the customers that request a non-scheduled service we called disruptions to the plan. These disruptions, coming from regular landing and departure, and other environmental factors originated from operations, whose tendency is to result in non-linear behavior have the potential to complicate committed plans thus making them difficult to accommodate events using a deterministic scheduling program.

The research in this area is extensive, starting with supply chain models for spare parts. There are several ways to categorize this research see for example Muckstadt (Muckstadt, 2005) for an extensive coverage in spare parts and *repairables*. The extensive research started with works such as those of Scarf with early work since 1962 (Clark & Scarf, 2004) and Iglehart (Iglehart, 1964) has also focused in the inventory management for the spare parts and the unique demand conditions for some of the components.

More recent work in the area of MRO in general has being done for trains such as those applications for models for deterministic scheduling without disruptions by Arts (Arts et al., 2012) Looking to present a solution for the resources planning and

maintenance with the particular case of rotables in the Netherlands. Or those also in train transportation for the resource programming and scheduling with disruptions were Chen et al (Chen, Yan, & Chen, 2010) working for the mass transit authority in Taiwan developed a model to schedule mixed deterministic and stochastic demands for the planning of carriage maintenance manpower supply.

The models very complete, tackle the problem from the stand point of a facility that has resources dedicated to internal maintenance and repairs of the transportation systems. Similar to what our model is looking yet their approach is from the operational level. In our case we continue to look for models in the strategic level, to answer to questions if makes sense, for example, to have rotables rather than repairs, to offer the product or service to external customers, and what should be the order acceptance policies to accept or reject.

These, questions have partial solutions from the third leg of research and the common subject of the different models depicted out of this document for the order acceptance policies. In this area the objective is to find conditions on the business typically expressed in markov chains, or queue theory, with underlying probability distributions and memory less states, where at any given business condition, there is a blocking probability and an optimal business performance that maximizes gain, if admission policies are set up to get into the optimal result. Work in this area has been done by Millhiser et al. (Millhiser & Burnetas, 2013)

From our document stand point, it is the combination of these three lines of research that give us the model we are looking after, that will help to answer in addition to the other formulated questions; At what price should it be bid?, finally if there are

conditions, when having an asset of one kind is recommend over another asset (think for example on the trade off between inventory vs. repair capability).

The contribution of our models will help to clarify the distinction between two approaches; the first maintaining a mixed facility at minimum cost, maximum gain, and the second maintaining an inventory on hand only facility, both of whom allow the provider to capture these non-scheduled customers—walk in's. The provider command a premium price and thus realize significant gain, if proper consideration is done to the asset and investment cost, when they can provide immediate exchange.

5.3 Basic Settings for the Models

We consider a MRO service provider that receives jobs with random intervals. An arriving job is an equipment or device that requires maintenance and overhaul. The MRO company can serve the customer by either overhauling the customer's original equipment (repair strategy) or by exchanging the incoming equipment with an in-stock ready-to-go equipment (exchange strategy). In the former case, the customer must wait until the service of its equipment is completed by the MRO process. This generally results with significant delays for the customer's operations (such as the operations of an aircraft) since the turnover times are usually long. In the latter case, the customer's equipment is immediately exchanged (or rented out) with another from the service provider's inventory which virtually eliminates the turnover times for the customer. The customer's original equipment - once serviced - is added to the MRO company's "exchange inventory" for a future exchange with another customer (or returned to the original owner and the rented equipment is returned back to the service provider's inventory). In practice

such inventory is also referred to as the *rotable* inventory. In this study, we use terms exchange inventory and *rotable* inventory exchangeably.

The exchange strategy based on rotable inventory is usually appealing to customers since they can bypass the waiting time for the MRO operations by means of a straightaway exchange. Typically, the downtime due to equipment overhaul is costly for the owners. Therefore, the equipment owners are willing to pay higher fees for quick turnovers, which may make the exchange option attractive also for the MRO service providers. On the other hand, in order to employ an “exchange” policy, the MRO service provider needs to continuously carry rotable inventory of overhauled equipment, which may incur substantial costs.

We investigate the value of the exchange approach from the perspective of the MRO firm. We study the threshold between the exchange strategy and the conventional overhaul strategy. We carry out our analysis for two policies: 1) pure exchange policy and 2) mixed policy. In the former case, the MRO adopts a policy on pure exchange where all arriving overhaul requests are responded by an exchange subject to availability of the finished rotable inventory. If there is no exchange equipment in the MRO company’s inventory, the arriving jobs cannot be accepted. In the mixed policy, once the exchange inventory is depleted, the subsequent job arrivals are still accepted for service, where the customer needs to wait for the completion of her equipment’s overhaul. Under this policy, the service fees depend on whether the customer receives an immediate exchange or waits for the whole overhaul process.

For illustration effects reader can think on the Avionics business where different landing gears and configurations have significantly different service times, however

initially many of the complexities of real life operations in the Avionics company will be for the moment set aside, as right now our interest is mostly to extend the basic model. So initially we will start with the unrealistic assumption that there is unlimited capacity in the shop so all demand can be met.

Also, for the purposes of this model we will consider a class of regular and standard customers that are already programmed for repair and refurbishment. Every day also, some of these engagements are finished successfully leaving the system. Systems or products under this process are assumed to remain in the process for whole number of days; we assume the system behaves holistically within a steady state determined probability distribution.

Equipment with similar refurbish and repair needs and characteristics is assumed to be homogeneous, which means that their processes take similar cycle times and probability distribution.

However, cycle times are assumed to be independent of one another and independent of the number of units under process in the system, the *occupied capacity*. Also we will need to quantify the effects such as the weekend activity build up and the staff availability by skill.

Another consideration will be that there is no re-scheduling. Jobs are executed, and once initiated, they have to be finished. Re-scheduling is possible when no capacity is available, and the buffer could be done in two ways, by leaving holes in the agenda or by adding extra capacity in the reserve.

When we have a customer walk-in, we assume that no jobs will be interrupted, only the next available slot will be allocated to the walk in if price tender—price

considerations are beyond the scope of this model—is accepted. Another assumption is that next job to be executed after the walk-in will be the next job on the scheduled program. Both assumptions will later be relaxed to allow insight on re-schedules and delivery issues.

In our analysis, we assume that the MRO requests arrive following a Poisson Process with an arrival rate denoted by λ . We also assume that the service time is exponential with a service rate denoted by μ . We consider a model with multiple servers where arriving jobs can be processed in parallel. We let c represent the number of parallel service lines operating under this setting. To model the resulting queuing system, we define the system state based on the number of exchange units in inventory, S .

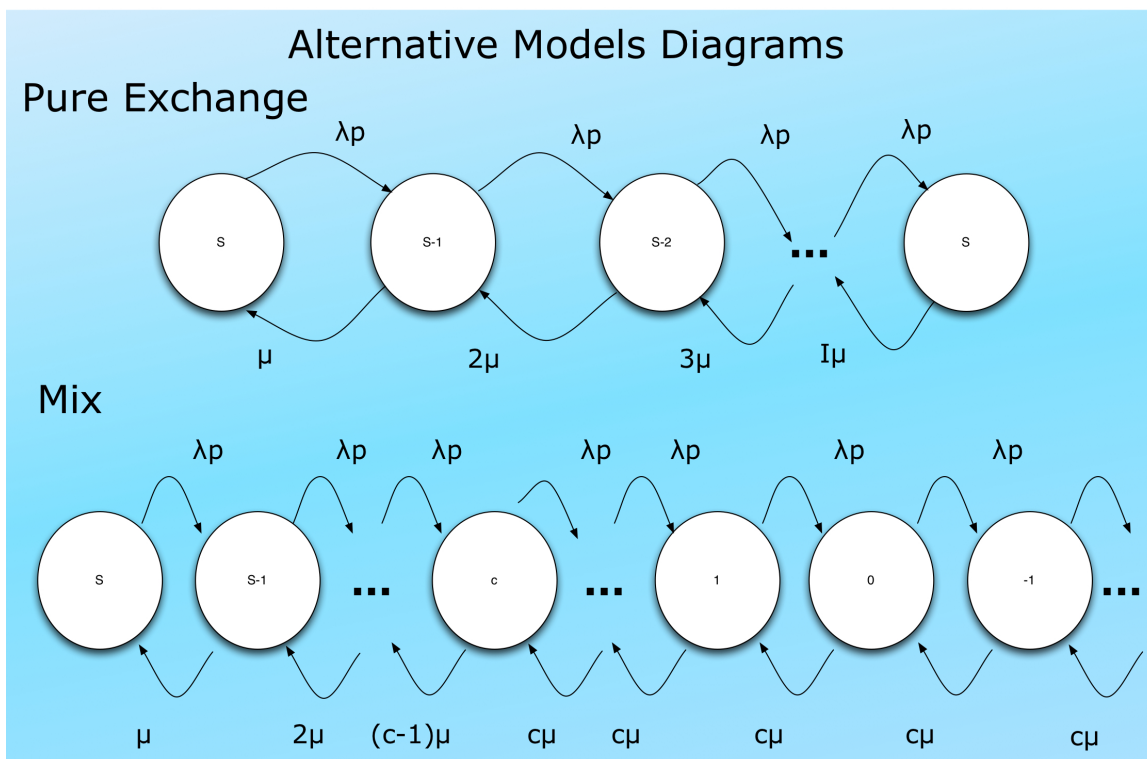


Figure 18. The Pure Exchange and Mixed order policy models.

5.3.1 Pure Exchange Model

We first investigate optimal policies for the pure exchange strategy, where the service requests are accepted only when exchange inventory is available. The service provider charges a fee, denoted by r , for each accepted demand for service. We consider a MRO service provider who maintains a rotatable inventory position for equipment exchange. We represent this quantity by S . The inventory position is composed of finished (overhauled) equipment inventory, S_f , and the work-in-process inventory, S_w . As such, the company has at all times $S=S_f+S_w$ units of equipment in the system. Each unit of equipment in inventory costs company h per time unit. In this section, we first propose a model to find optimal inventory position given exogenous parameters such as the arrival rate (λ), the service rate (μ), the number of service lines (c), and revenue per exchanged equipment (r). To avoid pathological cases, we assume that $r * \min(\lambda, \mu) > h$.

The markov process that represents the pure exchange setting depends on the comparison between the inventory position (S) and the number of service lines (c). Clearly, if $S \leq c$, the number of busy service lines cannot exceed S . Consequently, the service system effectively operates with S servers since demand beyond S will not be accepted by the service provider under this model. Moreover, there will be no service queue for the in-service exchange inventory.

On the other hand, if $S > c$, for $S_w > c$, there will be a queue for the service that replenishes (turns over) the exchange inventory. Based on this observation, we need to consider two different chains as depicted in Figure 18. We first solve the steady state parameters to generate the long run profit function for the server provider assuming abundant capacity (i.e., $S \leq c$), which is followed by the analysis of the case with scarce

capacity ($S > c$). Comparing two cases, we investigate optimality conditions for the server provider's exchange inventory position.

When capacity is abundant, the pure exchange policy turns into a truncated queue where no line forms. Using Kendall notation, we can denote this system as $M/M/c/c$, in such systems the steady-state probability distribution is captured by Erlang's first formula, namely,

$$\pi_n = \frac{\rho^n}{n! \sum_{i=0}^S \frac{\rho^i}{i!}} \quad (0 \leq n \leq S), \quad (64)$$

Where $\rho = \lambda/\mu$, which is understood to be the traffic intensity and π_n denotes the probability that there are n units of equipment in service (*i.e.*, $S_c = n$). The expected profit function depends on the effective arrival rate. As pointed out earlier, incoming orders are accepted as long as there is available inventory for exchange. When $S_c = S$, all inventory position is in service implying that no order can be accepted in this state. From (64), this probability is captured by π_S and referred to as the blocking probability. Consequently, the effective arrival rate is $\lambda(1 - \pi_S)$ and the associated revenue is $r\lambda(1 - \pi_S)$. Using eq.(64), we can write the profit function as follows

$$G_{S \leq c} = r\lambda \frac{A_{S-1}}{(\rho^S/S!) + A_{S-1}} - hS \quad (65)$$

where

$$A_x = \sum_{i=0}^x \frac{\rho^i}{i!} \quad (66)$$

In Eq. (65), the first term is the expected revenue whereas the second term is the cost of the inventory position. We first show that $G_{S \leq c}$ is concave in S .

Lemma 7. The steady state profit function $G_{S \leq c}$ is concave in S for $S \geq 1$.

Proof. To complete the proof, it is sufficient to show that $\frac{A_{S-1}}{(\rho^S/S!) + A_{S-1}}$ is concave in S . First we note that this term is a discrete function of S . A discrete function, $f(n)$ is concave in n a given region if for any n in the said region the following holds:

$$f(n + 1) + f(n - 1) - 2f(n) \leq 0. \quad (67)$$

Therefore, for concavity we need

$$\frac{A_I}{A_{I+1}} + \frac{A_{I-2}}{A_{I-1}} - 2 \frac{A_{I-1}}{A_I} \leq 0 \quad (68)$$

It is straightforward to observe that the left hand side of the above inequality returns a negative value for any $S > 1$ and $\rho > 0$ using a three dimensional graph. For the case where $S = 1$, the above inequality reduces to

$$- \frac{1}{(1+\rho)\left(1+\rho+\left(\frac{1}{2}\right)\rho^2\right)} \quad (69)$$

Which is clearly negative for $\rho > 0$. Consequently, we conclude that $G_{S \leq c}$ is concave in S for $S \geq 1$.



This result indicates that the profit function given in Eq. (65) has a unique maximizer and can be easily obtained using a line search. It is interesting to observe that the steady-state probability formula given in Eq. (64) is independent of the service distribution type and only depends on the average service time (Gross & Harris, 1998). As such, the optimal inventory position obtained from Eq. (65) is valid for any service time distribution.

When the number of service lines is small, (that is, assuming $S > c$) the resulting system becomes an $M/M/c/S$ queue. In this queue, the blocking probability, where the entire inventory position is in-service (that is, $S_c = S$ and $S_0 = 0$), is

$$\pi_S = \frac{(\rho/c)^S}{c!c^{-c}} \pi_0 \quad (70)$$

Where

$$\pi_0 = 1 / \left(\sum_{i=0}^{c-1} \frac{\rho^i}{i!} + \frac{1}{c!c^{-c}} \sum_{i=c}^S \left(\frac{\rho}{c} \right)^i \right) \quad (71)$$

From Eq. (70) and Eq. (71) we can re-write the blocking probability. Defining $z = \rho/s$, we get

$$\pi_S = \frac{c^c(1-z)z^S}{c!(1-z)A_c + zc^c(z^c - z^S)} \quad (72)$$

Hence, the expected profit function is

$$G_{S>c} = r\lambda \left(1 - \frac{c^c(1-z)z^S}{c!(1-z)A_c + zc^c(z^c - z^S)} \right) - hS \quad (73)$$

Next, we show that the above function is unimodular in S .

Lemma 8. The steady state profit function $G_{S>c}$ is unimodular with a unique maximizer for $S > 0$.

Proof. To prove unimodularity, we show that at any stationary point for $S > 0$ in (10), the second derivative returns a negative value implying that there must be a unique stationary point which maximizes the function. First let $X = c!(1-z)A_c + z^{c+1}c^c$. Then the first derivative of $G_{S>c}$ with respect to S is

$$G'_{S>c} = r\lambda \left(- \frac{c^c(1-z)\ln(z)z^S X}{(X - c^c z^{(S+1)})^2} \right) - h \quad (74)$$

At any stationary point, the above function should return zero. Now we look at the second derivative:

$$G''_{S>c} = r\lambda \left(-\frac{c^c(1-z)\ln(z)^2 z^S X(X+c^c z^{(S+1)})}{(X-c^c z^{(S+1)})^3} \right) \quad (75)$$

Evaluated at the stationary point (say $S=S^*$), the second derivative will return the following:

$$G''_{S>c}(S = S^*) = h \left(\frac{(X+c^c z^{(S^*+1)})}{(X-c^c z^{(S^*+1)})^3} \right) \ln(z) \quad (76)$$

First suppose $c \geq \rho$. In this case, since $z \leq 1$, $\ln(z) \leq 0$ and both the numerator and the denominator inside the above parenthesis are nonnegative. As such, the overall function returns a non-positive value. When $c < \rho$, $z > 1$ which implies that $\ln(z) > 0$. In this case, it is straightforward to see that the denominator inside the parenthesis is strictly negative since $c \leq S^*$. We need to take a closer look at the numerator to make conclusions about its sign. Using the incomplete upper gamma function, we can rewrite the numerator as follows:

$$X + c^c z^{(S^*+1)} = (1-z)e^\rho \Gamma(c+1, \rho) + c^c (z^{c+1} + z^{(S^*+1)}) \quad (77)$$

From the Gauss's continued fraction expansion of the incomplete gamma function (Jones, Thron, & Waadeland, 1982) we can observe that $\Gamma(c+1, \rho) < \rho^{c+1} e^{-\rho} / (\rho - c)$. Consequently, since $z > 1$,

$$X + c^c z^{(S^*+1)} > (1-z) \frac{\rho^{c+1}}{\rho-c} + c^c (z^{c+1} + z^{(S^*+1)}) \quad (78)$$

Where the right hand side reduces to ρ^{S+1}/c^{S-c+1} . Since this term is a lower bound for (77) and strictly positive, we can conclude that the equation in (77) always returns a positive value. Hence, the overall function in (76) must be strictly negative.

As such, we observe that any stationary point must be maximizing for the profit function implying that it must be unique.

From (65) and (73), now we can write the expected profit function for the MRO service provider in general form:

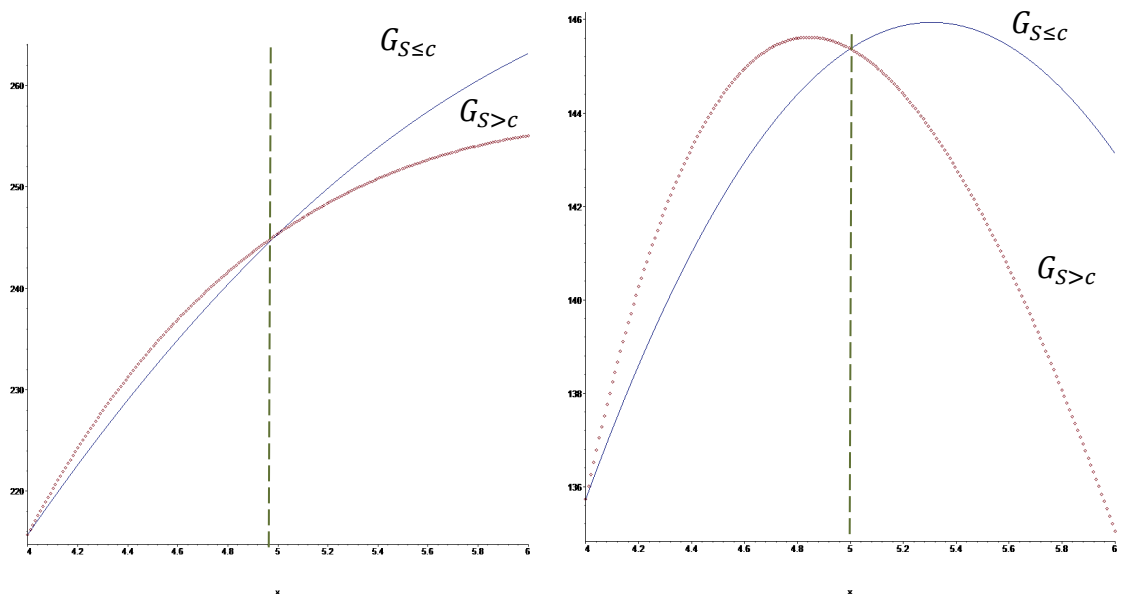
$$G(S) = \begin{cases} G_{S \leq c}, & \text{if } S \leq c \\ G_{S > c}, & \text{if } S > c \end{cases} \quad (79)$$

Given (79), we can investigate the optimal inventory position for the service provider.

We first need the following result:

Lemma 9. At $S = c$, the slope of $G_{S > c}$ is lower than the slope of $G_{S \leq c}$.

Proof. To see this result holds, we first note that at $S = c$ both functions intersect. Second, beyond this point (i.e., $S > c$) $G_{S \leq c}$ always returns higher values since this function is constructed based on the assumption that the maximum number of busy servers is S whereas in $G_{S > c}$ this number is c . As such the profit curve for the latter function should always be above the former one for any S such that $S > c$. This implies that at $S = c$, $G_{S > c}$ must have a lower slope than that of $G_{S \leq c}$.



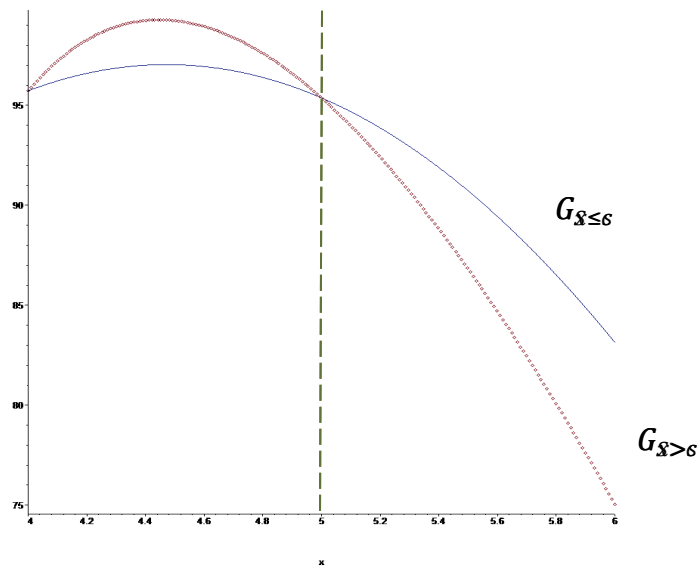


Figure 19. Expected gain curves.

The above result helps us deduce the possible scenarios for optimal inventory positioning. Basically, Lemma 9 implies that one of the three cases should occur: at $S = c$, either i) both curves are decreasing or ii) both curves are increasing or iii) while $G_{S \leq c}$ is increasing, $G_{S > c}$ is decreasing. All three cases are illustrated in Figure 19. In the first case, clearly the optimal inventory position is below c and computed based on $G_{S \leq c}$. In this case, the service provider's optimal exchange inventory is lower than the number of available servers and as such, there will be no queue for service. In the second case, the optimal inventory position must be above c implying that the optimal inventory level is above the number of available servers. The last case implies that at optimality, the exchange inventory level exactly matches the number of available service lines.

Based on this observation, we can derive the optimal inventory policies for the MRO service provider. We let S_L^* and S_H^* represent the integer inventory position values that maximize $G_{S \leq c}$ and $G_{S > c}$ respectively.

Proposition 5. For given arrival rates (λ), service rates (μ), and number of servers

(c) the optimal inventory position, S^* , is such that

$$i) \quad S^* = S_L^* \text{ where } S_L^* < c, \text{ if and only if} \\ h/r > \lambda \left(\frac{A_{c-1}}{A_c} - \frac{A_{c-2}}{A_{c-1}} \right) \quad (80)$$

$$ii) \quad S^* = c \text{ if and only if} \\ \lambda \left(\frac{A_{c-1}}{A_c} - \frac{A_{c-2}}{A_{c-1}} \right) \geq h/r \geq \lambda c^c z^c \left(\frac{c^c z^{c+1} - c!(z-1)A_c}{c!A_c(c!A_c + c^c z^{c+1})} \right) \quad (81)$$

$$iii) \quad S^* = S_H^*, \text{ where } S_H^* > c, \text{ if and only if} \\ h/r < \lambda c^c z^c \left(\frac{c^c z^{c+1} - c!(z-1)A_c}{c!A_c(c!A_c + c^c z^{c+1})} \right) \quad (82)$$

Proof. The inequality (80) is derived from $G_{S \leq c}(c-1) - G_{S \leq c}(c) > 0$. That is, the inequality holds if and only if the profit with $S = c-1$ is higher compared to $S = c$. This clearly indicates that in the latter case the profit function has a negative slope. From Lemma 9, we can conclude that $G_{S > c}$ must have a negative slope as well. Since both functions are known to be unimodular (as shown by Lemmas 7 and 8), the optimal inventory position must be strictly below s and hence, the optimal solution should be determined based on (65). If the inequality does not hold, from concavity, it is straightforward to see that optimal inventory position must be equal to or greater than s . Hence, the first part (i) holds.

The proof of part (iii) is similar. The inequality in (82) is directly reduced from $G_{S > c}(c+1) - G_{S > c}(c) > 0$. That is, the inequality holds if and only if the profit defined in (73) with $S = c+1$ is higher compared to $S = c$. This implies that at the latter point the profit function has a positive slope and as such, from Lemma 9, $G_{S \leq c}$ must have a positive slope as well. Since both functions are unimodular, the optimal inventory

position must be strictly above c and hence, the optimal solution should be determined based on (73). If the inequality does not hold, from concavity, it is straightforward to see that optimal inventory position must be equal to or lower than c .

The deduction obtained in the proofs of parts (i) and (iii) clearly implies that when (81) holds, the optimal inventory position is equal to the number of service lines.



The above result sums up the optimal inventory policy for the service provider, which depends on the ratio between the holding cost and revenue. As expected, when this ratio is above a threshold, the optimal inventory position will be below the available service lines, leading to a system with no queues and redundant servers. On the other hand, if the ratio is below another threshold, the optimal policy requires that the inventory level will be above the number of service lines.

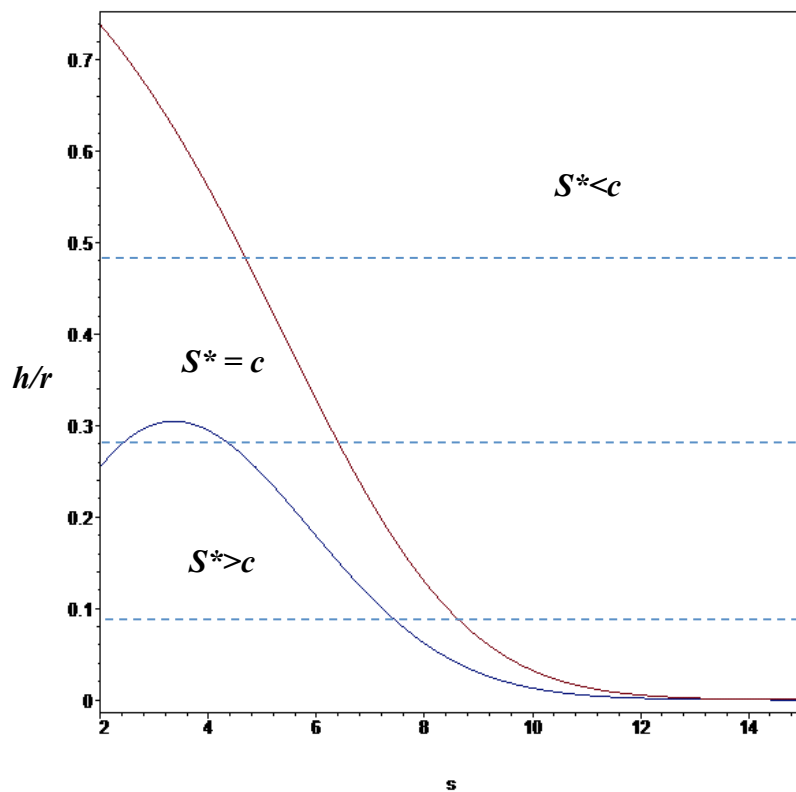


Figure 20. Threshold lines as functions of the number of servers

In between thresholds it is optimal for the service provider to match the inventory level precisely with the number of servers. In general, the optimal inventory position is non-decreasing in per unit revenue (r) and non-increasing in holding cost (h) as expected.

The threshold curves are illustrated in Figure 20. Typically, for a given ratio, as the number of server lines increases the optimal inventory level is realized below the number of server lines. Interestingly, in some cases while the optimal inventory level first matches the number of server lines, as the number of servers increases the optimal inventory will be first above the number of servers then back to be equal to the number of servers and finally falls below it. This case is illustrated by the middle dash line in Figure 20. This implies that the relation between the number of server lines and the optimal inventory level is not necessarily monotonic. The impact of system parameters on the choice of optimal inventory level is investigated in detail using a numerical analysis in the next section.

5.3.2 Exchange Model Numerical Examples.

In this section, we investigate the impact of model parameters on optimal inventory policies using a numerical analysis. To carry out the analysis, we compute optimal solutions for about 2,500 instances. We utilize the numerical instances to support our sensitivity analysis on varying value combinations across arrival rates, service rates, number of servers and holding cost.

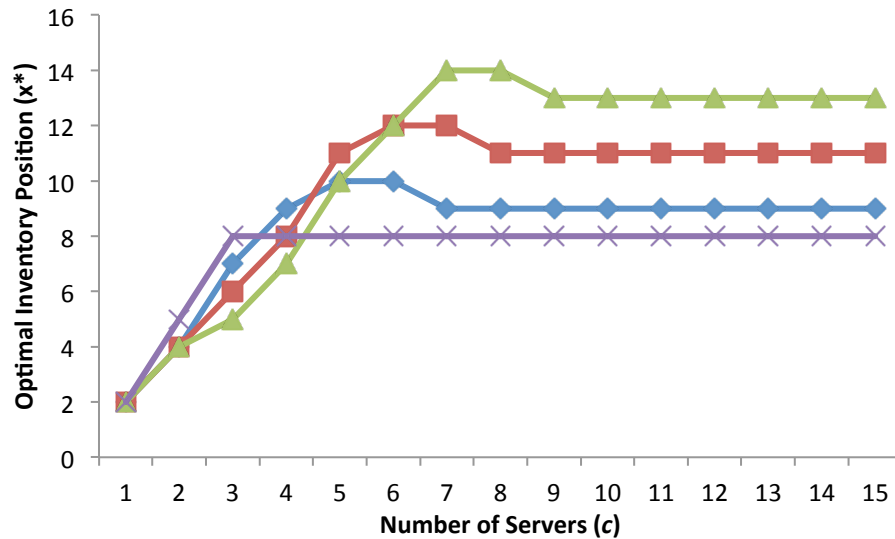


Figure 21. Optimal inventory as function of number of servers ($\mu = 4.5$ and $h = 25$)

Figure 21 illustrates the impact of the number of servers on the optimal inventory levels for varying values of the arrival rates. The optimal inventory curves indicate that for sufficiently low arrival rates the inventory level is non-decreasing with the number of servers. On the other hand, for relatively higher arrival rates the optimal inventory level first increases and then decreases. In all cases, the optimal inventory level converges to S_L^* . Basically, when there is limited number of servers, the service rather than the inventory becomes the bottleneck for the system. Increasing the number of servers enables more return on inventory investment up to a certain point. After that point, when service rate becomes less of a bottleneck for the system due to increased number of servers, fewer inventories are needed since inventory can be replenished rather quickly.

It is interesting to observe that when there are few servers, higher arrival rates lead to lower inventory units and when there are sufficiently high numbers of servers opposite is true. As mentioned above, in the former case service rate becomes more of a bottleneck as the arrival rates increases, which lowers the turnover for the inventory

leading to a decrease in inventory units. In the latter case, since there are sufficiently high number of servers higher arrival rates lead to higher inventory turnovers which results with increased inventory units. When the number of service lines is neither low nor high, the relation between the arrival rates and the optimal inventory is non monotonic. In order to better understand this pattern we first make the following observation regarding the trade-off between the arrival rates and the optimal inventory:

Lemma 10. Assuming $r\mu > h$, as $\lambda \rightarrow \infty$, $S^* \rightarrow c$.

Proof. First, note that as the arrival rate grows too large there is always demand for any completed job. Therefore, the effective demand is as high as the throughput of the overall service system. As such the effective demand rate is $S\mu$ when $S < c$ and $c\mu$ when $S \geq c$. In the former case, the mean profit increases in S since $r\mu > h$ whereas in the latter case it decreases with S since the revenue is a function of c while the total holding cost is increasing with S . Consequently, optimal inventory position, S^* , converges to the number of service lines, c . \diamond

This result implies that with sufficiently high demand rate, the optimal inventory position converges to the number of servers. This is intuitive in that at high arrival rate the system can admit arrivals as fast as its maximum throughput and as such, the service company cannot be better off by maintaining an inventory position either below or above c . In general, the impact of arrival rates on the optimal exchange inventory is illustrated in Figure 22.

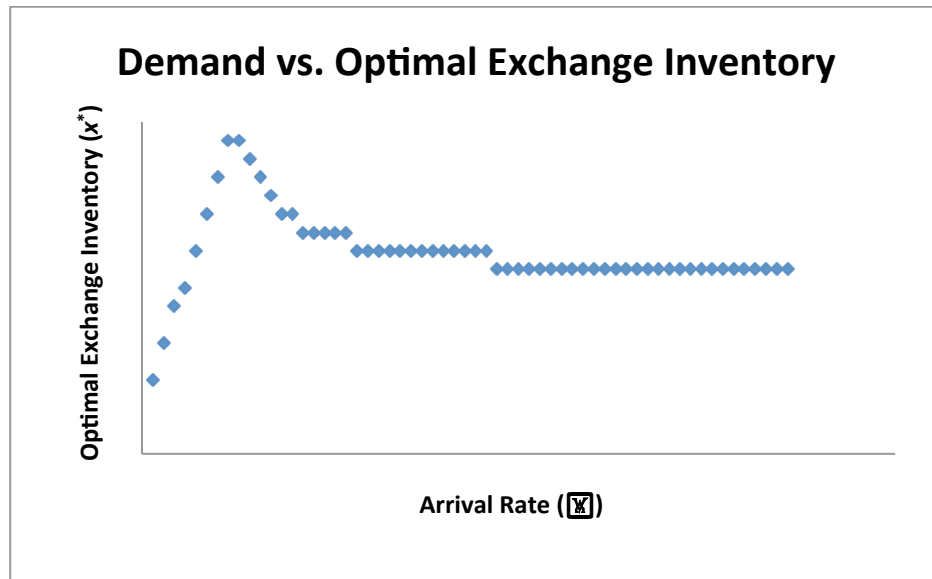


Figure 22. Arrival rate vs. optimal inventory ($c = 9, \mu = 4.5, h = 25$)

Given that all other parameters are constant, the optimal inventory first increases with λ to a certain point and then decreases until it converges to c . When the demand arrival rate is sufficiently low but increasing, the service company needs to increase its exchange inventory position in order to cope with increased demand. However once the traffic density reaches a tipping point (usually around $\lambda/(c\mu) = 1$), the arrival rates become too dense that the work-in-process inventory (S_w) and thus, the service queue inflate. In the steady state, the system will rarely have ready-to-exchange finished inventory (i.e., $S_f \sim 0$). Consequently, the overall service rate becomes the bottleneck for the system rather than the inventory. In this case, additional inventory will be rarely used and thus incur more cost than revenue. As the arrival rate further increases, the inventory level will converge to the number of servers.

Figure 23 provides another picture for the impact of number of servers. It depicts the optimal inventory curves for varying values of service rate, μ . Consistent to the

pattern present in Figure 21, optimal inventory level is first increasing and then decreasing with the number of servers, eventually converging to S_L^* in all cases.

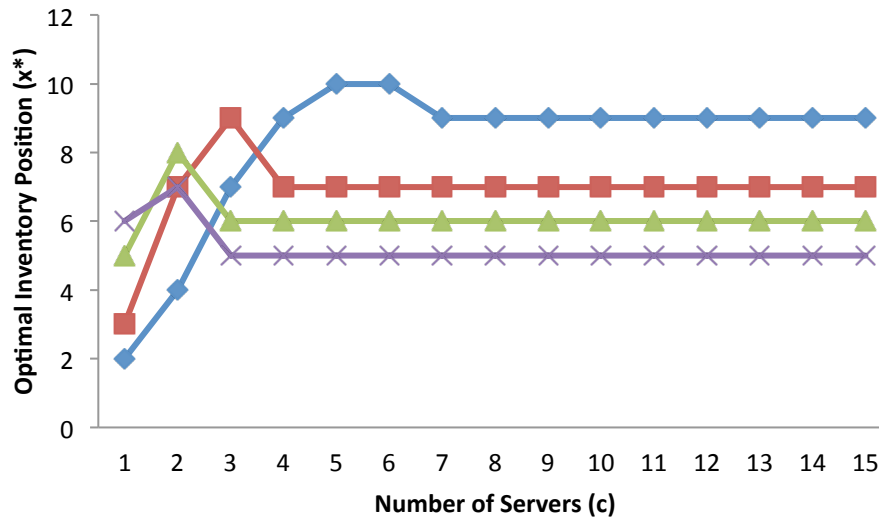


Figure 23. Optimal inventory as function of number of servers ($\lambda = 20$ and $h = 25$)

When few service lines are available, higher service rate results with higher inventory levels. Opposite is observed when the number of service lines is sufficiently high. In the former case, increasing service rate enables the use of more inventory units. With high number of service lines and increasing service rates leads the utilization will be too low to justify higher inventory levels. As such, the increase in replenishment rates result with lower inventory. Similar to the case with arrival rates, when the number of servers is neither low nor high the relation is non monotonic.

Further studying the influence of service rates, we first observe that the inventory level converges to one as service rate becomes too big.

Lemma 11. Assuming $r\lambda > h$, as $\mu \rightarrow \infty$, $S^* \rightarrow 1$.

Proof. The proof is straightforward from the fact that an infinite service rate implies instantaneous overhaul and hence replenishment of the exchange rate and as such, there will be no need to carry more than a single unit of inventory. \blacklozenge

Consistent to Lemma 11, in general, the optimal inventory level decreases as the service rate increases. However, as illustrated in Figure 24,

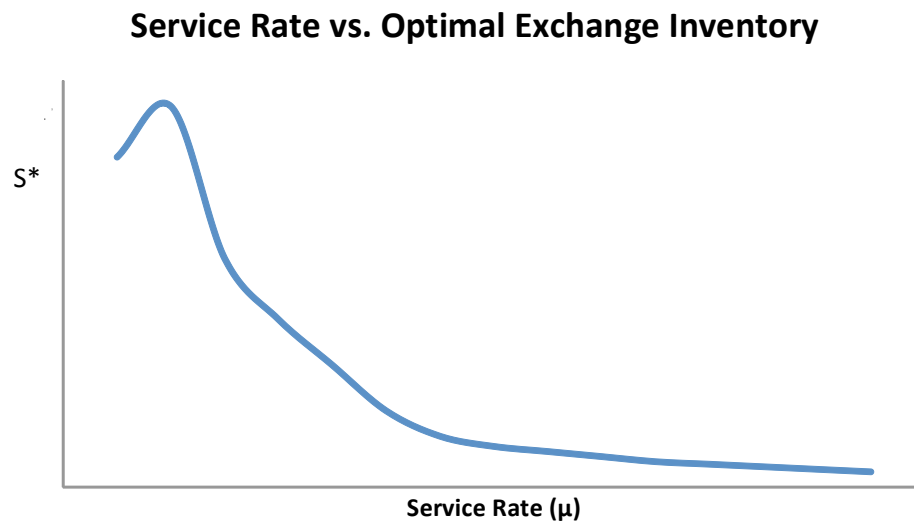


Figure 24. Service rate vs. optimal inventory ($c = 9$, $\lambda = 25$, $h = 25$)

in some cases when the service rate is too low, the inventory level increases first with the service level to a certain point. This is especially the case when the arrival rate is high, which as explained above, leads to lower finished overhaul inventory due to relatively slow turnover in service. At this point, as the high arrival rates are balanced by the increasing service rates additional inventory brings more turnover and hence revenues.

Consequently, higher inventory position improves the revenue. As the service rate continues to increase, effective utilization in the system and hence, the need for higher inventory position decreases.

In general, we conclude that the optimal inventory level is relatively lower with low and high loads. Typically, need for inventory peaks for intermediate load levels. When the system load (or traffic density) is too low, the return on investment in inventory does not justify keeping high number of equipment units for exchange. On the other hand, when the load is too large, service becomes a bottleneck for the system and substantially long replenishment times do not allow for turnover of large inventory at steady state.

5.3.3 The Mixed Policy for Exchange and Service

In this stage, we consider a mixed model that allows for conventional service where the customer needs to wait through the service process. As in the previous case, the service firm maintains a rotatable inventory for exchange. Moreover, when the whole rotatable inventory is work-in-process (that is, $S_f = 0$ and $S_w = S$), any arriving customer's equipment will be directly admitted into the service process. This customer will receive the first equipment that comes out of the service process. Customers who receive conventional service instead of equipment exchange are charged αr where $\alpha \in (0,1)$. As such, these customers pay a portion of the full exchange service fee since their service is not instant as in the case of direct exchange. It is assumed that all customers are willing to take the exchange option if available and are also willing to wait for service only if direct exchange is not possible in return of a discounted rate set by α .

This setting translates into an $M/M/c/\infty$ queuing system, where the service firm needs to determine the optimal inventory level for exchange. Similar to the previous case,

the long run gain function is conditional on whether the inventory level is below the number of servers or not. We first consider the case $S \leq c$. Revenue from direct exchange is realizable when there is at least one finished inventory available for exchange. As such, the effective rate for exchange revenue is $\lambda \sum_{i=0}^{S-1} \pi_i$. Since $S \leq c$, the steady state probabilities in that range are

$$\pi_i = \frac{\rho^i}{i!} \pi_0 \quad (83)$$

where.

$$\pi_0 = 1 / \left(\sum_{i=0}^{c-1} \frac{\rho^i}{i!} + \frac{1}{c!c^{-c}} \sum_{i=c}^{\infty} \left(\frac{\rho}{c} \right)^i \right) \quad (84)$$

Consequently, we can write the gain function for the mixed model as follows:

$$G_{S \leq c}^m = (1 - \alpha)r\lambda A_{S-1}\pi_0 + \alpha r\lambda - hS \quad (85)$$

where A_x is defined in (66). In contrast to the pure exchange case, the above gain function is not always concave. As explained below, concavity is attained under certain conditions and it mainly depends on the traffic density.

Lemma 12. *The steady state profit function $G_{S \leq c}^m$ is concave in S for $S \geq \rho$. The function is strictly convex for S in $[0, \rho)$.*

Proof. When we apply the inequality given in (67), we obtain the following condition for concavity:

$$A_S + A_{S-2} - 2A_{S-1} \leq 0. \quad (86)$$

It is straightforward to observe that the above inequality reduces to $S \geq \rho$. As such, the function is concave for $S \geq \rho$ and strictly convex otherwise. $S = \rho$ is the saddle point. \blacklozenge

The above result implies that the optimal solution is not necessarily identified by the stationary point in the gain curve. Let S_L^{m*} denote the optimal inventory level for the mixed policy where $S \leq c$.

Lemma 13. *There exists a threshold value for $h/(1 - \alpha)r$, say ϖ_L , such that if $h/(1 - \alpha)r > \varpi$ then $S_L^{m*} = 0$. Otherwise, optimal inventory level is the stationary point that corresponds to the inventory point above ρ , that is, $S_L^{m*} > \rho$.*

Proof. We first note that the result of Lemma 12 implies that the gain function is either i) strictly decreasing in $S \in [0, \infty)$ or ii) first decreasing, then increasing and then decreasing, or iii) first increasing and then decreasing. Clearly, in the first case the optimal inventory level is zero and in the third case the unique stationary point maximizes the gain function. In the second case, there are two stationary points: one is a local minimizer and the other a local maximizer. Clearly, the latter one is also the global maximum if and only if the value of the gain function is greater than $G_{S \leq c}^m(S = 0)$. At the stationary point we get

$$(1 - \alpha)r\lambda A'_{S-1}\pi_0 = h \quad (87)$$

Let S^* be the inventory level that satisfies the above equation which is also greater than ρ (i.e., S^* is the local maximizer). For this inventory level to be the global maximizer, the following inequality must hold:

$$G_{S \leq c}^m(S = S^*) - G_{S \leq c}^m(S = 0) = \frac{\lambda A_{S^*-1}\pi_0}{S^*} \geq h/(1 - \alpha)r \quad (88)$$

It is straightforward to observe from (87) that S^* decreases with $h/(1-\alpha)r$ since the function is concave at this point. Suppose that the inequality (88) holds. Then it can be shown that the left hand side is non-increasing in h/r . If the inequality does not hold, the

left hand side increases with $h/(1-\alpha)r$ with a lower rate (since A_{S-1} is concave in this region). Consequently, we conclude that there is a unique threshold for $h/(1-\alpha)r$ above which $S_L^{m*} = 0$ and below which $S_L^{m*} = S^*$. Moreover, we know from Lemma 12 that $S = \rho$ is a saddle point implying that $S^* > \rho$ must be true.

■

The above result indicates that when the ratio between the holding cost and revenue is above a unique threshold, the firm is better off with not offering any direct exchange service. We note the threshold itself is a function of h , r , and α , since these parameters determine the optimal inventory level. As expected, this threshold in general decreases in α or in other words it decreases as the revenue gap between direct exchange and conventional service closes up. Unfortunately we cannot derive a closed form equation for ϖL , however its numerical computation is fairly easy.

For the case of $S > c$, the steady state revenue rate and hence the long run gain function change as what follows:

$$G_{S>c}^m = (1 - \alpha)r\lambda\pi_0 \left(A_{c-1} + \sum_{i=c}^{S-1} \frac{\rho^i}{c^{i-c}c!} \right) + \alpha r\lambda - hS, \quad (89)$$

which can be rewritten as

$$G_{S>c}^m = (1 - \alpha)r\lambda\pi_0 \left(A_{c-1} + \frac{c^c}{c!} \frac{(z^c - z^S)}{1-z} \right) + \alpha r\lambda - hS. \quad (90)$$

Here, as earlier, z is defined as $\frac{\rho}{c}$. In what follows we show that the above function is strictly concave in S for $z < 0$.

Lemma 14. *The steady state gain function $G_{S>c}^m$ is strictly concave in S for $z < 0$.*

Proof . We prove the concavity from the second order derivative. The second derivative of $G_{S>c}^m$ with respect to S is

$$\frac{\partial^2 G_{S>c}^m}{\partial S^2} = -(1-\alpha)r\lambda\pi_0 \left(\frac{c^c (\ln z)^2 z^S}{c!(1-z)z} \right). \quad (91)$$

Clearly, since $\alpha < 1$, the above function always returns a strictly negative value if and only if $z < 1$. Hence, for these value ranges, $G_{S>c}^m$ is strictly concave in S .



Similar to the $S \leq c$ case, this function leads to a non-zero optimal inventory level if $h/(1-\alpha)r$ is below a threshold.

Lemma 15. *There exists a threshold value for $h/(1-\alpha)r$, say ϖ_H , such that if $h/(1-\alpha)r > \varpi$ then $G_{S>c}^m$ results leads to zero inventory at optimality. Otherwise, optimal inventory level is the stationary point that corresponds to the inventory point above c .*

Proof. The proof is similar to that of Lemma 13. With a closed form definition for ϖ_H .



From (85) and (90), we can lay out the overall expected gain function for the MRO service provider as follows:

$$G^m(S) = \begin{cases} G_{S \leq c}^m, & \text{if } S \leq c \\ G_{S > c}^m, & \text{if } S > c \end{cases} \quad (92)$$

Given (92), now we investigate the optimal inventory position for the service provider and derive the following conclusion. For the next result, we let $\varphi = \lambda\pi_0\rho^c/c!$

Proposition 6. *For given arrival rates (λ), service rates (μ), and number of servers (c) the optimal inventory position, S^{m*} , is such that*

i) $S^{m*} = 0$ if and only if

$$\frac{h}{(1-\alpha)r} > \varpi_L \quad (93)$$

ii) $S^{m*} = S_L^*$ where $S_L^* \in \{\lceil \rho \rceil, \lceil \rho + 1 \rceil, \lceil \rho + 2 \rceil, \dots, c\}$, if and only if

$$\varpi_L \geq \frac{h}{(1-\alpha)r} \geq \varphi \quad (94)$$

iii) $S^{m*} = \arg \max_{x \in \{\lfloor y \rfloor, \lfloor y \rfloor + 1\}} G_{S > c}^m(S = x)$ where $y = \frac{\ln\left(\frac{c!h(1-z)}{(1-\alpha)r\pi_0 c^c (-\ln z)}\right)}{\ln(z)}$ if and only if

$$\frac{h}{(1-\alpha)r} < \min(\varphi, \varpi_H) \quad (95)$$

Proof. We first note from Lemma 13 that keeping inventory is not optimal if (93)

holds. Now, we show that at $S=c$, both $G_{S \leq c}^m$ and $G_{S > c}^m$ are either increasing or decreasing together. We can easily see that at $S=c$

$$G_{S \leq c}^m(S = c + 1) - G_{S \leq c}^m(S = c) = G_{S > c}^m(S = c + 1) - G_{S > c}^m(S = c) = (1 - \alpha)r\lambda\pi_0 \frac{\rho^c}{c!} - h. \quad (96)$$

We can deduce from the above result that both functions are increasing or decreasing at c . Noting that $c > \rho$, at $S=c$, both functions are concave. Therefore, if equation (96) returns a non-positive value optimal inventory level must be at or below c . From (96), this happens only when $h/(1 - \alpha)r \geq \varphi$. From Lemma 13, we know that if $h/(1 - \alpha)r$ is below threshold ϖ , then the optimal inventory is in $(\rho, c]$. The ordering between φ and ϖ is not monotonic. Consequently, S^{m*} must be strictly greater than c if and only if $h/(1 - \alpha)r$ is below both φ and ϖ . In this case, the exact value of the optimal inventory is the next integer value that is either below or above y , which is computed from the first order optimality condition of $G_{S > c}^m$.



The above result indicates that the optimal inventory level is either zero or above the traffic density (ρ) depending on the trade-off between the holding cost, base price and premium for the exchange service. As expected, optimal inventory level increases in

service fee (r), and decreases in holding cost (h) and the discount rate ($1-\alpha$). Figure 25 illustrates all three scenarios. When the holding cost or revenue from non-exchange services is sufficiently high, there is no incentive to keep any inventory for exchange.

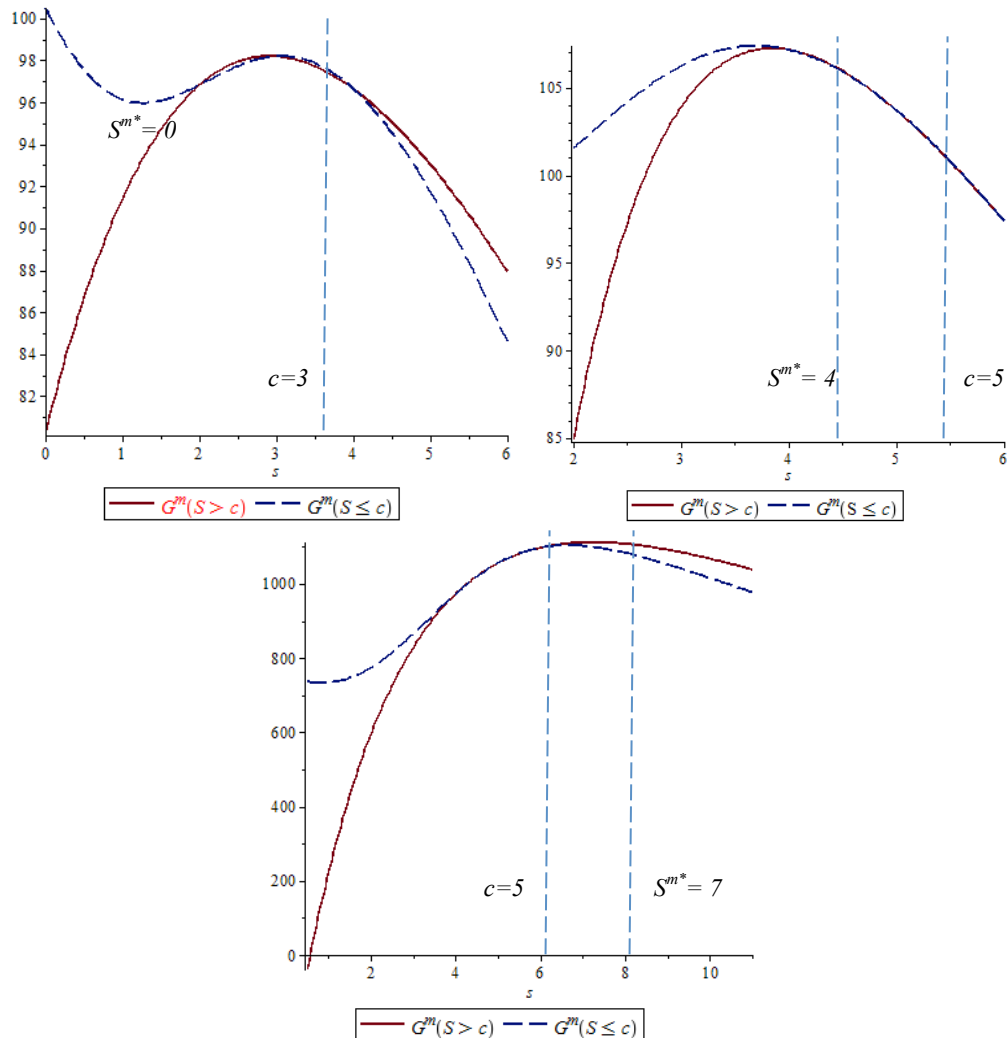


Figure 25. Expected gain curves under the mixed policy with optimal inventory levels of $S^{m*} = 0$, $S^{m*} \leq c$, and $S^{m*} > c$.

The cost of inventory and/or the premium for exchange does not justify the exchange policy. As the cost decreases or the gap for the exchange revenue increases, it is optimal for the firm to keep rotatable inventory for exchange. In this case, as shown in Lemma 13, it is never optimal for the firm to keep an inventory level that is below p . The

intuition is that at steady state, the probability that there will be less than ρ jobs is almost zero. This implies that if the inventory level is kept below ρ , on the long run, the firm will perform an exchange service almost never. Finally, when holding cost is sufficiently low or the premium from the exchange service is sufficiently high, the optimal inventory level extends beyond the number of servers.

As pointed out in the above discussion, the optimal solution depends on two types of thresholds whose orderings are not monotonic. That is, there are cases where one is larger than the other. The relation between φ and ϖ_L is non-linear in ρ . In general, as ρ gets bigger, φ becomes larger than ϖ . In the rest of the paper we will abuse the notation by representing both ϖ_L and ϖ_H with ϖ . We recall that while the former one applies to $G_{S \leq c}^m$, the latter one is used for $G_{S > c}^m$. The trade-off is further investigated in the next section.

5.3.4 Mixed Model Numerical Examples

In order to investigate the impact of system parameters on the optimal inventory policies under the mixed model, we further our study with a numerical study. More than 7,000 instances for various combinations of the arrival rate, service rate, revenue coefficient for non-exchange services, holding cost, and number of servers were generated.

Looking, first, the impact of the arrival rate (λ). Clearly, a higher arrival rate implies higher demand for service. Everything else is fixed, this also means higher load and lower π_0 for the overall system. Consequently, the impact of λ is not obvious for the optimal policy. Figure 26 depicts a typical relation between λ and the thresholds presented in Proposition 6. The dashed curve sets the boundary for positive inventory.

When $h/(1-\alpha)r$ is above this curve the firm is better off with carrying no inventory for exchange. The other curve sets the boundary between $S^{m*} \leq c$ and $S^{m*} > c$. Thus, for any $h/(1-\alpha)r$ that falls in between these two curves, the optimal inventory level will be positive but no more than the number of servers. For other values, which fall under both curves, the inventory level will exceed the number of servers.

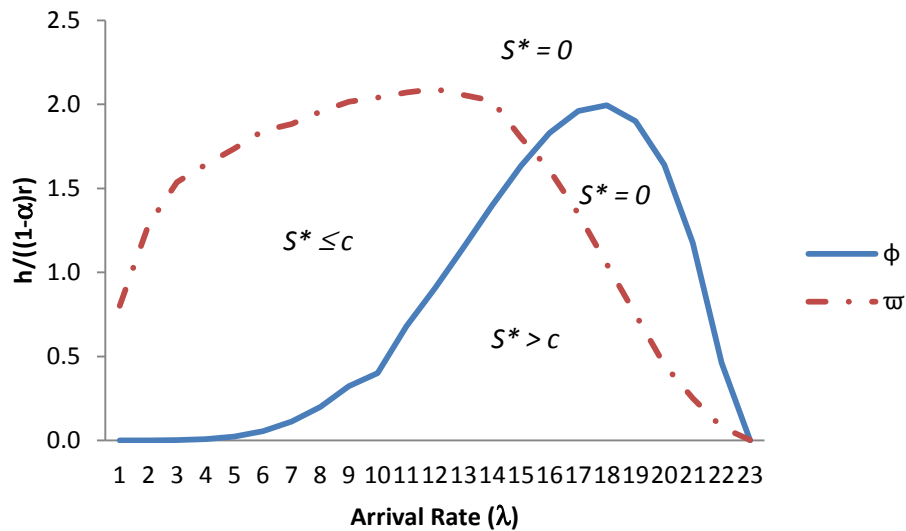


Figure 26. Threshold curves as functions of the arrival rate ($r=100, h=55, \alpha=0.7, \mu=4.5, c=5$).

In general we observe two patterns: 1) the ordering of the thresholds changes as λ increases and specifically φ becomes larger than the zero-inventory threshold (ϖ), and 2) both thresholds first increase and then decrease with λ .

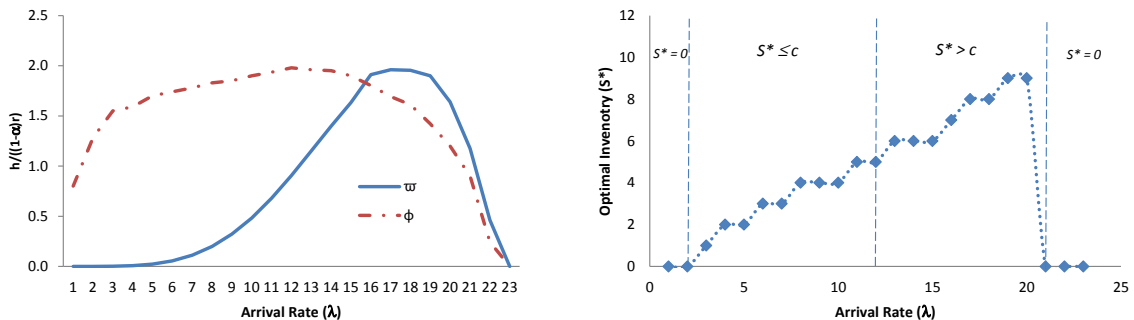


Figure 27. Threshold curves and optimal inventory levels as functions of the arrival rate ($r=100$, $h=55$, $\alpha=0.5$).

The first observation can be generalized to the traffic density of the system. Typically, φ becomes larger than the zero-inventory threshold as z increases. The second observation hints that for very low and very large values of λ , the optimal inventory must be zero. This pattern is confirmed by our numerical results and illustrated in Figure 27.

Clearly, when the arrival rate is too low (that is, demand is very infrequent), high inventory cost does not justify the risk of carrying inventory for the exchange service. After a tipping point, the inventory level typically increases with the arrival rate. Once the arrival rate is too high, in which case, the system load (z) is approaching to 100%, the firm will hardly have any occasion for exchange as it will have not much chance to process the equipment for inventory. Most equipment that come out of the MRO service process will be delivered to a customer that has already arrived earlier and waiting in the line. As such, the frequency of exchange services and thus the return from exchange inventory diminishes to a point that does not offset the inventory cost. Consequently, beyond another tipping point, the optimal inventory level jumps down to zero. We note that the optimal inventory level may jump down to zero before reaching to the $S^{m*} > c$ range. This happens especially when α is large, where the added revenue margin for the exchange service revenue is slim. We note that the observation for the mixed policy regarding the impact of λ is consistent with the pure exchange case in some sense. In the pure exchange case, we observe that the optimal inventory decreases in λ once λ becomes large. However, the optimal inventory converges to the number of servers instead of zero as is the case for the mixed policy.

Clearly, in the former case, the firm earns money only from the exchange service and high demand requires that all servers are utilized. In this case, the large demand can generate revenue via the conventional service with continuous flow of customers as such carrying costly inventory is not needed to enhance overall revenues.

In general, long-run gains increase with the arrival rate. However, interestingly, we observe that there are exceptions. Figure 28.a illustrates a case where higher demand in fact degrades company profits in a brief region, where $S^{m^*} > c$. To help understand this better, Figure 28.b provides a closer look for the turnover of the exchange inventory.

As mentioned above, when the arrival rates become large, the system will be crowded with service requests, which result in lower turns for the exchange inventory. After a point, the frequency of exchange services decrease with λ , which leads to lower bang-for-buck for the inventory. The diminished returns from inventory reach to an extent that hurts the overall profits. As shown in the Figure, with further increase in λ , the optimal inventory drops to zero and after that point gains continue to increase absent exchange services. We note that, optimal $G_{S>c}^m$ exhibits a quasi-concave pattern with λ in general. That is, as demonstrated in Figure 28, it is decreasing in λ after reaching to its peak.

However, in most cases, the optimal policy enters into the zero-inventory zone before the peak is reached. The situation demonstrated in Figure 28 arises when α is small (that is, the premium for exchange services is high) which elongates the region for $S^{m^*} > c$.

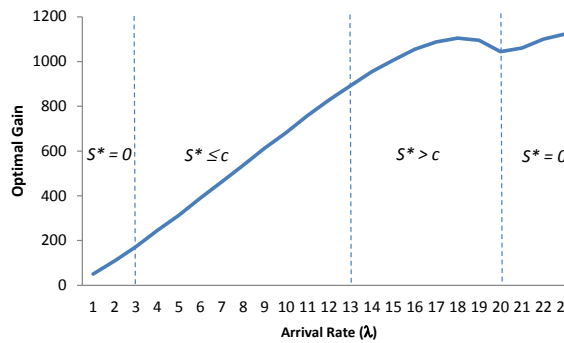


Figure 28.a. Arrival rate vs. optimal gain

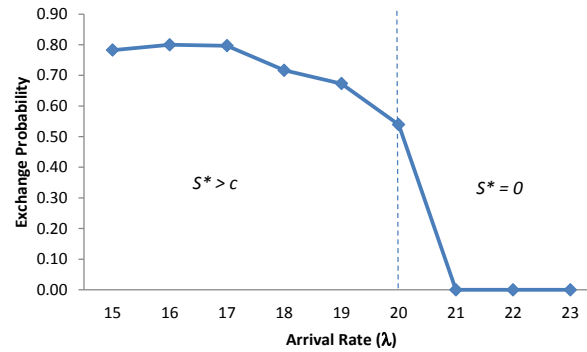


Figure 28.b. Arrival rate vs. exchange probability

Figure 28. Impact of the arrival rate on the optimal gain ($r=100, h=55, \alpha=0.5, \mu=4.5, c=5$).

Next, we investigate the influence of the service rate (μ). Similar to the pure exchange case, we can have the following general conclusion regarding the mixed policy for high service rates:

Lemma 16. *Assuming, $(1-\alpha)r\lambda > h$, as $\mu \rightarrow \infty, S^{m*} \rightarrow 1$.*

Proof. As explained in the pure exchange case, when the service rate becomes sufficiently high, the service is performed almost instantaneously, where π_0 approaches to 1. As such, more than often times, the firm serves to a single customer at each time. Since $\alpha \leq 1$, it is optimal for the firm to carry only one exchange inventory which can be turned as fast as λ (rate with which the orders arrive). In other words, the mixed policy converges to a pure exchange policy in this case.



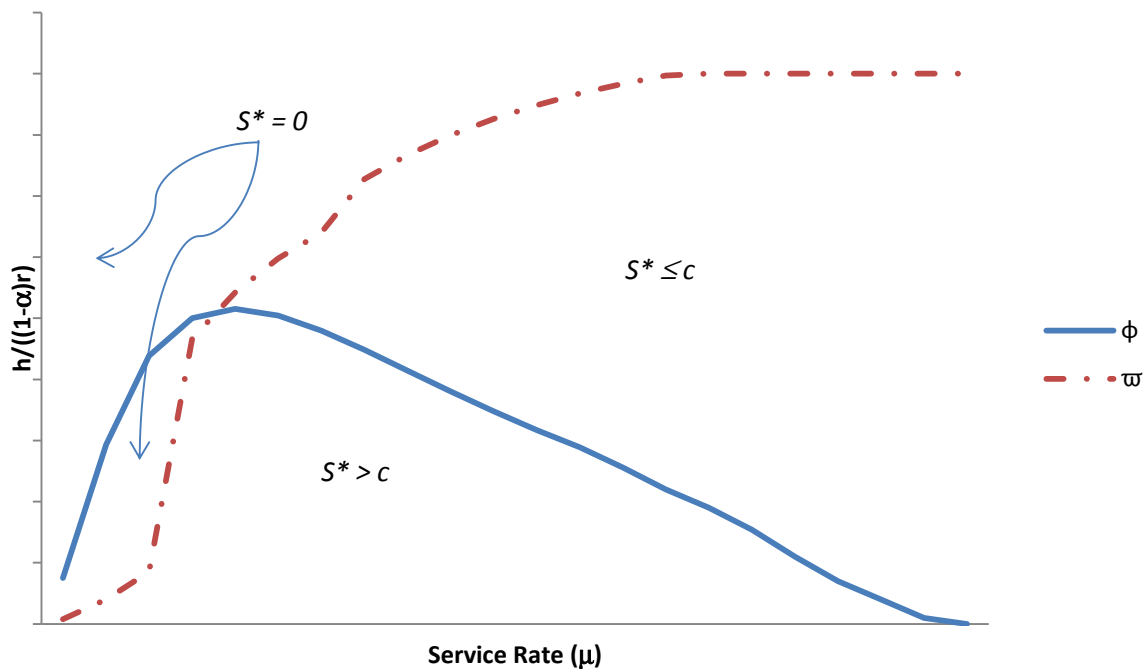


Figure 29. Threshold curves as functions of the service rate ($r=100$, $h=70$, $\alpha=0.7$, $\lambda=20$, $c=4$).

The numerical results suggest that it is not optimal to carry any exchange inventory if the service rate is too low. This is indicated by the threshold curves in Figure 29. Typically, when the service is slow, both threshold values are too low leading to $S^{m*} = 0$. Small μ causes the system load to be high and hence, the firm experiences the conditions discussed above for large λ . The low turnover on equipment does not justify the risk of carrying any inventory for exchange. Consistent to our previous observations, as μ increases the system load decreases and as such, the thresholds first increase and then decrease as shown in Figure 29. This leads to a pattern in optimal inventory levels illustrated by Figure 30.

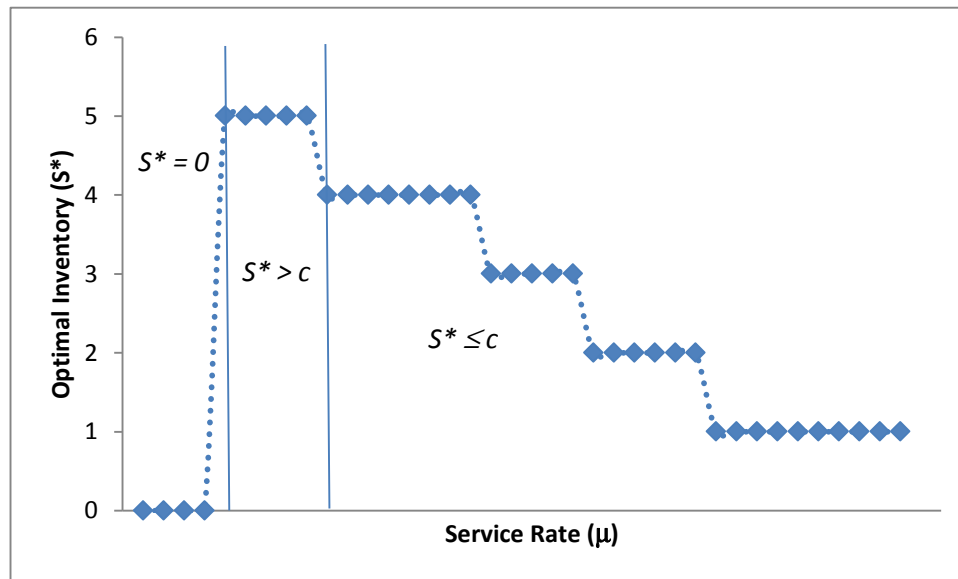


Figure 30. Service rate and optimal exchange inventory ($r=100$, $h=70$, $\alpha=0.7$, $\lambda=20$, $c=4$).

As the service gets faster, the inventory level jumps up, after which it monotonically declines eventually converging to 1. We note that the optimal inventory level may not always jump to the $S^{m*} > c$ region. This is typically the case when the holding cost, α , and/or the number of servers are high. The firm profits always increase as the service gets faster.

The impact of the number of servers is a bit more ambiguous. A typical case demonstrating the influence of the number of servers on the threshold curves is given in Figure 31.

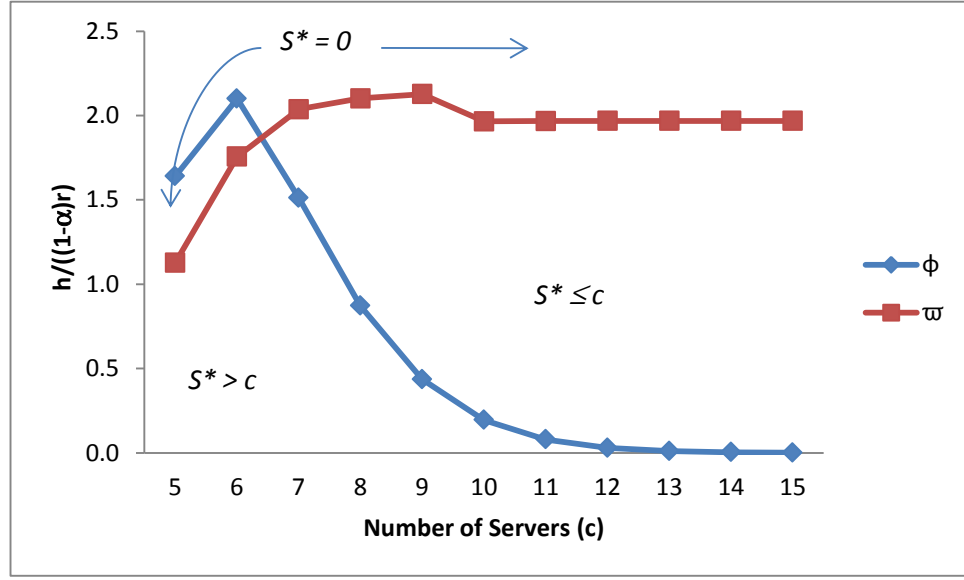


Figure 31. Threshold curves as functions of number of servers ($r=100$, $h=40$, $\alpha=0.5$, $\lambda=20$, $\mu=4.5$).

Basically, small number of servers implies higher load on the system, which results with φ being larger than ϖ , indicating that the optimal solution is realized either at $S^{m*} = 0$ or $S^{m*} > c$. The exact starting point (e.g., at $c = 1$) depends on $h/(1-\alpha)r$. As the number of servers increases φ typically decreases while ϖ increases. This implies that the optimal inventory eventually enters into the $S^{m*} \leq c$ region. Since in this region $G_{S \leq c}^m$ is concave, it can be easily deduced from (85) that S^{m*} increases in c . Consequently, we can make the following conclusion:

Lemma 17. *As $c \rightarrow \infty$, the optimal inventory converges to a unique level that solves the following equation:*

$$(1 - \alpha)r\lambda A'_{S-1} = h \quad (97)$$

Proof. We first note that $c \rightarrow \infty$ means $\rho \rightarrow 0$ and $\pi_0 \rightarrow 1$. Since c is too large, the long-run gains are captured by $G_{S \leq c}^m$. In this case, all feasible inventory choices will be greater than ρ , which from Lemma 6 implies that $G_{S \leq c}^m$ is strictly concave for $S > 0$.

Consequently, $G_{S \leq c}^m$ reduces to the function given in (85) with $\pi_0 = 1$. Equation (97) is simply the first order optimality condition for this function.



As in the case of service rate, the optimal gain always increases with the number of servers. In general, keeping traffic density (ρ) at a constant level, increase in the arrival and service rates lead to higher profits and inventory levels. As shown in Figure 32 with both thresholds increasing, the gap between φ and ϖ grows.

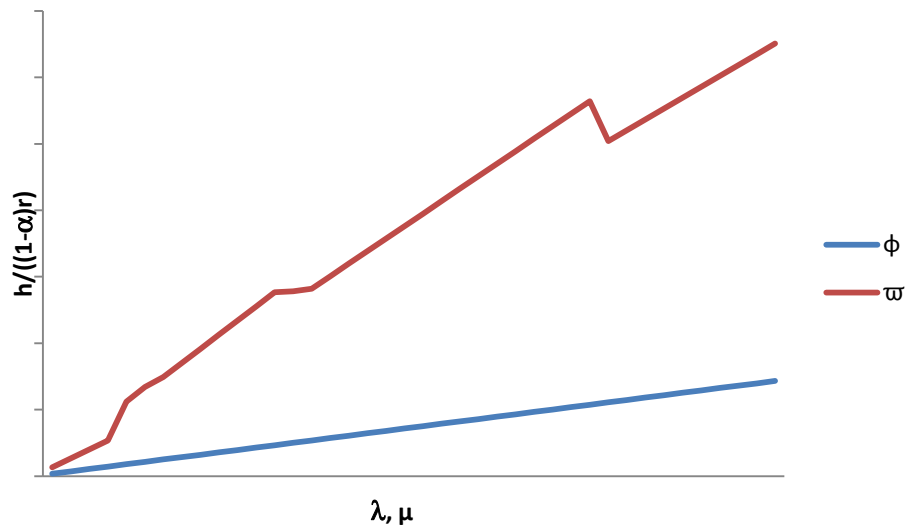


Figure 32. Threshold curves under same rate increases in λ and μ ($r=100$, $h=25$, $\alpha=0.5$, $c=5$, $\rho=2$).

5.4 Pure Exchange vs. Mixed Model Comparison

Direct comparison with the two models will help us to understand better the insight captured in the observations of our previous section. The two models agree in a significant number of cases where both models suggest the same solution. As stated in the gain functions G depends on S . The dependent variable axis in Figure 33 indicates the

gain denoted by $G(S^*)$ in the Mix Model, against the corresponding $G(S^*)$ in the Exchange Model on the abscises axis.

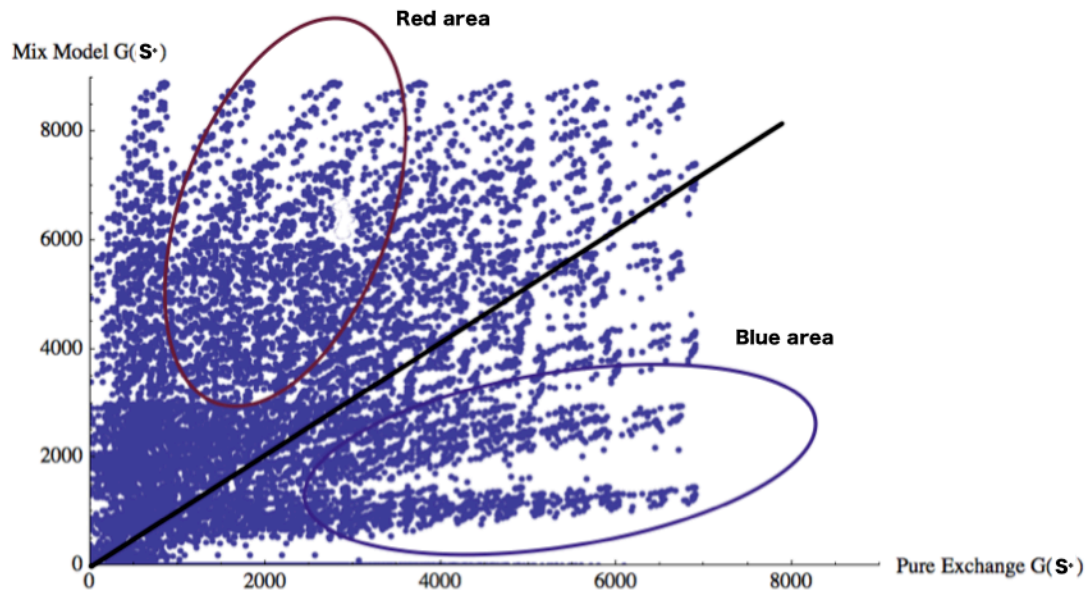


Figure 33. Gains Mixed vs. Exchange Models evaluated on I^*

Although the figure is busy, it gives an idea where the 17,000+ combinations of arrivals, service rates and other parameters lie. There are 3 identifiable areas that emerge which we will comment. Looking to the figure, it allows the observation of two distinctive regions coming from the results from both models, each model has a whole optimal region. First region, encircled by an ellipse see red area in figure 33, all of these cases, above the line that would be demarked by a diagonal black straight line represent those cases where the Mix Model will bring better results. Along the line, the second area, the two models deliver similar solutions as commented and observed in the previous section. The region in the bottom represented by the ellipse in the blue area, under

identical business conditions, the Pure Exchange model delivered more gains than the Mix Model.

As expected, the gain increases with inventory available for exchange. It is important to observe that the gain is limited after a number of pieces in inventory to a maximum gain, then if business conditions expressed by arrival rate, service rate, cost, price and premium assume certain values, additional inventory will actually be detrimental to the gains.

One obvious factor that determines the trade-off between the pure exchange and mixed service policies is the premium that the customers pay for the exchange service, namely $(1-\alpha)r$.

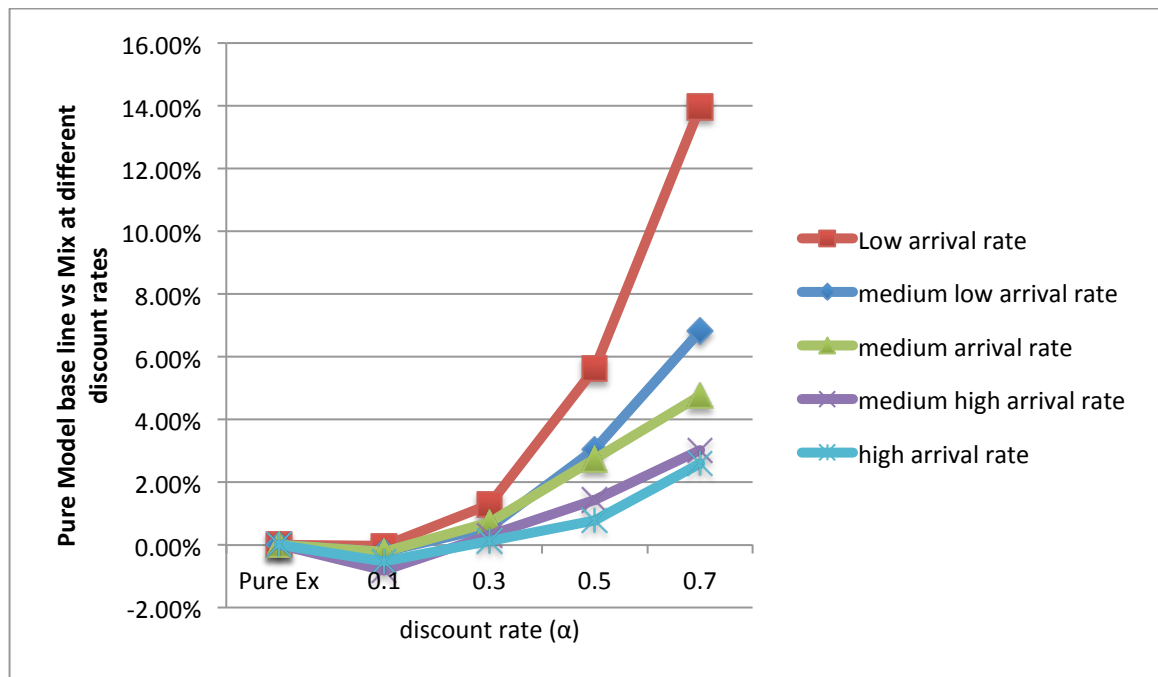


Figure 34. Trade-off between Pure and Mix Policy (λ, μ, h, c constant)

As such, as α increases the incremental benefit of the exchange service diminishes. In one extreme, as α approaches to 1, the exchange service becomes

indistinguishable from the conventional service. Clearly, then, there is no need to carry any inventory which is costly. On the other hand, when α is too small, conventional service brings very low income while keeping the system busy. In this case, the firm should focus on the exchange service by not accepting customers for non-exchange services.

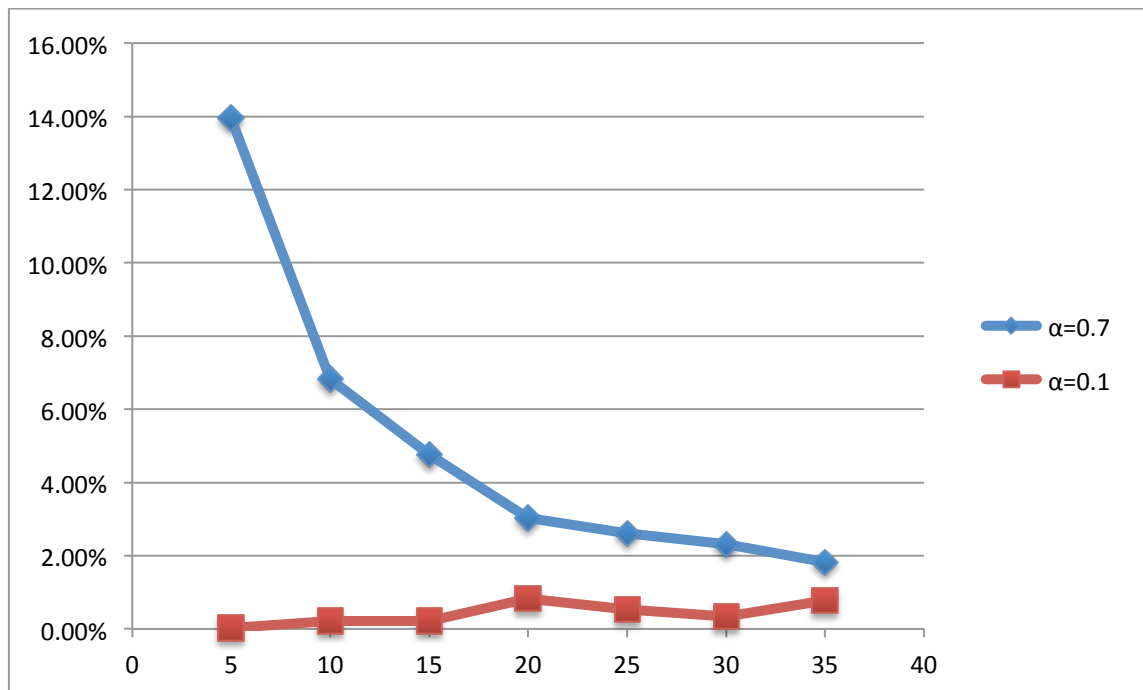


Figure 35. Effect of ΔG_{Mix}^* vs. G_{Pure}^* at high and low levels of discount.

Consequently, given that all other parameters are equal, we can easily conclude that there is a single threshold for α , above which the firm is better off with the mixed policy. The pure exchange policy is preferable if the opposite is true. See for example figure 34 to illustrate the threshold holding constant different values of $\lambda, \mu, h,$ and c going from $\alpha = 0.1$ to 0.7 , you can observe that Pure exchange performs at 0.1 , even though in this particular case the hold cost is set at the high value of 70 .

The insight then is that indeed, a policy where holding inventory is prescribed could be useful if a proper Premium is set for such a service, under a given business traffic density, and will help the decision maker to obtain additional gains otherwise not available.

For small λ , as observed in figure 35. mixed policy generates higher profits as it does not require risk of carrying inventory which faces low turnover. As λ grows pure exchange becomes more profitable as the firm can be focused on more revenue generating exchange services. The overall gains become close at high λ as α gets bigger. See figure 36 to see a similar effect caused by increasing cost. When λ is small the inventory levels are close. When it is neither small nor large, the mixed policy leads to higher inventory levels.

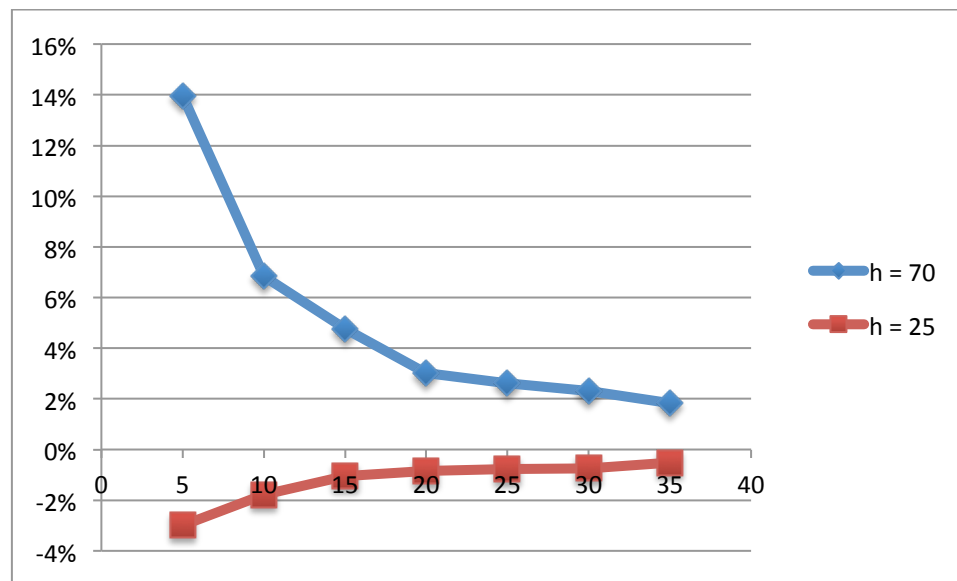


Figure 36. ΔG_{Mix}^* vs. G_{Pure}^* for different arrival rates and high and low cost levels

The risk of carrying inventory is alleviated by the opportunity of receiving income from regular services. In the pure exchange case, as arrival rates get too big the inventory

count converges to the number of servers whereas in the mixed policy it drops all the way down to zero. This situation creates advantage towards the pure exchange policy. So the bottom line is, the pure exchange policy is especially appealing when the demand is relatively high or low service levels are in place.

As the service rate increases, the mixed policy converges to the pure exchange policy.

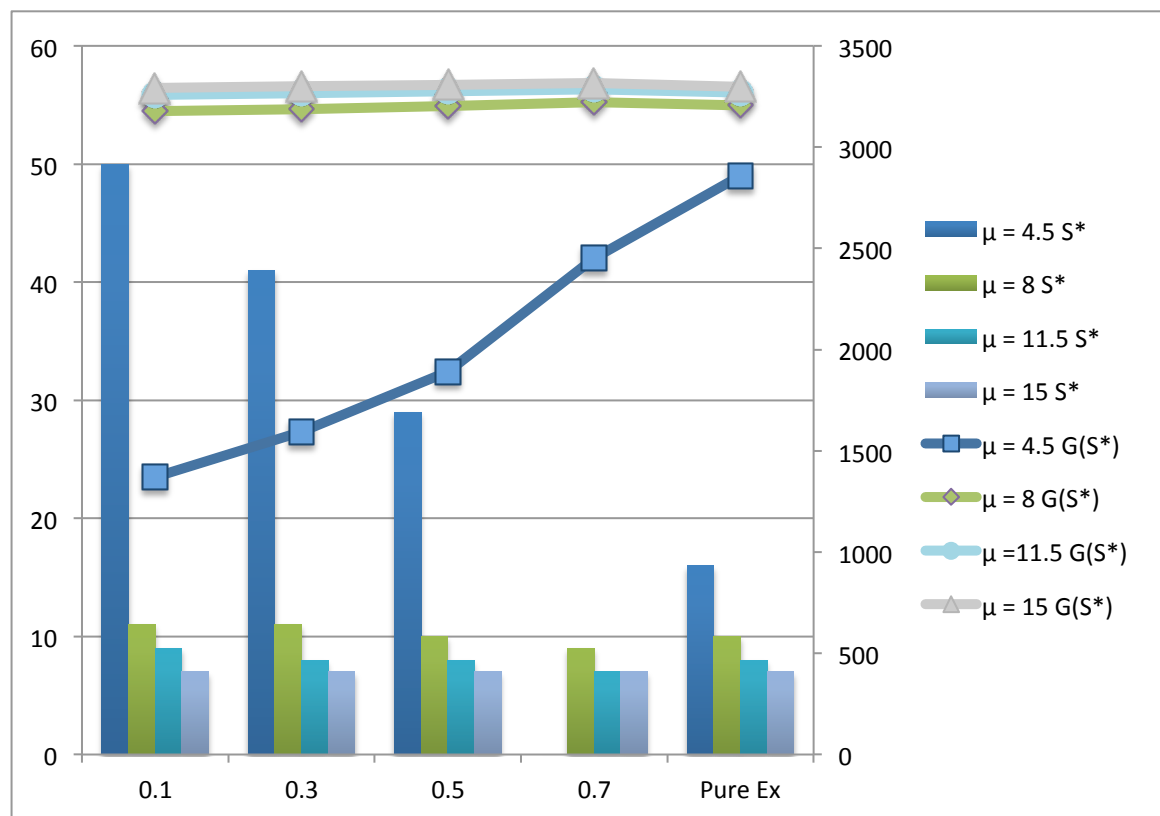


Figure 37. Service rate vs alpha Mix Model and Pure Ex S^* and $G(S^*)$

Under low service, see figure 37, pure exchange may have an edge against the mixed service since the system load becomes large. This is consistent with the case of large arrival rates. In both cases the system load is high. Consequently, we can conclude that the pure exchange policy is especially attractive when the offered load for the system is

high (under high demand and/or slow service). Consequently pure exchange seems like a good option when service is either very slow or very fast. In between it depends on α . In general, mixed policy results with higher inventory levels except when service is fast when pure exchange may result with higher inventory under high α conditions.

In general, the inventory level is higher in the mixed policy with the increase in the number of servers. When α is sufficiently small the pure exchange is better with lower number of servers. In fact, in general, the difference between the mixed strategy gains and pure exchange gains increases with the number of servers. That is as the firm has more servers available mixed strategy becomes more attractive. This is consistent with above observations as the system load gets lower.

5.5 Discussion on the Order Acceptance with Walk-in Model

This model provides, from Proposition 1 and 6, three readily available insights, when parameters are provided and the amount of inventory to maximize gain is determined. Although we only have a closed form solution when $S > c$, a simple line search on Eq. (82) delivers the optimal amount for $S \leq c$.

When $\mu > \lambda$, the range of $\rho < 1$, the model estimates what quantities (S landing gears, by reminding the reader this was our particular motivation for the model) it would be better to stock and deliver for exchange at a premium.

For the case with $\rho > 1$ there is full capacity and enough business at hand to spread the cost, thus the model suggests the gain will be at most the in house repair capacity, as no trade off between stock cost and capacity wasted exists.

For the case with $S > c$ the mixed model provides results for those cases where additional servers come into play. In such cases S could be found as an optimal result directly from Eq. (103) with $S \in \text{reals}$.

The application of the model helps to support decisions for exchange or repair. The model is especially tuned to the typical business dynamic for high value, high margin items, whose arrival rate resembles the dynamic portrayed in the Pure Exchange Model. Optimal S increase as premium increases, that is $1 - \alpha$, but up to a point delimited by carrying cost, price and repair capacity—represented by variables, h , r , c , and z in the model.

One implication of the model for the Pure Exchange is that inventory is limited and only a number of walk-ins will be accepted. All other clients must be rejected once the exchange inventory is gone. Those accepted will have immediate access to a repaired part. On the other hand, with the Mixed Model representing repair in house, all clients are accepted, although all of them must face additional waiting until the repair is completed.

If Pure Exchange is the only policy chosen, it is recommended only for businesses with the highest margin, as premium will be a major player on the model, the restricted capacity either because low service rate or high arrival rate as compare to capacity tend to favor the results on the Pure Exchange model as reviewed, as such those conditions make for a superb application of the model. High arrival rates, coupled with high service rate, low margin business should apply mixed models better as another alternative available, as the more capacity the business has the better the Mixed Model will perform.

As more and more businesses, try to minimize their asset foot print, either by subcontracting capacity or by reducing overall investment, a pure exchange strategy

might be a good fit for the business, as the model will help the decision maker compose appropriate response to the business conditions, with certitude of the outcome. Also the implementation allows for daily or weekly review of the policy if the frequency of the business so requires as the computation for policies is straight forward and fast.

Chapter 6. Comments and Conclusions

The present work is the foundation to continue this line of research, looking for policy solutions to order acceptance problems, in the variants of using dynamic premiums in combination of differentiated services, to provide to customers.

The policies uncovered so far, allowed the decision-making regarding accepting small jobs and when faced with uncertain conditions to receive more profitable big jobs.

In our three cases we have gone from:

1) representing a constrained reality to understand the monotonicity and prove it, when a decision needs to be made and we found you can have a unique maximizer and an optimal result, setting the price for a big deal vs. small jobs.

2) We then extended our work to find an optimal way to set up the premiums of the big job, under small job conditions and find the optimal pricing solutions based on the arrival rates for the small and big jobs, their differentiated service rates, and the number of customers waiting. This allows the decision-maker to have a set of premiums that can be charged taking into consideration the loss of the steady business and the price sensitivity.

3) In our final model, we took our knowledge on the previous models and developed a optimal solution to the question of whether or not, when facing big job conditions, would it be adequate to hold inventory, and in choosing to do so, in which optimal quantity. This model could help us, to set up a trade off with different assets—in the utilization of the scarce monetary resources, between when to hold inventories vs. capacity for example as was in our demonstration. Our main finding and contribution is that under given business conditions, might be advisable to hold inventory and if when the decision

maker needs to resort to a mix usage, of inventory and capacity, how the policy can be implemented with optimal overall gain results.

Given the fact these order acceptance policies apply to different settings, we like to think we still have a great challenge and opportunity ahead to continue bringing optimal solutions to order acceptance problems.

Chapter 7. Next Steps

During the execution of this line of research, we have encounter multiple instances where the business dynamics, specially the arrival rates, become very sparse, and in some instances showed, skewedness of the values, towards low values, with sudden picks.

When the demand or arrivals are represented by multiple number of zeros, high variability and nervousness, the exponential distribution models whose results are contained in this work, might not apply, due to their generalization (Ross, 2003) they are too ample to ensure a narrow solutions as the necessity to have large t on the periods to ensure the normal distribution convergence discussed on the chapter 5, page 326 of the Ross probability book.

So our work applies, as long as those conditions are in place since the markov theory is supported by this important factor.

The walk-in we so far have analyzed in this work, and some of the order acceptance policies, could be considered lumpy, if certain conditions are in place, yet those conditions must ensure large t periods and the normal distribution convergence.

Doing a relatively modest Literature Review in the subject, there are some pointers of what is considered, by most authors, a random and lumpy phenomena—by following the definition of long tail and multiple instances where you find 0 as value, one such definition proposes lumpy as the point where there is equal probability than non-zero and zero values showing up on the frequency distribution (Cornacchia & Shamir, 2012)—since a natural extension of these work directly applies to situations where customer demand is not frequent, no-shows and similar events could also be treated as lumpy demand events in our models—to be able to get useful insights.

Such insights might be as simple as tests for statistical modeling to fine tune parameters for solution applications.

Also, during the review it was our impression there are several sources for lumpiness. The first source for lumpiness lies in the decision criteria that sets the size for the order driven by when the when the product or service will be needed. This customer demand is not frequent. (Donald S. Allen, 1996)

Second source of lumpiness, might be coming from systematic patterns, people unconsciously executes during a period of time, for example the time when e-mails are read, or send, the placement of orders during the day, with unconscious decision driven by work schedules, or work place management. (Malmgren, Hofman, Amaral, & Watts, 2009)

Third source for lumpiness, this one is associated the probability that a certain timing will be lapsed before an order at given price, the delivery time actually offered, and monetary reserves of the ordering party at the time of the order, with this factors in addition to the typical factors of lumpiness considered in the spare parts and maintenance environment.(Kukreja & Schmidt, 2005)

Even from our own observation in setting the pricing sensibility function (see for example page 23 for our treatment for $e^{-\beta p}$). The pricing function in itself another exponential, frequently used to describe customer reaction to prices.

Many of these authors have a consensus in that, all the functions representing the phenomena, are similar in nature, and that most if not all of them can be represented by poisson arrivals. Also, most of the authors concluded, the variables in their works could be treated as compound poisson arrivals.

So for us then has become clear that lumpiness, if all of these three sources are exponential can then all be represented by poisson arrivals, and compound poissons to model its variability and nervousness.

With that we have come to the idea to entertain the questions: Is there a model we can use to describe mass probability distribution for lumpy events where few and sparse events might not have long histories? Can they be explained by compound poisson, poisson arrivals and exponential functions? This we wonder: Can they then, be represented by the long tail distribution associated with Yule-Simon and all of them parameters can then be represented by a compound poisson?

As such we continue to look for solutions on the Order Acceptance Policy and the “Street Walker” tradition with models and parameters, which typically represent lumpy demand conditions

Other research suggest two other ways to eliminate decision variables and relax some assumptions, by using data from other sources that can then turn into an input for the model, thus reducing the dependency on the forecasting of the lumpy demand. By understanding more upstream conditions, unpredictable variability can be reduced with a positive impact on the system outcomes. (Romeijnders et al., 2012)

Also, keeping in mind the similarities on equipment, parts, failure modes, if data characteristics accommodate for this possibility, that is dropping the identical assumption, could then it be relaxed in the model to help to understand the lumpy demand effects on cycle time variability.(Gallivan et al., 2002)

For the future models when doing labor planning there is the need to quantify the effects such as the weekend activity build up and the staff availability by skill. This will

relay heavily on the research done in the area of *rotables* Arts *et al.* (Arts et al., 2012), where we would like to expand the service differentiation by category.

Along the resource management line, both assumptions in our last model in chapter 4, no job is interrupted and new queue starts on small jobs when a big deal is accepted, that is all customers are drop until big deal is finish. Can latter be relaxed to allow insight on re-schedules and delivery issues.

All the demand once taken will then use the equipment and labor for a number of units of time, although if needed, the time fraction is a natural extension—the demand is associated with a customer. These additions will help to develop a very robust model for “walk-in”, unexpected arrivals and continue to provide tools for order acceptance policies in this area.

Adding the price sensitivity restriction to premiums is another opportunity area, since the mathematical work support premiums per state, this could actually limit the number of states under consideration, as the infeasibility of a too high or two low premiums will then help to restrict the space to analyze.

Furthermore, from some of the studies mentioned, we also incorporate the idea of predictable variability, which suggests that we can compensate for factors such as seasonal trends, business cycles, and customer patterns. If in the system design such data can be found, it can the be used to minimize asset usage or compensate with the appropriate capacity and of course that would be another area of interest to expand.

Just to reiterate over the point made about the relaxation of some other assumptions, by using data from other sources that can then turn into an input for

the model replace variables, thus might have an impact on reducing the dependency on the forecasting of the variables.

Also in particular to extend the work done on the Pure vs. Mix Model work, again in chapter 5, a natural extension is to make a consideration for the system capacity. For example, make a relevant policy regarding the number of orders admitted and link the acceptance to the discount, capacity and pricing decisions.

Due to the nature of the MRO business, another potential extension also falls naturally in understanding the effect of multiple products using the available servers (workstations for repair).

Bibliography

- Allen, Donald S. (1996). Aggregate dynamics of lumpy agents. Federal Reserve Bank of St. Louis.
- Allsop, R. E. (1972). Estimating the traffic capacity of a signalized road junction. *Transportation Research*, 6(3), 245-255.
- Arts, J., Flapper, S., & Vernooij, K. (2012). Aggregate overhaul and supply chain planning for rotables. (). Technische Universiteit Eindhoven.
- Aviation Safety Reporting System. (2012). ASRS Data base. Retrieved from <http://asrs.arc.nasa.gov/index.html>
- Banks, J. (2010). Discrete-event system simulation (5th ed.). Upper Saddle River: Prentice Hall.
- Bitran, G. (2003). An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3), 203-229.
- Blaisdell, C. J., Weiss, S. R., Kimes, D. S., Levine, E. R., Myers, M., Timmins, S., & Bollinger, M. E. (2002). Using seasonal variations in asthma hospitalizations in children to predict hospitalization frequency. *The Journal of Asthma : Official Journal of the Association for the Care of Asthma*, 39(7), 567-575.
- Canada. Ministry of Transport. (2009). Landing gear maintenance "best practices". *Aviation Safety Letter*, (3/2009), 25-26.
- Chen, C., Yan, S., & Chen, M. (2010). Short-term manpower planning for MRT carriage maintenance under mixed deterministic and stochastic demands. *Annals of Operations Research*, 181(1), 67-88. doi:10.1007/s10479-010-0689-y
- Çil, E. B., Örmeci, E. L., & Karaesmen, F. (2009). Effects of system parameters on the optimal policy structure in a class of queueing control problems. *Queueing Systems*, 61(4), 273-304. doi:10.1007/s11134-009-9109-x
- Clark, Peter L. (2010). Dirichlet series. *University of Georgia*, , 1-18.
- Clark, A. J., & Scarf, H. (2004). Optimal policies for a multi-echelon inventory problem. *Management Science*, 50(12 SUPPL), 1782-1795.
- Cohen, M., & Wille, J. (2006). Implications for service parts management in the rapidly changing aviation MRO market.
- Cornacchia, E., & Shamir, J. (2012). Mastering lumpy demand. USA: Tools Group.

- Iglehart, Donald L. (1964). The dynamic inventory problem with unknown demand distribution. *Management Science (Pre-1986)*, 10(3), 429.
- FAA. (2012). STC certification definition. Retrieved from http://www.faa.gov/aircraft/air_cert/design_approvals/stc/
- Federal Aviation Administration (Ed.). (2012). Aviation maintenance technician handbook - general (Latest Edition ed.). USA: Federal Aviation Administration.
- Feng, Y., & Pang, Z. (2010). Dynamic coordination of production planning and sales admission control in the presence of a spot market. *Naval Research Logistics*, , NA. doi:10.1002/nav.20398
- Gallego, G., & van Ryzin, G. (1994). Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8), 999-1020. doi:10.1287/mnsc.40.8.999
- Gallivan, S., Utley, M., Treasure, T., & Valencia, O. (2002). Booked inpatient admissions and hospital capacity: Mathematical modelling study. *British Medical Journal*, 324(7332), 280-2.
- Gans, N., & Savin, S. (2007). Pricing and capacity rationing for rentals with uncertain durations. *Management Science*, 53(3), 390-407.
- Ghobbar, A. A., & Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: A predictive model. *Computers and Operations Research*, 30(14), 2097-2114.
- Ghoneim, H. A., & Stidham Jr., S. (1985). Control of arrivals to two queues in series. *European Journal of Operational Research*, 21(3), 399-409.
- Gross, D., & Harris, C. M. (1998). Fundamentals of queueing theory (3rd ed.). New York: Wiley.
- Hillston, J. (2005). Process algebras for quantitative analysis. 239-248.
- Ishii, N., Takano, Y., & Muraki, M. (2013). An order acceptance strategy under limited engineering man-hours for cost estimation in Engineering–Procurement–Construction projects. *International Journal of Project Management*
- Isken, M. W., Ward, T. J., & Littig, S. J. (2011). An open source software project for obstetrical procedure scheduling and occupancy analysis. *Health Care Management Science*, 14(1), 56-73.

- Johansen, S. G. (1994). Optimal prices of an M/G/1 jobshop. *Operations Research*, 42(4), 765-774.
- Jones, W. B., Thron, W. J., & Waadeland, H. (1982). Analytic theory of continued fractions: Proceedings of a seminar-workshop held at loen, norway, 1981. Berlin ; New York: Springer-Verlag.
- Kukreja, A., & Schmidt, C. P. (2005). A model for lumpy demand parts in a multi-location inventory system with transshipments. *Computers and Operations Research*, 32(8), 2059-2075.
- Lewis, M. E., Ayhan, H., & Foley, R. D. (2002). Bias optimal admission control policies for a multiclass nonstationary queueing system. *Journal of Applied Probability*, 39(1), 20-37.
- Lippman, S. A., & Ross, S. M. (1971). The streetwalker's dilemma: A job shop model. *SIAM Journal on Applied Mathematics*, 20(3), 336-7.
- M Bichler, J Kalagnanam, K Katirciglu, & A J King. (2002). Applications of flexible pricing in business-to-business electronic commerce. *IBM Systems Journal*, 41(2).
- Malmgren, R. D., Hofman, J. M., Amaral, L. A. N., & Watts, D. J. (2009). Characterizing individual communication patterns. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France. 607-616.
- Miller, B. L., & Buckman, A. G. (1987). Cost allocation and opportunity costs [2]. *Management Science*, 33(5), 626.
- Millhiser, W. P., & Burnetas, A. N. (2013). Optimal admission control in series production systems with blocking. *IIE Transactions*, 45(10), 1035.
- Muckstadt, J. A. (2005). Analysis and algorithms for service parts supply chains. New York, NY: Springer.
- Newsmax. (04/02/2011). Grounding planes likely cost Southwest millions. News clip.
- Nishimura, S. (1982). Monotone optimal control of arrivals distinguished by rewards and service time. *Japan Operations Research Society*, 25, 217.
- Örmeci, E. L., & Burnetas, A. (2005). Dynamic admission control for loss systems with batch arrivals. *Advances in Applied Probability*, 37(4), 915-937.

- Pachon, J., Erkoç, M., & Iakovou, E. (2007). Contract optimization with front-end fare discounts for airline corporate deals. *Transportation Research Part E: Logistics and Transportation Review*, 43(4), 425-441.
- Regattieri, A., Gamberi, M., Gamberini, R., & Manzini, R. (2005). Managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management*, 11(6), 426-431.
- Robinson, D. (1978). Optimizing of priority queues: A semi-markov decision chain approach. *Management Science (Pre-1986)*, 24(5), 545.
- Romeijnnders, W., Teunter, R., & Van Jaarsveld, W. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research*, 220(2), 386-393.
- Ross, S. M. (2003). Introduction to probability models (8th ed.). San Diego, CA: Academic Press.
- Roth, M., Yanishevsky, M., & Beaudet, P. (1992). Failure analysis of aircraft landing gear components. *Failure Analysis Techniques and Applications (Proceedings of the First International Conference on Failure Analysis)* J.I. Dickson, E. Abramovici, and N.S. Marchand, Ed., ASM International, USA.
- Simchi-levi, D., & Caro, F. (2012). Optimal static pricing for a tree network. *Annals of Operations Research*, 196(1), 137-152.
- Weber, R. R., & Stidham, S. (1987). Optimal control of service rates in networks of queues. *Advances in Applied Probability*, 19(1), 202-218.
- Wen, U., & Chen, Y. (2010). Dynamic pricing model for substitutable products. *International Journal of Operations Research and Information Systems*, 1(1), 35-51.
- Wu, C. (2003). Flight schedule punctuality control and management: A stochastic approach. *Transportation Planning and Technology*, 26(4), 313-330.
- Yoon, S., & Lewis, M. E. (2004). Optimal pricing and admission control in a queueing system with periodically varying parameters. *Queueing Systems*, 47(3), 177-199.
- You, P., & Wu, M. (2007). Optimal ordering and pricing policy for an inventory system with order cancellations. *OR Spectrum*, 29(4), 661-679.
- Ziya, S., Ayhan, H., & Foley, R. D. (2008). A note on optimal pricing for finite capacity queueing systems with multiple customer classes. *Naval Research Logistics (NRL)*, 55(5), 412-418.