Essays in Service Operations Management

by

Vincent W. Slaugh

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at
Carnegie Mellon University
Tepper School of Business
Pittsburgh, Pennsylvania

Dissertation Committee:

Sridhar R. Tayur (Chair)	R. Ravi
Alan Scheller-Wolf	Onur Kesten
Bahar Biller	M. Utku Ünver

July 14, 2015

Abstract

The first two chapters of this thesis focus on the operation of rental businesses, which provides a novel and increasingly significant application of a supply chain and service operation. According to IBISWorld industry reports, at least fifteen different rental industries in the United States have annual revenues exceeding $1 billion. Furthermore, the consumer trend of "access over ownership" has spurred the creation of new and disruptive business models such as Rent the Runway, which rents designer dresses by mail to over 2 million users. However, existing operations management literature offers little support to rental operations subject to complex demand characteristics and the loss of rental units due to customer damage or purchase.

Motivated by new and innovative rental business models, we study the operation of a rental system with random loss of inventory due to customer use. We use a discrete-time model in which the inventory level is chosen before the start of a finite rental season, and customers not immediately served in each period are lost. Demand, rental durations, and rental unit lifetimes are stochastic, and sample path coupling allows us to derive structural results that hold under limited distributional assumptions. Considering different "recirculation" rules — i.e., which rental unit to choose to meet each demand — we prove the concavity of the expected profit function and identify the optimal recirculation rule under two different models of a rental unit's state: the number of times rented out or its condition. A numerical study clarifies when considering rental unit loss and recirculation rules matters most for the inventory decision: Accounting for rental unit loss can increase the expected profit by 7% for a single season and becomes even more important as the time horizon lengthens. We also observe that the optimal inventory level in response to increasing loss probability is non-monotonic. Finally, we show that choosing the optimal recirculation rule over a commonly used policy suggests that more rental units should be added, and the profit-maximizing service level increases by up to six percentage points.

The second chapter extends our rental model to include the problem of admission control through accepting and rejecting reservation requests. We use a stochastic model of a rental operation to study the problem of whether the firm should accept each reservation request over the course of a rental season. In our model, the firm must balance a desire to serve more customers, thereby achieving a higher utilization of its rental assets, with the risk of being unable to serve a reservation that was previously accepted. Service variability is considered through using exponentially distributed service times, and each acceptance decision is made only with knowledge of the number of rental units that are currently busy and the list of accepted reservation requests. We discuss challenges in proving the optimal policy for admitting reservations, and propose an easy-to-implement newsvendor-style heuristic for accepting reservations. We show that the heuristic and two extensions perform well for test cases motivated by three different rental businesses, and compare its performance to bounds and simple heuristics. Furthermore, numerical results reveal that increasing the notice time — i.e., the time between when customers make a reservation request and service begins — decreases the expected profit.

The third chapter of the thesis addresses both strategic and tactical policy questions for the operation of a statewide adoption network, which matches children in state custody with prospective adoptive families. The Pennsylvania Adoption Exchange (PAE) helps case workers who represent children in state custody by recommending prospective families for adoption. We describe PAE's operational challenges using case worker surveys and analyze child outcomes through a regression analysis of data collected over multiple years. A match recommendation spreadsheet tool implemented by PAE incorporates insights from this analysis and allows PAE managers to better utilize available information. Using a discrete-event simulation of PAE, we justify the value of a statewide adoption network and demonstrate the importance of better information about family preferences for increasing the percentage of children who are successfully adopted. Finally, we detail a series of simple improvements that PAE achieved through collecting more valuable information and aligning incentives for families to provide useful preference information.

Acknowledgments

I am deeply grateful to Sridhar Tayur, Alan Scheller-Wolf, and Bahar Biller for their support and valuable advice as my advisers and collaborators. The opportunity to learn from them has been a dream realized and the highlight of my time in Pittsburgh. I have great respect for not just their skills as researchers, but also for their virtuous life outlook and proclivity for investing in those around them. I also thank my other co-authors, Mustafa Akan, Onur Kesten, and Utku Ünver, for their efforts to guide our project as we explored a new and important topic. Finally, I thank R. Ravi for serving on my committee and for providing helpful feedback.

I also wish to thank Jane Johnston, Pennsylvania Adoption Exchange Division Manager, and the managers and staff of the Pennsylvania Statewide Adoption and Permanency Network for their extensive support of this project. I readily came to admire their dedication to help children in need and their perseverance in finding families for those children.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis, we study two innovative service operations. In the next two chapters, we use classical methods from stochastic inventory theory to study a relatively new class of business models, online rental operations. In Chapter 4, we utilize the relatively new field of market design to study the ancient practice of child adoption.

1.1 Rental Business Models

Advances in online commercial models have produced a new generation of innovative businesses built upon renting goods. For an increasing variety of products, the promise of flexibility and affordability has led to rental businesses specializing in just about every aspect of our business and personal lives. Besides traditional rental products such as movies, cars, and hotel rooms, less common goods available to rent range from bicycles to jets, cribs to coffins, and furniture to camping gear. According to IBISWorld industry analysts, the annual revenue of fifteen different rental industries in the United States each exceeded $1 billion in 2013, while the annual revenue of each of the car, heavy equipment, and industrial equipment rental industries surpassed $25 billion.

Asking "Is Owning Overrated?," some commentators have connected the rise of the rental economy to a "a growing, post-recession movement to value experiences over possessions.(Miller 2014)" Technological advances have allowed online startups to efficiently rent items from their own inventory or serve as middlemen to connect owners and renters. Luxury goods have received particular attention as fertile ground for rental businesses that make those goods available to new customer classes. For example, Rent the Runway is a company that allows customers to rent high-fashion dresses for either four or eight days at approximately 10% of the retail price of a dress (Wortham 2009). Customers can view the selection of dresses and their availability through a website, and receive style and fit advice from Rent the Runway consultants and customer reviews. Dresses are shipped to customers and returned by mail.

The unique characteristics of rental inventory problems provide three additional challenges beyond traditional inventory problems:

1. Each unit of inventory must be tracked over time as service begins and completes for each customer. If subject to variability in the service duration or return time, the

rental unit return process introduces supply-side uncertainty.

2. Individual rental units may have qualities that evolve over time, and it may be valuable to manage which rental units are assigned to which customers.

3. Customers tend to be time-sensitive if their demand is tied to a specific calendar event, which motivates a lost sales model and introduces reservations as a means for customers to have greater certainty that their demand is fulfilled.

Chapters 2 and 3 of this thesis study important decisions that must be made before and during the rental season, respectively. In Chapter 2, we consider the critical decision regarding the number of dresses that will comprise Rent the Runway's seasonal rental inventory and how the loss of rental units during the season interacts with the inventory decision. This decision must be made shortly after pre-season fashion shows, which are several months in advance of the rental season (Binkley 2011). In Chapter 3, we introduce reservations to our model so that demand is known in advance of the service start time. Chapter 3 focuses on the admission decision; i.e., whether each reservation should be accepted or rejected.

1.2 Child Adoptions

According to the most recent report of the Children's Bureau of the US Department of Health and Human Services (2014), approximately 397,000 children in the United States are living in the foster care system, with 102,000 of them waiting for adoptive placement. In 2012, while 50,000 children were successfully adopted from foster care, approximately 23,000 were discharged due to emancipation as they reached the age of 18 without receiving a permanent home. As cataloged by Howard and Brazin (2011), numerous studies have shown that children who spend significant time in foster care or "age out" of foster care without finding a permanent family suffer from alarming levels of unemployment, homelessness, early parenthood, and incarceration. For example, Reilly (2003) reports that 41% of respondents between the ages of 18 and 25 who had aged out of the foster care system have spent time in jail. Dworsky and Courtney (2010) also find that 53% of females who had exited foster care had experienced at least one pregnancy by the age of 19, which is more than double the rate of 20% by the age of 19 for females in the general population.

Despite the growing prominence and urgency of finding families for children in state custody, little is known about how to define, analyze, and improve the family finding process. The fundamental supply-and-demand imbalance for children of different demographic characteristics was first described using an economics framework by Landes and Posner (1978). An empirical study by Baccara et al. (2014) identifies biases in preferences of prospective adoptive families for infant adoption. They show that an adopted child's desirability to prospective families depends heavily on the child's age, gender, and race, with some of the greatest disparities accounted for by the child's race. Hanna and McRoy (2011) characterize the goals of the matching process and document tools to assess the quality of a family-child match that other governmental and nonprofit organizations have developed concurrent to our efforts.

1.3 Thesis Contributions and Outline

Relatively little work has focused on the management of inventory for rental businesses, and Chapters 2 and 3 of this thesis address fundamental elements of rental systems that have not previously received attention. In Chapter 2, we study usage-based loss in which rental units retire from circulation (i.e., due to damage or customer purchase) and provide structural results for the relationship between the profit and initial inventory level. When rental units have an increasing failure rate over a finite horizon, the rules for allocating rental units to customers must be specified, and we demonstrate the optimal recirculation rule. Finally, numerical results show the value of accounting for these model elements in practice.

In Chapter 3, we are the first to study admission control for customers requesting reservations in a rental model with stochastic service times. We discuss this model and its properties to show difficulties in obtaining structural results. We provide bounds for the profit and show properties of simple heuristics. Based on insights from this analysis and numerical results, we develop more advanced heuristics that perform well compared to an optimal upper bound for known arrival epochs.

In Chapter 4, we are the first to study child welfare processes from an operations management or market design perspective. We characterize the system's operations and the challenges that it faces, as well as simple improvements implemented by a statewide adoption network. We use case worker surveys to identify challenges and receptiveness to improvements to a match suggestion algorithm used to find families for a child in state custody. Analysis of child outcome data over multiple years helps to classify factors affecting a child's chances of a successful adoption and guides weights used in the match suggestion algorithm. A discrete-event simulation indicates the value of a statewide network compared to counties operating in isolation and motivates the need to collect additional information that managers believe would be valuable for predicting whether a match would be successful. Finally, we describe changes implemented to the matching algorithm to reduce families' incentives to overstate the types of children that would be acceptable.

We conclude with a summary of insights for these three chapters in Chapter 5. We also discuss future work related to rental systems and child adoptions, as well as an application that combines capacity planning and non-profit operations management.

Chapter 2

Managing Rentals with Usage-Based Loss

With Bahar Biller and Sridhar Tayur

2.1 Introduction

Choosing the number of rental units to procure before the start of a rental season without the possibility of replenishment during the season is an important problem that many rental businesses face. For online start-up businesses, how efficiently the capital-intensive rental inventory is managed influences the need to raise additional capital and determines key metrics (e.g., the average number of rental cycles that can be performed by each rental unit) presented to potential investors. The availability of inventory can affect the reputation of the rental service and customer retention rates, and failing to rent to a customer because the rental unit was damaged by a previous customer can result in a challenging customer service encounter. Furthermore, inventory management that accounts for loss relates to key strategic decisions. For example, Rent the Runway has recognized its laundry operation as a core competency and brought it in-house. Also, while customers may not purchase the high-fashion dresses at the end of a rental, the company has introduced a subscription rental service that allows customer to buy other clothing items instead of returning them.

Despite the seemingly fundamental nature of this problem, operations management literature offers very little analytical support when lost sales and discrete time periods — natural assumptions for many rental systems — are considered. In this chapter, we analyze a single-product rental system using a discrete-time framework. We focus on the usage-based loss of rental units over a finite rental horizon during which no additional rental units may be ordered, e.g., when long procurement lead times or high administrative costs prohibit in-season reordering. In particular, we consider each rental unit to have a random lifetime, which is characterized by a general probability distribution on the number of times the unit can be rented before its retirement from the rental inventory. Our goal is to understand the role of this uncertainty arising from the usage-based loss of rental units on the management of rental inventory.

In addition to Rent the Runway, whose dresses are susceptible to both destructive incidents and wearing out over time, other rental systems face the challenge of losing inventory that can be difficult to replace in the middle of the rental season. For example, a Paris-based bicycle sharing program that began with 20,600 bicycles in 2007 had more than 8,000 bikes stolen and another 8,000 bikes severely damaged and in need of replacement within two years (Erlanger and De La Baume 2009). Inventory loss can also occur when customers exercise an option to purchase a product. Users of Redbox, an automated movie and game rental kiosk, rent a DVD for $1.50 a day. If the DVD is not returned in 17 days, then the customer pays $25.50 for the accrued daily rental charge and owns the DVD. Another example is Rent-A-Center, a company with over $3 billion in revenue in 2012 and which rents furniture, appliances and electronics to customers who can own the item if it is rented beyond a certain duration. In its 2012 annual report, Rent-A-Center states that approximately 25% of its rental agreements result in customer ownership.

Existing work supporting capacity planning for rental businesses relies primarily on queueing models. Although Poisson or compound Poisson arrival processes may adequately represent demands for some rental businesses, better choices may exist for modeling demand in rental systems characterized by discrete rental time slots. For example, business travelers occupy a hotel room for a discrete number of days and are more likely to begin renting a hotel room on Monday night than a Saturday night. At Rent the Runway, for example, whose customers primarily rent dresses for events on Fridays and Saturdays, a discrete-time demand model with a period of one week more accurately represents a customer demand pattern than a Poisson arrival process. Therefore, extending the discrete-time inventory theory to include loss of rental inventory offers an advantage for a rental system like Rent the Runway. We develop a model that makes no distributional assumptions and captures (a) operational details such as random rental unit lifetimes (with constant, increasing or decreasing failure rates) and random rental durations, (b) very general demand models with features such as seasonality, auto-correlations, and forecast uncertainty and (c) recirculation rules that are used in practice for choosing among available rental units to satisfy demands.

We make the following contributions regarding the inventory management of rental systems:

1. **Model and Framework:** To the best of our knowledge, we are the first to consider the loss of rental units according to distributions over the number of times that each unit can be rented before loss. Thus, our model includes either a state variable that represents the number of times that a rental unit has been rented out (i.e., a "count-based" model) or a state variable that represents a rental unit's condition (i.e., a "condition-based" model). It also accommodates an arbitrary demand process and general distributions for lifetime and duration of each rental unit.

2. **Structural Results**

   (a) We establish the concavity of the expected profit function in the initial inventory of rental units for geometric lifetime distributions. Not surprisingly, this structural property holds independent of the rental unit recirculation rule as the loss probability is constant over time.

   (b) For general lifetime distributions, it becomes necessary to consider the recirculation rule for allocating rental units to satisfy customer demand for both count-based or condition-based models.

(c) **Count-Based Model:** We establish the concavity of the expected profit function for the "static priority" recirculation rule; i.e., the units to be rented are prioritized according to a list that does not change over the rental horizon. We show that the concavity of the expected profit function also holds for a policy that spreads the rental load evenly over all units, allocating the rental unit that has been rented out the fewest number of times. Referring to this recirculation rule as the "even spread" policy, we prove its optimality when rental unit loss probabilities are non-decreasing in the number of times that the unit has been rented.

(d) **Condition-Based Model:** We demonstrate analogous results for the condition-based model, showing the concavity of the expected profit function for the "best-first" policy in which the rental unit in the best condition receives the highest allocation priority. Also, we prove that the best-first policy is optimal when the state transition probability matrix is totally positive of order 2, a condition that implies that the rental unit failure rate is increasing as its condition worsens.

3. **Managerial Insights from Numerical Study**

(a) Failing to account for usage-based loss of rental inventory leads to a significant reduction in the expected profit. For a 5% probability of loss each time a unit is rented, we find that ignoring the loss of rental units reduces the expected profit by 7.3% and 33.0% for a half-year and a full-year rental horizon, respectively.

(b) The optimal response to the increasing loss probability is to first increase the number of rental units, then decrease the number of rental units and finally stock zero rental units.

(c) For a rental unit lifetime distribution with increasing loss probability, the rental unit recirculation rule plays an important role according to the rate at which the loss probability increases. We focus on the count-based model, as similar results apply for the condition-based model, and compare the even spread policy to the static priority recirculation rule. Choosing the even spread policy increases the optimal initial inventory level with a corresponding increase of up to 6 percentage points in the profit-maximizing service level.

The remainder of the chapter is organized as follows. Section 2.2 reviews the rental inventory management literature. Section 2.3 introduces our rental inventory model. We establish the structural properties of this model for geometric lifetime distributions in Section 2.4 and for general lifetime distributions in Section 2.5, where we further identify the optimal rental unit recirculation rule under certain conditions. The numerical analysis follows in Section 2.6. We conclude with a summary of findings in Section 2.7. All proofs appear in Appendix A.

2.2 Literature Review

Early research on rental inventory management exclusively uses queueing models as a foundation for analysis. The initial advances in queueing theory by Takács (1962) and Riordan (1962) for the telephone trunking problem — finding the stationary probabilities of a multi-server pure loss system — have sparked two seminal papers on the problem of sizing a fleet of rental equipment. Tainiter (1964) formulates an optimization problem for $M/G/c/c$ and $G/M/c/c$ rental systems based on the limiting distributions of the system states derived by Takács (1962). The decision variable is the capacity of the rental system and the problem is studied both asymptotically and over a finite horizon. Whisler (1967), on the other hand, shows that the optimal policy structure for a rental system with lost sales, periodic reordering, and nonstationary state transition probabilities — as in Riordan (1962) — has upper critical values above which inventory should be discarded and lower critical values below which inventory should be ordered. Our work differs from these studies by its focus on the inventory decision prior to the rental season, the challenge of handling random usage-based loss of rental units and stochastic rental duration, and the use of a discrete-time model for demand representation.

The early research on rental inventory management with lost sales is followed by an extensive study of the $M/M/c$ queueing model with backlogged demands. Specifically, the problem is posed as finding the optimal number of servers to employ in a multi-server queuing system, where servers represent rental units and service time corresponds to the rental duration; see Huang et al. (1977), Jung and Lee (1989), Green et al. (2001), and Zhang et al. (2012). Motivated by the time-specific nature of customers' rentals, however, we restrict our focus to lost sales models in this chapter. Table 2.1 compares our rental inventory model to the other rental inventory models that also make the assumption of lost sales. In addition to the continuous-time rental inventory models of Tainiter (1964) and Whisler (1967) tabulated here, Papier and Thonemann (2008) build on the $M/M/c/c$ queueing model in Harel (1988), where approximations, as well as lower and upper bounds, are developed for the lost sales rate as a function of the system capacity. Extending this model to account for a compound Poisson arrival process, Papier and Thonemann (2008) conduct a stationary queueing analysis to obtain structural results for a fleet sizing problem and provide an approximation suitable for implementation. Adelman (2008) also uses an approximation of the $M/M/c/c$ or $M/G/c/c$ queueing model to study the capacity decision for a rental system. However, our work is different from this stream of research by our consideration of a discrete-time rental model with a finite rental season, random usage-based inventory loss, and an arbitrary demand model possessing the ability to capture any distributional characteristic.

In contrast to the continuous-time queueing models, Cohen et al. (1980) use a discrete-time model to represent a return process to a blood bank with the goal of determining an optimal order-up-to level in every period. Reflecting hospitals' tendency to order significantly more units of blood than needed, a constant percentage of the quantity rented by hospitals is returned to the blood bank and the rest is consumed after a rental duration of a fixed number of periods. A constant percentage of the inventory leftover at the blood bank is, on the other hand, considered to have decayed. The problem of finding the optimal inventory level under a periodic review policy is formulated as a dynamic program and an

Table 2.1: Comparison of lost sales rental inventory models.

	CONTINUOUS TIME			DISCRETE TIME		
	Taniter (1964)	Whisler (1967)	Papier and Thonemann (2008)	Cohen et al. (1980)	Baron et al. (2011)	Our Paper
Inventory Decision	One Time	Repeated	One Time	Repeated	One Time	One Time
Time Horizon	Finite	Finite	Infinite	Finite	Finite	Finite
Demand Process	IID Interarrival Times	IID Interarrival Times	Compound Poisson Stationary; Also Nonstationary	General IID	Arbitrary	Arbitrary
Rental Duration	General IID	General IID	General IID	Deterministic	General IID with a Restricted Return Process	General IID
Inventory Loss	N/A	N/A	N/A	Constant Decay	N/A	Usage-Based Random Loss

approximate solution is provided. In comparison, we examine the one-time pre-seasonal ordering problem and consider the loss of inventory as random, instead of being a constant proportion. Furthermore, we do not require the assumption of an independent and identically distributed demand process, and we allow randomness in the rental duration.

Closest to our model is Baron et al. (2011), who determine the optimal pre-season order quantity for a video rental store with lost sales but no inventory loss. In particular, Baron et al. (2011) consider a return process that is monotone; i.e., the percentage of the rental units rented in period t and returned by period n is always greater than or equal to the percentage of the units rented in period $t+1$ and returned by the same period n. The key result is the concavity of the expected number of rentals in the number of rental units procured. We are, on the other hand, the first to establish this result for a rental system with random usage-based inventory loss. We also address the issue of rental unit inventory allocation, which arises only in our rental inventory model as a result of accounting for random lifetimes of the rental units.

While we focus on allocating rental units based on their state, others focus on allocation policies for choosing among different customers based on customer class. Miller (1969) analyzes the admission decision to an Erlang loss queue when customers belong to classes based on the revenue received through service, and Savin et al. (2005) show the impact of the class-based allocation policies on the optimal fleet size. Papier and Thonemann (2010) and Levi and Shi (2011) build on this model to consider customers that request reservations in advance. Gans and Savin (2007) consider how to price admission to the queue, and Tang and Deo (2008) study the pricing decision when customers have uncertain return times. How to prioritize rental customers who pay a monthly rental subscription fee and are heterogenous in their rental duration is also the subject of research by Bassamboo et al. (2009) and Jain et al. (2015).

To analyze models in which rental unit lifetimes do not follow a geometric distribution, we use sample path analysis in a very general setting to prove the two main results of our work: the concavity of the expected profit function and the optimal rental unit recirculation rule. This approach has been used in various settings to model complexities of production and inventory systems. Examples include the number of customers and their utilities for a model with dynamic substitution by Mahajan and van Ryzin (2001) and the processing times for multi-station production lines by Muth (1979) and Tayur (1993). Also, our proofs of concavity bear similarities to that of Shanthikumar and Yao (1987) in their study of systems with multi-server stations.

2.3 Rental Model: A Sample Path Approach

In this section, we describe a sample path approach to modeling a rental inventory system that allows us to analyze the system — including rules for recirculating rental units — under general assumptions about the demand process. Motivated by the problem of selecting the number of rental units to procure before the start of a finite rental season, we begin our analysis with the following two-stage model of a single-product, discrete-time rental inventory system with lost sales. In the first stage, the size of the rental inventory is chosen to be y. Each rental unit is procured before the start of the season and has a value at the end of the rental season that depends on whether the rental unit retires from the inventory before the end of the season. A rental unit's retirement from inventory may decrease the salvage value due to damage or may increase its value if it is sold to a customer (or if a penalty is charged to the customer.) Hence, the unit procurement cost accounts for not just the purchase price, but is adjusted to also include the salvage value for a dress in "good" condition and the cost of holding the item for the duration of the rental season. In the second stage, demands occur over N periods and the units purchased in the first stage are rented to satisfy the customer demands. Each customer is assumed to rent a single unit, and for simplicity we begin by considering the case in which each rental lasts for a random duration before considering a model with random rental duration. Thus, if the duration of some rental is a periods, fulfilling the demand requires that one unit of the inventory is withdrawn for the period in which the demand is received and for the $a-1$ succeeding periods.

A critical aspect of rental inventory planning is to account for the loss of rental units. Misuse by customers, customer options to purchase rented items or simply the deterioration of the rental unit's quality over time present reasons for why a unit would be retired from the rental inventory.

The duration of each customer rental is a random variable denoted by $A_{m,i}$. We define its realization $a_{m,i}$ over rental unit m's lifetime as the rental duration for the ith demand served for $i \geq 1$ and $m = 1, 2, \ldots, y$. Each rental lasts for any number of periods between a minimum of A_{min} and a maximum of A_{max}; i.e., $a_{m,i} \in \{A_{min}, A_{min}+1, \ldots, A_{max}\}$. We assume that $A_{m,i}$ is independent and identically distributed according to a general probability mass function characterized by $h_A := \mathbb{P}\{A = a\}$, $a = A_{min}, A_{min}+1, \ldots, A_{max}$.

Each rental unit fails after a random number of rentals (i.e., its "lifetime" $\mathbf{l}_m$), with a loss probability $f_{i,a}$ corresponding to the ith rental served by unit m and a rental duration a. Upon completion of its $\mathbf{l}_m$th rental, unit m satisfies no further demands, although it might still have a salvage value that is earned at the end of the horizon or other reward in the event of loss due to customer purchase. The marginal lifetime distribution $\ell_i = Pr(\mathbf{l}_m = i)$

can then be defined recursively as

$$\ell_i = \left(1 - \sum_{j}^{i-1} \ell_j\right) \sum_{a=A_{min}}^{A_{max}} h_a f_{i,a}.$$

We note that our model accommodates correlation between the loss probability and the duration of each rental.

Demand d_n is received in period $n \in \{1, 2, \ldots, N\}$. Taken together, the demands $d_1, d_2, \ldots, d_N$ and the rental unit lifetimes $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_y$ comprise a sample path, which we denote by $\boldsymbol{\xi}$; i.e., $\boldsymbol{\xi} = \{d_1, d_2, \ldots, d_N, \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_y\}$. When rental unit loss probabilities change based on the number of times rented, we must also specify the recirculation rule γ to fully characterize the system's operation. The recirculation rule determines which rental unit is chosen given a set of available rental units for each customer served, and may affect the system's profit by changing the pattern at which rental units are lost over the rental horizon.

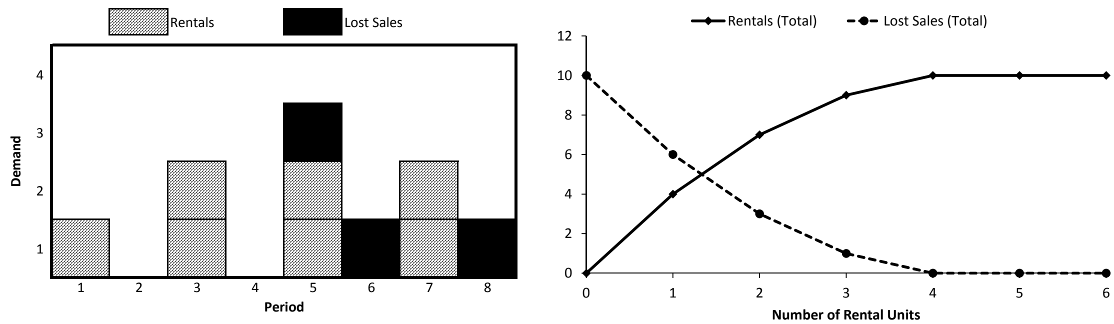
We use the notation $R_n^\gamma(y, \boldsymbol{\xi})$ for the number of units rented and $L_n^\gamma(y, \boldsymbol{\xi})$ for the number of sales lost in period n as a function of the initial inventory of y rental units and the sample path $\boldsymbol{\xi}$ of demands and rental unit lifetimes for a recirculation rule γ. For convenience, the total number of rentals and lost sales over the entire horizon are defined as $\mathcal{R}^\gamma(y, \boldsymbol{\xi}) := \sum_{n=1}^{N} R_n^\gamma(y, \boldsymbol{\xi})$ and $\mathcal{L}^\gamma(y, \boldsymbol{\xi}) := \sum_{n=1}^{N} L_n^\gamma(y, \boldsymbol{\xi})$, respectively. We also let $W_n^\gamma(y, \boldsymbol{\xi})$ denote the number of units that are successfully returned to the system in the beginning of period n and available to be rented again in that period.

Accounting for different possible rental durations, we use $R_{a,n}^\gamma(y, \boldsymbol{\xi})$ to denote of the number of rentals of duration a that begin in period n, and $\mathcal{R}_a^\gamma(y, \boldsymbol{\xi})$ denote the number of rentals of duration a that occur over the entire rental horizon. We allow $W_{a,n}^\gamma(y, \boldsymbol{\xi})$ and $Z_{a,n}^\gamma(y, \boldsymbol{\xi})$ to represent the number of rental units returned and lost, respectively, in period n after a rental duration of a periods with $\mathcal{Z}_a^\gamma(y, \boldsymbol{\xi}) := \sum_{n=a+1}^{N+a} Z_{a,n}^\gamma(y, \boldsymbol{\xi})$. The total number of rentals in a period that was previously defined can be written as $R_n^\gamma(y, \boldsymbol{\xi}) := \sum_{a=A_{min}}^{A_{max}} R_{a,n}^\gamma(y, \boldsymbol{\xi})$, with $W_n^\gamma(y, \boldsymbol{\xi})$ and $Z_n^\gamma(y, \boldsymbol{\xi})$ defined analogously. Furthermore, we let $a_{m,i}$ denote the realized rental duration of the ith demand served by the rental unit m. It is important to note that the sample path $\boldsymbol{\xi}$ now consists of not only the demand realizations d_n, $n = 1, 2, \ldots, N$, and the rental unit lifetimes $\mathbf{l}_m$, $m = 1, 2, \ldots, y$, but also the rental durations $a_{m,i}$, $i \geq 1$ and $m = 1, 2, \ldots, y$. To explicitly state the number of lost rental units in any period, we define $Z_n^\gamma(y, \boldsymbol{\xi}) := \sum_{a=A_{min}}^{A_{max}} \left(R_{n-a,a}^\gamma(y, \boldsymbol{\xi}) - W_{n,a}^\gamma(y, \boldsymbol{\xi})\right)$ as the number of rental units that would have been returned in period n but were lost.

The rental system operates for period n of the second stage as follows:

1. Of all the items rented in period $n - A$, $W_n^\gamma(y, \boldsymbol{\xi})$ units are returned while $Z_n^\gamma(y, \boldsymbol{\xi})$ retire from the rental inventory. After returns are received but before rentals are made, the total inventory available to rent out in period n is $I_n^\gamma(y, \boldsymbol{\xi}) := y - \sum_{t=1}^{n-1} R_t^\gamma(y, \boldsymbol{\xi}) + \sum_{t=1}^{n} W_t^\gamma(y, \boldsymbol{\xi})$.

2. The demand D_n is realized as d_n. If $d_n \leq I_n^\gamma(y, \boldsymbol{\xi})$, then d_n units are rented out. Otherwise, $I_n^\gamma(y, \boldsymbol{\xi})$ units are rented out. More succinctly, $R_n^\gamma(y, \boldsymbol{\xi}) := d_n \wedge I_n^\gamma(y, \boldsymbol{\xi})$, where $a \wedge b$ denotes the minimum of a and b. The rental unit recirculation rule deter-



(a) Illustration with two rental units ($y = 2$).

(b) Effect of number of rental units y.

Figure 2.1: Number of rentals and lost sales for Example 2.1 with a rental duration of two periods ($A = 2$).

mines which rental unit is allocated to satisfy each unit of demand, and consequently determines $W_n^\gamma(y, \boldsymbol{\xi})$ and $Z_n^\gamma(y, \boldsymbol{\xi})$.

3. Excess demand $L_n^\gamma(y, \boldsymbol{\xi}) := (d_n - I_n^\gamma(y, \boldsymbol{\xi})) \vee 0$ is lost, where $a \vee b$ denotes the maximum of a and b. This expression can be alternatively written as $L_n^\gamma(y, \boldsymbol{\xi}) = d_n^\gamma - R_n^\gamma(y, \boldsymbol{\xi})$.

Therefore, given the sample path $\boldsymbol{\xi}$, the dynamics of the rental system's operation can be represented recursively as follows, where $I_0^\gamma(y, \boldsymbol{\xi}) = y$ and $R_t^\gamma(y, \boldsymbol{\xi}) = 0$ for $t \leq 0$:

$$
\begin{aligned}
I_{n+1}^\gamma(y, \boldsymbol{\xi}) &= I_n^\gamma(y, \boldsymbol{\xi}) - R_n^\gamma(y, \boldsymbol{\xi}) + W_{n+1}^\gamma(y, \boldsymbol{\xi}). \\
R_{n+1}^\gamma(y, \boldsymbol{\xi}) &= d_{n+1} \wedge I_{n+1}^\gamma(y, \boldsymbol{\xi}). \\
L_{n+1}^\gamma(y, \boldsymbol{\xi}) &= d_{n+1} - R_{n+1}^\gamma(y, \boldsymbol{\xi}).
\end{aligned}
\tag{2.1}
$$

Example 2.1. *Figure 2.1 illustrates an example rental system with a demand sequence of* $\{d_1, \ldots, d_8\} = \{1, 0, 2, 0, 3, 1, 2, 1\}$ *for eight periods* ($N = 8$)*. Each rental lasts for a deterministic two periods; i.e., a unit that is rented in period* n *will next be available to be rented again in period* $n + 2$*. In this example, we assume that there is no rental unit loss. Thus,* $W_n^\gamma(y, \boldsymbol{\xi}) = R_{n-2}^\gamma(y, \boldsymbol{\xi})$ *and the recirculation rule does not matter, as rental units have infinite lifetimes. If the system would operate with only one rental unit (i.e.,* $y = 1$*), then that unit would be rented in periods* 1, 3, 5, *and* 7 *for a total of four rentals, while six units of the demand would be lost. Figure 2.1a shows how the demand is divided into rentals and lost sales for a system with* $y = 2$ *rental units. Thus, the addition of the second rental unit allows an additional unit of demand to be satisfied in periods* 3, 5, *and* 7*, so that there are now* 7 *units of fulfilled demand and* 3 *units of lost sales. Figure 2.1b shows how the number of rentals and lost sales change with the number of rental units* y*. We observe that the number of rentals is concave in* y *and that the number of lost sales is convex in* y *on this sample path. In other words, the number of additional rentals produced by one additional rental unit (i.e., the slope of the rentals curve) is decreasing in* y*.*

As the return process depends on the specific recirculation rule γ, we will describe $W^{\gamma}_{n+1}(y, \boldsymbol{\xi})$ and $Z^{\gamma}_{n+1}(y, \boldsymbol{\xi})$ as needed when referring to specific rules. To account for the return of rental units that are rented in periods $N - A + 1, N - A + 2, \ldots, N$, we define $W^{\gamma}_{N+1}(y, \boldsymbol{\xi}), W^{\gamma}_{N+2}(y, \boldsymbol{\xi}), \ldots, W^{\gamma}_{N+A}(y, \boldsymbol{\xi})$ as the returns and $Z^{\gamma}_{N+1}(y, \boldsymbol{\xi}), Z^{\gamma}_{N+2}(y, \boldsymbol{\xi}), \ldots, Z^{\gamma}_{N+A}(y, \boldsymbol{\xi})$ as the lost units in each of the corresponding periods. The total number of lost rental units is denoted by $\mathcal{Z}^{\gamma}(y, \boldsymbol{\xi}) := \sum_{n=A+1}^{N+A} Z^{\gamma}_{n}(y, \boldsymbol{\xi})$.

One way to model rental unit loss is to consider geometrically distributed rental unit lifetimes. The memorylessness of the geometric distribution leads to a constant probability of rental unit loss over time. However, if a rental unit does indeed have a higher probability of wearing out over time, then a rental unit lifetime distribution with an increasing failure rate (i.e., a loss probability increasing with the number of times the unit has been rented) would be a suitable choice. Bikes, cars and large equipment are examples of assets for which an increasing loss probability as a function of the number of rentals could be used to model the rental unit lifetime. Furthermore, lifetimes that are deterministic — when enforced by safety regulations that require their disposal after a certain number of uses — can be analyzed as a special case of an increasing loss probability.

Next, we define the revenues and costs in our models. A reward r_a is earned every time a unit is rented for a duration of a periods, and c is the unit cost of a lost sale. The salvage value of a rental unit that retires from the rental inventory during the rental season may differ from the salvage value of a unit that is still functional at the end of the season. The salvage value only depends on whether the rental unit is lost, and does not depend on the duration of the rental during which the loss occurred or the number of customers served before the customer associated with the loss. We separately define the procurement cost k^S for a unit that can be still rented at the end of the rental season (i.e., "survives") and the procurement cost k^L for a unit that has already retired from circulation (i.e., is "lost"). The relation $k^S \leq k^L$ indicates that the unit retiring from the rental inventory has been damaged. Hence, it has lost a portion of its value. The relation $k^S \geq k^L$ may, on the other hand, represent the purchase of the rental unit by the customer who is renting it as discussed in Section 1 for the rental companies Redbox and Rent-A-Center. We note that this model can also be easily extended to allow for the salvage value as a random variable, and the results hold using the expected salvage value.

To account for the cost of inventory loss in the objective function of our rental inventory model, the reduction in the salvage value of a lost rental unit $(k^L - k^S)$ is multiplied by the number of lost rental units and subtracted from the revenue as part of the profit function, which we denote by $\Pi^{\gamma}(y, \boldsymbol{\xi})$. Consequently, we obtain the profit function on any sample path as follows:

$$\Pi^{\gamma}(y, \boldsymbol{\xi}) = \sum_{a=A_{min}}^{A_{max}} r_a \mathcal{R}^{\gamma}_a(y, \boldsymbol{\xi}) - c\mathcal{L}^{\gamma}(y, \boldsymbol{\xi}) - k^S y - (k^L - k^S)\mathcal{Z}^{\gamma}(y, \boldsymbol{\xi}).$$

We are now ready to formulate the rental inventory optimization problem as the maximization of the expected profit function $\pi^{\gamma}(y) := \mathbb{E}[\Pi^{\gamma}(y, \boldsymbol{\xi})]$ subject to $y \geq 0$. We investigate the concavity of this expected profit function in the initial inventory of y rental units for geometric lifetime distributions in Section 2.4 and for general lifetime distributions in Section 2.5.

When rental duration is random, the convexity of the total number of lost sales, $\mathcal{L}(y, \boldsymbol{\xi})$, in y and thus, the concavity of the total number of rentals, $\mathcal{R}(y, \boldsymbol{\xi})$, in y, might not hold for every sample path $\boldsymbol{\xi}$. As an example, we consider the addition of two rental units to our inventory system, where the first additional unit fulfills one customer demand with a very long duration and the second additional unit fulfills several customer demands with short rental durations. In this case, the number of additional customer demands satisfied by one extra rental unit is not necessarily non-increasing in y. Therefore, we proceed by analyzing the structural properties of the expected number of lost sales and the expected number of rentals. We are the first to consider this modeling aspect simultaneously with random loss of rental inventory in the following section. It is worth noting that the random rental duration accounts for each customer's decision to keep the rental unit for a different number of periods, but it can also include the random service time needed to repair the rental unit depending on its condition upon return.

2.4 Rental Inventory Loss with Geometric Lifetime Distributions

This section considers a model in which each rental unit m experiences a loss probability of p with each rental. Specifically, the random variable $\mathbf{l}_m$, which denotes the number of times the unit $m \in \{1, 2, \ldots, y\}$ is rented before retiring from the rental inventory, follows a geometric distribution with an expected value of $1/p$. We assume that $\sum_{a=A_{min}}^{A_{max}} r_a h(a) + c \geq p(k^L - k^S)$. This condition implies that the expected benefit, $\sum_{a=A_{min}}^{A_{max}} r_a h(a) + c$, of converting a lost sale into a rental is greater than or equal to the expected cost, $p(k^S - k^L)$, of the rental unit loss. Due to the constant loss probability, the recirculation rule has no effect on $\mathbb{E}\left[L_n^{\gamma}(y, \boldsymbol{\xi})\right]$, $\mathbb{E}\left[R_n^{\gamma}(y, \boldsymbol{\xi})\right]$, or $\mathbb{E}\left[W_n^{\gamma}(y, \boldsymbol{\xi})\right]$ for any n. Therefore, we omit the superscript in the notation used in this section.

We establish the concavity of the expected profit function by first presenting a condition related to the rental return process for which the expectation of the number of lost sales, $\mathbb{E}[\mathcal{L}(y, \boldsymbol{\xi})] := \sum_{n=1}^{N} \mathbb{E}[L_n(y, \boldsymbol{\xi})]$, is convex. Correspondingly, the expectation of the number of rentals, $\mathbb{E}[\mathcal{R}(y, \boldsymbol{\xi})] := \sum_{n=1}^{N} \mathbb{E}[R_n(y, \boldsymbol{\xi})]$, is concave in the initial inventory of y rental units for this condition.

Lemma 2.1. *If the expected number of rental units returned over the rental horizon*

$$\mathbb{E}\left[\sum_{t=1}^{n} \sum_{a=A_{min}}^{A_{max}} W_{a,t}(y, \boldsymbol{\xi})\right]$$

is concave and non-decreasing in y for $n = 1, 2, \ldots, N$, then the expected number of lost sales $\mathbb{E}[\mathcal{L}(y, \boldsymbol{\xi})]$ is convex and non-increasing while the expected number of rentals $\mathbb{E}[\mathcal{R}(y, \boldsymbol{\xi})]$ is concave and non-decreasing in y.

We prove Lemma 2.1 using recursive substitution of the forward differences of the state equations (2.1). After eliminating all $\Delta R_n(y, \boldsymbol{\xi})$ terms, we can see that the concavity of the return process is a sufficient condition for the concavity of the expected number of rentals. We then use an argument by induction and Lemma 2.1 to show that the expected number of returns is indeed concave in y, which implies that the expected profit function is also concave in y.

Proposition 2.1. *When rental unit lifetimes are geometrically distributed, the expected profit $\pi(y)$ is concave in y for any rental unit recirculation rule and $y \geq 0$.*

2.5 Rental Inventory Loss with General Lifetime Distributions

When the lifetimes of the rental units follow a general distribution, the number of rental units returned in any period n may depend on the policy used to choose among available rental units to satisfy the demand in previous periods. Also, the number of rentals $\mathcal{R}(y, \boldsymbol{\xi})$ and the number of lost sales $\mathcal{L}(y, \boldsymbol{\xi})$ might not necessarily be concave and convex, respectively, in y due to a rental unit that has a particularly long or short lifetime. Therefore, we investigate whether it is possible to establish the concavity of the expected profit in the initial inventory of y rental units.

Because we have not yet found a direct algebraic proof, we compare sample paths via coupling, as described in Chapter 4 of Lindvall (1992). A coupling approach allows us to compare the value of an additional rental unit in two systems that differ only in the number of rental units. Due to the rental unit lifetime distributions and the recirculation rule, analysis of the change in the expected number of rentals would otherwise be extremely difficult. Our approach uses the following steps:

1. Establish demand values $d_1, d_2, \ldots, d_N$, which do not require any distributional assumptions.

2. Operate the system with y rental units, each of which has a lifetime $\mathbf{l}_m$, $m = 1, \ldots, y$ to indicate the number of customers that can be served by each rental unit before it retires from circulation. The ith demand, $i \geq 1$, served by rental unit m has a duration of $a_{m,i}$ periods.

3. Add an additional rental unit — the $(y+1)$st unit to the system — that has a lifetime $\mathbf{l}'$ and serves demands with durations $\{a'_1, a'_2, \ldots\}$. To be clear, the system has rental units with lifetimes $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_y, \mathbf{l}'$.

4. To the system described in Step 2 (i.e., ignoring Step 3), add a $(y+1)$st unit that has a lifetime $\mathbf{l}_{y+1}$ and serves demands with durations $\{a_{y+1,1}, a_{y+1,2}, \ldots\}$.

5. To the system described in Step 4, add an additional rental unit — the $(y+2)$nd unit — so that the system has rental units with lifetimes $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_y, \mathbf{l}_{y+1}, \mathbf{l}'$. This additional rental unit has the same lifetime $\mathbf{l}'$ and serves demands with same durations $\{a'_1, a'_2, \ldots\}$ as the additional unit added to the system in Step 3.

For notational convenience, we define $\boldsymbol{\xi}(y)$ as the sample path consisting of the demands $d_1, d_2, \ldots, d_N$ of all N periods, the rental unit lifetimes $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_y$, and the demand durations $\{a_{m,1}, a_{m,2}, \ldots\}$ for $m = 1, 2, \ldots, y$, as well as the lifetime $\mathbf{l}'$ and rental durations $\{a'_1, a'_2, \ldots\}$ for an additional rental unit. For example, $\boldsymbol{\xi}(y)$ and $\boldsymbol{\xi}(y+1)$ contain all of the sample path information necessary to analyze the systems described in Steps 3 and 5, respectively.

We consider two types of decisions for rental unit allocations. First, we examine a "count-based" rental unit state in which the allocation decision is based on the number of times that each unit has been rented. Then, we study a "condition-based" rental unit state in which the allocation decision is based on the current state of each rental unit. Each of these models may be relevant for Rent the Runway. Specifically, the dress's physical condition may not be observed — requiring a count-based model — if it is not carefully inspected or if the cause of a dress failure is difficult to observe as the dress's condition degrades. For instance, a zipper may be more likely to fail over time even if indications of impending failure may not be observed. On the other hand, a dress's physical condition may be observed if it relates to the condition of the fabric. Satin dresses are susceptible to developing minor damage to individual threads due to their loose weaves, and the repeated ironing of silk taffeta dresses may cause them to lose their ideal appearance around pleats and seams. A condition-based model would then be more appropriate for this setting.

2.5.1 Count-Based Rental Unit State

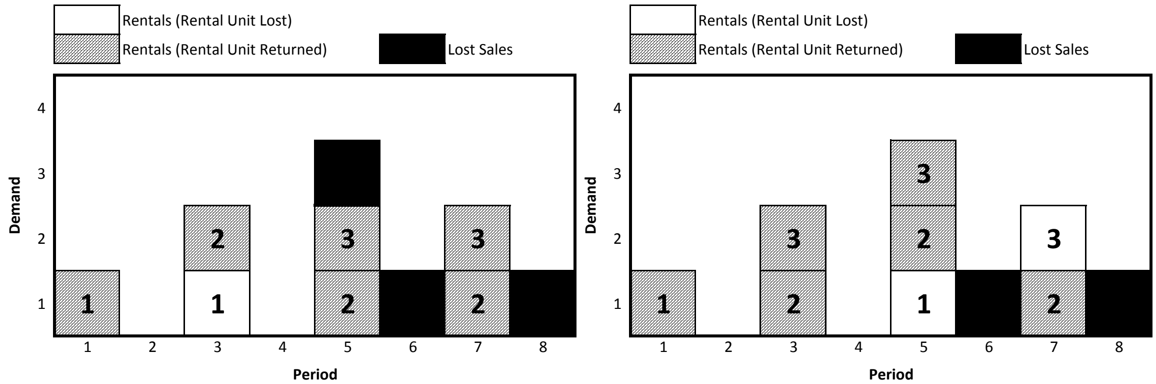
For the analysis in this section, we assume that $\sum_{a=A_{min}}^{A_{max}} r_a h_a + c \geq (k^L - k^S)\ell_i$ for $1 \leq i \leq N/A$. This condition implies that the expected benefit of an additional rental to a customer (i.e., $\sum_{a=A_{min}}^{A_{max}} r_a + c$) is greater than or equal to the reduction in the salvage value due to loss multiplied by the loss probability.

We denote by $\mathcal{C}$ the set of all policies for choosing an available rental unit to satisfy a demand based on knowledge of the number of times that each unit has been rented. In any period n, a sequence of $R_n(y, \boldsymbol{\xi})$ allocation decisions must be made. For each decision, we allow $\mathcal{A}$, $\emptyset \subset \mathcal{A} \subseteq \{1, 2, \ldots, y\}$, to denote the set of available rental units from which the rental unit to allocate, m^γ, is chosen. At the time of the decision, the number of times that each rental unit has been rented is η_m for $m = 1, 2, \ldots, y$; for the chosen rental unit m^γ, the rental unit is removed from $\mathcal{A}$ and η_m^γ is increased by one. If the rental unit is lost (i.e., $\mathbf{l}_{m^\gamma} = \eta_m^\gamma$), the rental unit is not returned to $\mathcal{A}$. Otherwise, it is returned to $\mathcal{A}$ at the beginning of period $n + a_{m^\gamma, \eta_{m^\gamma}}$.

In this section, we examine two recirculation rules: the even spread policy (denoted by "ES") and the static priority policy (denoted by "SP"). In the even spread policy, each demand is served by an available rental unit that has been rented out the fewest number of times among all available rental units. We note that the priority is assigned to rental units in the order of increasing hazard rate under the even spread policy. The static priority policy, on the other hand, allocates rental units according to a priority list that does not change over the course of the rental horizon. These two policies can be defined as selecting some rental unit m^{ES} or m^{SP} such that

$$m^{ES} \in \arg\min_{m \in \mathcal{A}} \eta_m,$$
$$m^{SP} \in \arg\max_{m \in \mathcal{A}} \eta_m.$$

Because the static priority rule is easier to analyze, we begin by proving that policy's structural properties and then consider the even spread policy.

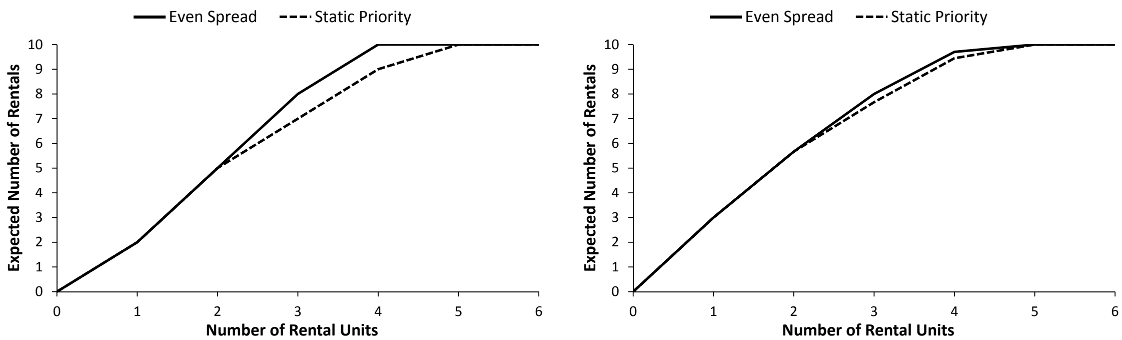


(a) Static Priority Policy (b) Even Spread Policy

Figure 2.2: Rental unit recirculation schemes for Example 2.2 with $y = 3$. The number inside a box identifies the rental unit that satisfies a demand.

Example 2.2. *Figure 2.2 shows how different rental unit recirculation schemes can affect the number of rentals and lost sales for an example sample path $\boldsymbol{\xi}$. Using the same demand values as in Example 2.1, we now state the lifetimes of available rental units as $\{\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_5\} = \{2, 4, 3, 4, 2\}$. For $y = 1$ and $y = 2$ with $\mathbf{l}_1 = 2$ and $\mathbf{l}_2 = 4$, both the even spread and static priority policies satisfy the same number of demands; i.e., $\mathcal{R}(1, \boldsymbol{\xi}) = 2$ and $\mathcal{R}(2, \boldsymbol{\xi}) = 5$. However, when $y = 3$, the even spread recirculation rule enables one more rental over the rental horizon than the static priority rule. Under the static priority rule, rental unit 1 is lost after serving a demand in period 3, while it is lost after serving a demand in period 5 under the even spread rule. This allows one extra demand to be served in period 5 for the even spread rule because it has one more rental unit available than the static priority rule. Figure 2.3a shows that the even spread rule also serves one more demand than the static priority rule when $y = 4$ and that both policies serve all ten units of demand when $y \geq 5$. Figure 2.3a also demonstrates that the number of rentals is not necessarily concave in y; i.e., the addition of rental unit 1 with lifetime $\mathbf{l}_1 = 2$ satisfies fewer additional units of demand than the addition of rental unit 2 with lifetime $\mathbf{l}_2 = 4$.*

In Figure 2.3b, we use the demand values from Example 2.2 but instead let the lifetime of each rental unit be a discrete uniform random variable between 2 and 4 (i.e., $\ell_2 = \ell_3 = \ell_4 = 1/3$), and estimate the expected number of rentals with a simulation executed for a sufficiently large number of replications so that the standard error of the experiment is negligible. Even though concavity is violated on individual sample paths, the expected number of rentals is revealed to be a concave function of the number of rental units. The even spread and static priority policies result in the same number of rentals regardless of the sample path for $y \leq 2$ and $y \geq 5$. However, the expected number of rentals for the even spread policy exceeds that of the static priority policy by 0.33 when $y = 3$ and by 0.26 when $y = 4$. When $y = 3$, the even spread policy results in at least one more rental than the static priority policy on 44.1% of the sample paths, and the static priority policy exceeds the even spread policy on 10.9% of all sample paths. As an example of a sample path of rental unit lifetimes in which the static priority policy outperforms the even spread policy, the static priority policy produces one more rental than the even spread policy when $y = 3$



(a) Example sample path. (b) Expected number of rentals.

Figure 2.3: Effect of the number of rental units and rental unit recirculation rule for Example 2.2.

and $\{\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3\} = \{4, 3, 2\}$.

The Static Priority Recirculation Rule

The static priority recirculation rule, denoted by the superscript SP, guides the selection of rental units to satisfy demands according to a constant priority list. That is, when a rental unit is needed to satisfy demand, the one with the highest priority among the set of available rental units is chosen. Therefore, when the rental duration is deterministic, the static priority recirculation rule is the same as the policy selecting the rental unit that has been rented the most.

Proposition 2.2. *In a rental system with general rental unit lifetime distributions, the expected profit* $\pi^{SP}(y)$ *is concave in the initial inventory of* y *rental units for the static priority recirculation rule.*

To prove this property of the expected profit function, we utilize first forward differences of the state equations and sample path coupling to compare the expected value of an additional rental unit for a system with initial inventory level y to the expected value of a an additional rental unit for a system with initial inventory level $y + 1$. By assigning the lowest priority to the additional rental unit, we are able to analyze the system without changing any existing allocations of rental units to customers.

The Even Spread Recirculation Rule

We now consider the even spread recirculation policy, which satisfies a demand with the rental unit that has been rented the fewest number of times. Defining $R^{\gamma}_{(n,m)}(y, \boldsymbol{\xi})$ as the number of times that rental unit m is rented in period n under some policy γ, an even spread compliant policy selects a rental unit m to satisfy a demand based on available rental units that minimize $\sum_{t=1}^{n-1} R^{\gamma}_{(t,m)}(y, \boldsymbol{\xi})$. When the loss probability for rental units is non-decreasing in the number of times rented, the even spread recirculation priority corresponds to a hazard rate ordering. We assume that ties are broken by some static priority list for allocating rental units. We first investigate the concavity of the expected profit function in the initial rental inventory (Proposition 2.3). We then demonstrate the optimality of the even spread policy to maximize the expected profit when the loss probability of each rental unit increases with the number of times that the unit has been rented (Proposition 2.4).

Proposition 2.3. *In a rental system with general rental unit lifetime distributions, the expected profit $\pi^{ES}(y)$ is concave in the initial inventory of y rental units under the even spread recirculation rule.*

The proof of Proposition 2.3 is challenging as the additional rental units may change which demand is served by the existing rental units. When comparing the effect of an additional unit on systems with y and $y+1$ rental units, we restrict any rental units from serving certain customer demands in the system with y units so that corresponding rental units serve the same customer demands in the two systems. We then show that relaxing the restriction so that the additional unit for the system with y units has a more positive effect on the expected profit than an additional unit for a system with $y+1$ units.

When the rental unit loss probability is increasing in the number of times that the unit has been rented, we identify the even spread policy as the optimal rental unit recirculation rule to maximize the expected profit.

Proposition 2.4. *If the loss probability of each rental unit increases with the number of times that the unit has been rented, then the profit from the even spread recirculation rule is stochastically larger than that of any other count-based recirculation rule; i.e., $\Pi^{ES}(y, \boldsymbol{\xi}) \geq_{st} \Pi^{\gamma}(y, \boldsymbol{\xi})$, $\gamma \in \mathcal{C}$.*

Our key argument in this proof is a pairwise interchange argument in which iteratively switching each instance that an allocation violates the even spread policy to conform to the even spread policy increases the expected number of rentals. We require additional notation to compare sample paths in our argument, which we describe along with an overview of the steps of the proof:

1. Find the first allocation decision over the rental horizon that violates the even spread policy. We denote this existing policy with the superscript V for "violating." Assume that this violating decision occurs in some period n. Specifically, a rental unit j is allocated to demand when some other rental unit i is available and $\sum_{t=1}^{n-1} R^V_{(t,j)}(y, \boldsymbol{\xi}) > \sum_{t=1}^{n-1} R^V_{(t,i)}(y, \boldsymbol{\xi})$. The availability of rental units i and j implies that $\sum_{t=1}^{n-1} R^V_{(t,j)}(y, \boldsymbol{\xi}) < \mathbf{l}_j$ and $\sum_{t=1}^{n-1} R^V_{(t,i)}(y, \boldsymbol{\xi}) < \mathbf{l}_i$.

2. Consider a switched forward allocation path of units i and j in periods $n, n+1, \ldots, N$ so that rental unit i is allocated instead of rental unit j. We refer to this allocation with the superscript S for "switched."

3. Change values in $\boldsymbol{\xi}$ related to the lifetimes and rental durations after period n for units i and j with two new partial sample path vectors $\xi_{(1)}$ and $\xi_{(2)}$. Specifically, we generate two sets of random durations $(a_{(1),1}, a_{(1),2}, \ldots)$ and $(a_{(2),1}, a_{(2),2}, \ldots)$ for demands after period n served by two different rental units and inverse probability mass function values $\eta_{(1)}$ and $\eta_{(2)}$ for the conditional lifetime distributions of the two

rental units. The latter information allows the determination of the lifetimes $\mathbf{l}_{(1)}$ and $\mathbf{l}_{(2)}$.

4. Calculate the number of rentals over the entire horizon under four scenarios (with corresponding notation for the total number of rentals used for convenience): (1) $\mathcal{R}^V(\xi_{(1)}, \xi_{(2)})$ for the violating allocation with $\xi_{(1)}$ applied to rental unit i and $\xi_{(2)}$ to rental unit j; (2) $\mathcal{R}^V(\xi_{(2)}, \xi_{(1)})$ for the violating allocation with $\xi_{(2)}$ applied to unit i and $\xi_{(1)}$ to unit j; (3) $\mathcal{R}^S(\xi_{(1)}, \xi_{(2)})$ for the switched allocation with $\xi_{(1)}$ applied to unit i and $\xi_{(2)}$ to unit j; and (4) $\mathcal{R}^S(\xi_{(2)}, \xi_{(1)})$ for the switched allocation with $\xi_{(2)}$ applied to unit i and $\xi_{(1)}$ to unit j.

5. Compare scenarios to observe that $\mathbb{E}\left[\mathcal{R}^S(y)\right] \geq \mathbb{E}\left[\mathcal{R}^V(y)\right]$, which implies that $\mathbb{E}\left[\Pi^S(y)\right] \geq \mathbb{E}\left[\Pi^V(y)\right]$ under certain assumptions on the cost parameters.

6. Go to Step 1 and repeat until the switched allocation is equivalent to the even spread allocation.

2.5.2 Condition-Based Rental Unit State

We now study a different model of rental units in which each rental unit m has a known state $s_m \in \{1, 2, \ldots, S\}$ that may change after each time that the unit is rented. On a sample path $\boldsymbol{\xi}$, we define $s_{m,i}$ as the state of rental unit m after it is rented for the ith time, $m \in \{1, 2, \ldots, y\}$ and $i \in \{1, 2, \ldots, \mathbf{l}_m\}$. The initial state of each rental unit is defined as $s_{m,0} = 1$ and a rental unit's retirement from recirculation corresponds to $s_{m,\mathbf{l}_m} = S$. A transition probability matrix P_a governs the evolution of each rental unit's state upon each instance in which the unit is rented with duration a. We define $P_a(i,j)$ as the probability that a rental unit transitions from state i to state j after each rental with $i, j \in \{1, 2, \ldots, S\}$. We also assume that $\sum_{a=A_{min}}^{A_{max}} r_a h_a + c \geq (k^L - k^S) \sum_{a=A_{min}}^{A_{max}} P_a(i, S)$ for $i = 1, 2, \ldots, S-1$ so that the expected value of offering a rental is never negative. For convenience, we define an overall transition probability matrix P with $P(i,j) := \sum_{a=A_{min}}^{A_{max}} h_a P_a(i,j)$.

One simple recirculation policy based on the observed rental unit state is to allocate the rental units in increasing order of their state. In other words, the rental unit that is in the best condition is given the highest allocation priority. We label this policy as the "best-first" policy, denoted by the superscript BF. Similarly, the "worst-first" policy, which we denote as WF, gives the highest priority to the rental unit in the worst condition for which it can still be rented out. Specifically, for each policy, the rental unit selected from the set of available rental units $\mathcal{A}$ obeys

$$m^{BF} \in \arg\min_{m \in \mathcal{A}} s_m,$$
$$m^{WF} \in \arg\max_{m \in \mathcal{A}} s_m.$$

For both policies, we show that the expected number of rentals is concave in the initial inventory level. We use $\mathcal{D}$ to represent the set of all recirculation rules for choosing a rental unit to allocate solely based on the condition and availability of each rental unit.

Proposition 2.5. *For the best-first and worst-first recirculation rules, the expected profit $\pi^\gamma(y)$, $\gamma \in \{BF, WF\}$, is concave in the initial inventory of y rental units.*

We next consider the optimal rental unit recirculation policy when rental unit selection decisions are based on the rental unit condition. We assume that the transition matrix is totally positive of order 2; i.e., that $P(i,j)P(i',j') \geq P(i,j')P(i',j)$ for all $i < i'$, $j < j'$. Brown and Chaganty (1983) show that this property implies that the first passage time from state 1 to some state $C_j = \{i : i > j\}$ has an increasing failure rate for $j = 1, \ldots, S-1$.

Proposition 2.6. *If the transition matrix P is totally positive of order 2, then the profit from the best-first policy is stochastically larger than that of all other condition-based recirculation rules; i.e., $\Pi^{BF}(y, \boldsymbol{\xi}) \geq_{st} \Pi^{\gamma}(y, \boldsymbol{\xi})$, $\gamma \in \mathcal{D}$.*

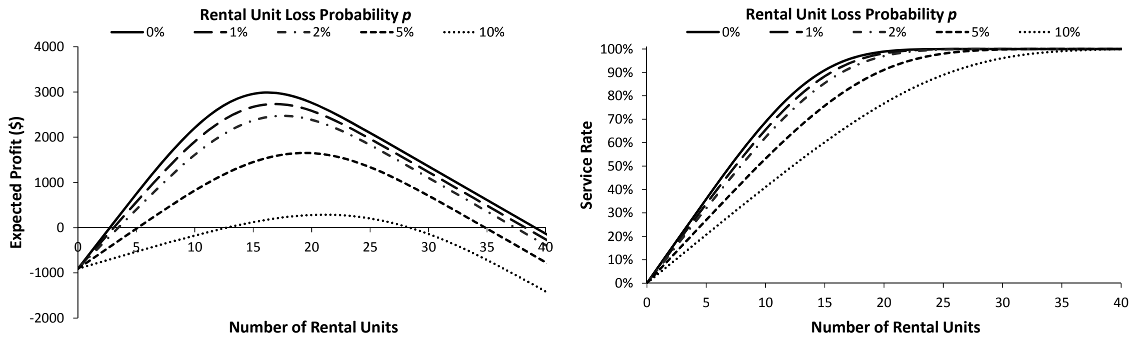
2.6 Case Study: Rent the Runway

Motivated by the high-fashion dress rental business Rent the Runway, we introduce the model parameters representing a rental system with usage-based loss of inventory in Section 2.6.1. We discuss the impact of the rental inventory loss on the optimal procurement decision in Section 2.6.2 and the effect of the rental unit recirculation rule on rental inventory management in Section 2.6.3. All numerical testing is performed via sample average approximation, as described in Kleywegt et al. (2002).

2.6.1 Rental Model Parameters

The product we consider is a "middle-tier" dress as described in Eisenmann and Winig (2012); i.e., a full-price rental provides a net revenue of \$59, which is the difference between \$90 in revenue and \$31 in costs of cleaning, shipping, packaging and credit card processing. However, customers are allowed to rent a second style for \$25 and a second size for free; thus, a unit may not achieve \$59 in net revenue every time it is rented. We assume that these three scenarios for a rental — renting as the primary dress with net revenue of \$59, renting as the secondary dress with net revenue of \$20, and renting as the free second size with net cost of \$5 — occur with probabilities 50%, 20%, and 30%, resulting in an expected net revenue of $r = \$32$ per rental.

Eisenmann and Winig (2012) report that Rent the Runway purchases a middle-tier dress with a retail price of \$750 for \$226. We assume an annual unit holding cost that is equal to 20% of the purchase price of the dress to account for the cost of storage and the cost of capital. At the end of a fashion season, dresses in a variety of conditions are sold in New York City at what is known as a "sample sale." Based on websites such as Yannetta (2013) that report on these sales, we let a dress in good condition sell for $80\% - 85\%$ off of the \$750 retail price and a dress in bad condition (i.e., a dress that retires from the rental inventory) to sell for 95% off of the retail price. Adjusting these sample sale prices for staging and transaction costs, we assume a dress that does not retire from the rental inventory by the end of the season to have a salvage value of \$100 and a dress that retires from the rental inventory to have a salvage value of \$30. We also calculate procurement costs separately for these two types of dresses by combining their purchase prices, holding costs and salvage values. For a 26-week horizon, the cost of procuring a dress is $k^S := \$149$, which consists of a purchase price of \$226, a holding cost of \$23 and a salvage value of \$100. A dress retiring from the rental inventory incurs an additional penalty of \$70, resulting in a procurement



(a) Expected profit. (b) Service rate.

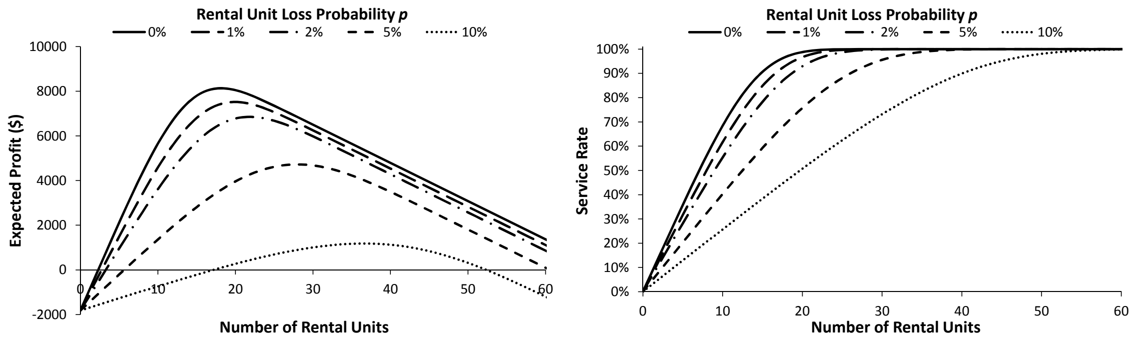
Figure 2.4: The optimal inventory level increases with the loss probability for a 26-week system, and ignoring inventory loss significantly reduces expected profit compared to the optimal inventory level.

cost of $k^L := \$219$. Finally, we choose $c := \$5$ as a customer goodwill penalty for the loss of a sale. With these parameters, a dress must be rented five times (seven times, on average, if the possibility of loss exists) to break even based on the ratio k^S/c (k^L/c).

With each period corresponding to a week, we consider Poisson distributed demand with a mean of $\lambda = 7$ per week and a rental horizon of $N = 26$ weeks, which corresponds to one of two major fashion seasons each year. We will also consider a longer rental horizon of $N = 52$ for a dress that could be in style for two consecutive seasons. We model each rental duration as lasting for a constant of $A = 2$ periods; i.e., the rented dress will be unavailable during the weekend for which it is rented and the weekend either preceding or following that weekend, depending on the day of the week on which the rental begins. A more granular representation of the rental duration in terms of the individual days is certainly possible. However, we believe that weekly periods adequately represent the system under the assumption that customers of Rent the Runway rent dresses primarily for weekend events.

2.6.2 Rental Inventory Loss

We first investigate the importance of accounting for the possibility of usage-based loss when choosing the initial inventory of rental units. Allowing the lifetime of each rental unit to follow a geometric distribution with a loss probability of $p \in \{0, 0.01, 0.02, 0.05, 0.10\}$, we illustrate the expected profit as a function of the initial inventory of rental units for the short rental horizon of $N := 26$ periods in Figure 2.4, and compare it to the "state of the art" in rental inventory management represented by Baron et al. (2011), which does not include inventory loss. Consistent with Proposition 2.1, we observe the expected profit function to be concave in the number of rental units to procure in the beginning of the rental season. In the system with no inventory loss ($p = 0$), we identify the optimal solution as 16 units with a corresponding service rate — i.e., the percentage of customers that are served — of 93.5%. However, when there is the possibility of inventory loss (i.e., $p > 0$), we find the optimal number of rental units to increase in the rental unit loss probability.



(a) Expected profit.

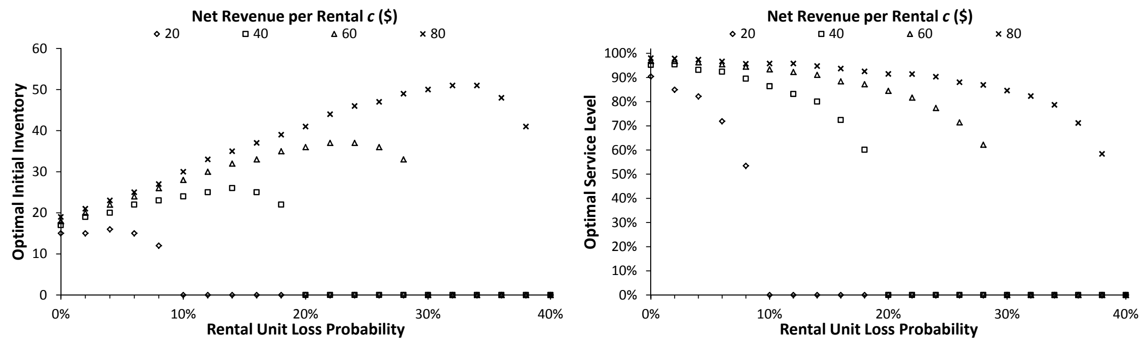
(b) Service rate.

Figure 2.5: Accounting for inventory loss is more important in terms of effect on expected profit for a 52-week system.

Specifically, for a 5% loss probability, the optimal policy is to add three rental units to the initial inventory. Hence, ignoring inventory loss and using 16 rental units instead of the optimal 19 rental units results in a reduction of 7.3% in the expected profit. Furthermore, the service rate would only be 79.4% instead of the 88.7% corresponding to the optimal number of rental units for the system with $p = 5\%$.

Figure 2.5 shows that the impact of ignoring inventory loss is more dramatic for the longer rental horizon covering 52 weeks than for the shorter rental horizon with 26 weeks. This can be explained by the availability of fewer rental units to rent towards the end of the longer rental horizon. The comparison of Figure 2.5 to Figure 2.4 reveals more asymmetry in the expected profit as a function of the initial inventory of rental units for the longer horizon. More specifically, the slope of the expected profit function for a lower value of the number of rental units is steeper because each rental unit averts more lost sales in a long horizon than in a short horizon. Furthermore, the higher optimal service rate for the system with the longer horizon than the system with the shorter horizon reflects the higher value of a marginal rental unit. In other words, the consequence of having too few rental units is more severe in the longer horizon.

For rental systems considered in Figure 2.4 and Figure 2.5, the optimal policy is to always add more rental inventory to account for the loss of rental units; i.e., the profit-maximizing inventory level is increasing in p for $p \in \{0, 0.01, 0.02, 0.05, 0.10\}$. However, if the loss probability is sufficiently high, then the units will not be rented enough to justify having any stock at all, which means that the optimal policy is to not stock any rental units. Figure 2.6 illustrates such a policy by considering $c \in \{20, 40, 60, 80\}$ for the net revenue per rental to represent varying levels of profitability per rental, $N = 26$ weeks for a rental horizon, and $\lambda = 7$ for the mean demand. We observe that the optimal response to an increasing inventory loss probability is to initially increase the inventory of rental units until we reach a certain value of the loss probability p associated with the optimal number of rental units $y^\star$ to procure in the beginning of the rental horizon. As p continues to increase, the optimal number of rental units decreases and the optimal service level also appears to be non-increasing in the loss probability. Eventually, a loss probability $\hat{p}(c)$ is reached such



(a) Optimal initial inventory. (b) Optimal service rate.

Figure 2.6: The optimal inventory level is non-monotonic in the rental unit loss probability p.

Table 2.2: Performance of simple policies based on upper bound for number of lost rental units.

p	0%	1%	2%	5%	10%
y^*	16	17	18	19	21
y^{UB1}	16	18	20	26	37
π^{UB1}/π^*	100.0%	98.2%	95.0%	69.2%	-20.8%
y^{UB2}	16	18	20	25	34
π^{UB2}/π^*	100.0%	98.2%	95.0%	75.0%	8.1%

that $y^\star = 0$ for all $p \geq \hat{p}(c)$. Naturally, the optimal service level and $\hat{p}(c)$ increase with c because a dress with a higher net revenue per rental requires fewer rentals to be profitable.

We also provide two simple bounds on the the number of lost rental units when the loss probability is p. Both of these bounds also translate to easy-to-implement policies that managers may use to compensate for usage-based loss by adding extra rental units. In particular, a manager who has chosen the optimal initial inventory level without usage-based loss — possibly through Baron et al. (2011) or by approximating the system as an M/G/c/c queue — can simply increase the initial inventory level by the approximate number of lost rental units.

First, we present a bound (UB1) based on the assumption that all demand is served; i.e., $\mathbb{E}\left[\mathcal{R}(y,\boldsymbol{\xi})\right] \leq \mathbb{E}\left[\sum_{n=1}^{N} D_n\right]$, where $D_1, D_2, \ldots, D_N$ are random variables representing the demand over the horizon. The expected number of lost rental units is then $\mathbb{E}\left[\mathcal{Z}(y,\boldsymbol{\xi})\right] = p\mathbb{E}\left[\mathcal{R}(y,\boldsymbol{\xi})\right] \leq p\mathbb{E}\left[\sum_{n=1}^{N} D_n\right]$.

A second bound (UB2) is derived from assuming that all rental units have 100% utilization resulting from demand that is sufficiently large in each period. Denoting the expected rental duration by $\bar{a}$, the expected maximum number of customers that a rental unit can serve is $\lceil N/\bar{a} \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. Thus, $\mathbb{E}\left[\mathcal{R}(y,\boldsymbol{\xi})\right] \leq y\lceil N/\bar{a} \rceil$ and

$\mathbb{E}\left[\mathcal{Z}(y,\boldsymbol{\xi})\right] = p\mathbb{E}\left[\mathcal{R}(y,\boldsymbol{\xi})\right] \leq py\lceil N/\bar{a}\rceil$.

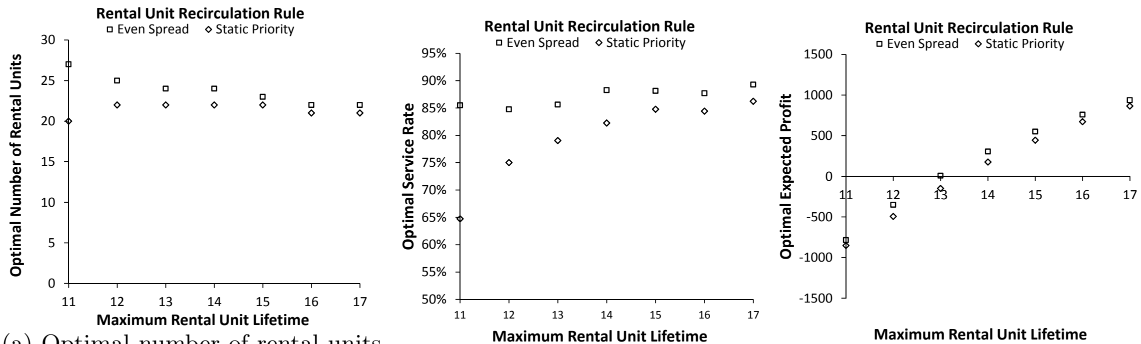
We can also compare the optimal initial inventory level to a simple policy using the optimal initial inventory level $y = 16$ when $p = 0$ combined with the two upper bounds on the number of lost rental units. If all demand is assumed to be satisfied (UB1), the expected number of rentals is 162 and the expected number of lost rental units is $162p$. If all rental units are fully utilized (UB2), each rental unit will be rented out exactly 13 times, which results in a maximum number of rentals of 208 and a maximum expected number of lost rental units of $208p$. Table 2.2 shows the policies and their performance in relation to the optimal policy. As expected given that the policies are upper bounds on the number of lost rental units, the initial inventory level for both bounds exceeds that of the optimal policy. For example, when $p = 5\%$, the initial inventory should be increased from 16 to 19 for the optimal policy due to inventory loss. However, under $UB1$, the initial inventory is increased to 26, achieving only 69.2% of the optimal profit. Under $UB2$, the initial inventory is increased to 25 and achieves 75% of the optimal profit.

2.6.3 Rental Unit Recirculation Rules

Different recirculation rules employed during the rental horizon may result in different numbers of units available near the horizon's end. We expect that the importance of the rental unit recirculation policy varies according to factors such as the horizon length, rental unit lifetime distribution, and demand characteristics. Of concern to us is a horizon that is short enough that some rental units are still functional by the end of the last time period but long enough that some rental units have already retired from the rental inventory during the season. In this section, we compare the even spread and static priority policy for the count-based model, omitting similar managerial insights and results for the best-first and worst-first rules of the condition-based model.

Executives at Rent the Runway indicate that the policy used in practice more closely resembles the static priority policy than the even spread policy. Out of convenience, dresses that have just returned from cleaning after a rental may be selected to satisfy the next rental. However, because individual units are not tracked, there may be an element of randomness in dress selection as workers select a dress to rent out. The goal of this section is to quantify the effect of using the even spread policy for rental unit recirculation over the static priority recirculation rule. For an adequate representation of the role of the rate at which the loss probability is increasing, we consider the lifetime of a rental unit to be a discrete uniform random variable that takes values between 1 and $A_{max} \in \{10, 11, \ldots, 20\}$ rentals. As before, we consider a rental horizon of 26 periods and a mean demand of 7 units, with all other parameters remaining the same.

The rental system illustrated in Figure 2.7 is only profitable when $A_{max} \geq 13$ for the even spread policy and when $A_{max} \geq 14$ for the static priority policy due to the costs incurred when rental units are lost. Consistent with Proposition 2.4, the even spread policy achieves a higher expected profit than the static priority policies. This performance difference can be explained by the nature of the even spread policy to delay the failure of rental units until later periods; thus, the even spread policy satisfies higher demand in later periods compared to the static priority recirculation policy. On the other hand, the static priority policy causes failures to occur earlier in the rental horizon, limiting the system's ability to



(a) Optimal number of rental units $y^{\star}$.

(b) Optimal service rate.

(c) Optimal expected profit.

Figure 2.7: Choosing the even spread policy over the static priority policy increases the optimal initial inventory level, service rate, and expected profit.

meet higher quantities of demand in later periods.

The optimal number of rental units for the even spread, as well as the corresponding service rate, exceeds that of the static priority policy. Choosing the even spread policy instead of the static priority policy allows for rental units to be profitably added, thereby increasing the service rate. For example, the optimal initial inventory level is two units higher for even spread policy than the static priority policy when $A_{max} = 14$, and the service rate is 6.0 percentage points higher for the even spread policy.

2.7 Conclusions

As rental industries continue to grow in size and the scope of products rented, inventory management techniques that account for the complexities of rental systems become critical for achieving profitability and service goals. We develop a discrete-time rental model with random usage-based loss of inventory that also includes arbitrarily distributed customer demands and random rental durations, and identify structural properties for this model. The concavity of the expected profit function in the initial inventory of rental units is shown to hold for geometrically distributed rental unit lifetimes regardless of the rental unit recirculation rule. When rental unit lifetimes are generally distributed, we also show the concavity of the expected profit function in the initial inventory of rental units for simple rental unit recirculation rules which are count-based or condition-based. We further demonstrate the optimality of the even spread policy in the count-based setting and the best-first policy in the condition-based setting to maximize the expected profit when the loss probability of each rental unit increases with the number of times it is rented.

Several important insights emerge from a numerical analysis of our rental inventory management solutions for a high-fashion dress rental business. First, we find that the possibility of inventory loss during the rental season can significantly affect profitability, even with a small probability of loss each time that a unit is rented. Choosing the number of rental units to procure in the beginning of the rental season by ignoring the effect of rental inventory loss can reduce the expected profit by 7%. Second, we examine how the optimal inventory policy responds to the increasing loss probability. We show that the optimal policy is to first procure additional rental units, then decrease the number of rental units to be procured and eventually procure zero rental units. Finally, we consider rental unit lifetime distributions with loss probabilities that are increasing in the number of rentals. For horizon lengths and lifetime distributions in which the recirculation rule affects the expected profit, we show that choosing the even spread policy allows for more inventory to be profitably obtained and can increase the service level by up to 6 percentage points.

Many categories of products available to rent, including dresses at Rent the Runway, present customers with the option to substitute if their first choice is unavailable. Thus, a future research direction in the study of rental inventory management is the case of multiple products with stock-out based substitution. Potential future work also includes advance reservation acceptance policies, in-season reordering and maintenance decisions, and rental pricing.

Chapter 3

Reservation Admission Control in Rental Systems

With Alan Scheller-Wolf and Sridhar Tayur

3.1 Introduction

When businesses provide customers with the temporary use of a reusable asset, they may offer customers the option to reserve the asset in advance to better match supply and demand. Reservations are particularly important when customer demand is time sensitive or customers are highly averse to waiting. In many rental businesses, either the customer or the firm requires a commitment to provide service in advance of the service start time. Examples include rental businesses that allow specific rental units such as high-fashion dresses to be reserved, restaurants that allow reservations for tables, and hospitals in which doctors or machines times may be booked through appointments. Decisions about the number of reservations to offer and policies surrounding changes and cancellations affect the percentage of overall demand captured, the utilization levels of the reusable assets, and the overall service level.

At Rent the Runway, customers may place reservations for a dress up to four months in advance. For each dress size of each style, customers can see a calendar of dates for when they can make reservations. Customers may cancel reservations with a full refund up to thirty days in advance of the delivery date. Between fourteen and thirty days, a canceling customer receives credit to her account. A customer may cancel within fourteen days of the event if she pays a fee of $9.95. In the event that Rent the Runway is not able satisfy a reservation due to damage or another customer's late return, customers report that a company stylist tries to work directly with them to find a substitute item.

In this chapter, we study the optimal reservation admission decision and discuss challenges related to showing the optimal policy. To the best of our knowledge, we are the first to provide results for a stochastic model in which acceptance decisions are made over time upon the arrival of reservation requests. We also present a heuristic for this decision and numerically evaluate its performance. After reviewing related literature in Section 3.2, we describe our model in Section 3.3. We discuss challenges related to structural results in

Section 3.4 and heuristic policies in Section 3.5. In Section 3.6, we provide numerical results related to three different rental operations, and present improvements to the heuristic in Section 3.7. We conclude in Section 3.8 with summary of our work and a discussion of possible future topics of investigation.

3.2 Related Literature

Early work on reservations primarily focused on characterizing the blocking probability of a loss system with reservations or the delay distribution when customers accept the next available service time, and was motivated by bandwidth management in telecommunications applications. Emstad and Feng (1990) characterize the blocking probability of a discrete-time model with stationary demand, stationary notice distribution, and service times of one period for both loss systems and systems in which customers will wait. Kaheel et al. (2006) characterizes the blocking probability for a similar loss model but allows for service times to follow a geometric distribution, and Zhu and Veeraraghavan (2008) studies a related model in which customers have multiple acceptable starting times. The blocking probabilities of continuous-time reservations systems are studied for a single resource by Coffman Jr et al. (1999) and for multiple resources by Lu and Radovanović (2007) and van de Vrugt et al. (2014).

The tactical decision of admission control has also attracted attention, with most work motivated by bandwidth sharing networks. Luss (1977) and Virtamo and Aalto (1991) analyze models for accepting and rejecting customer reservations for a discrete-time model. In their models, all customers arrive before the start of a finite service horizon and request a reservation for a start time that is uniformly distributed over the horizon. Greenberg et al. (1999) allow for reservations to be placed far in advance in a system with random holding times, and propose an algorithm for admissions control of requests in which service is to begin immediately. Levi and Shi (2011) study a problem in which arrival rate, notice time, and revenue generated vary by customer class, and propose a policy for class-based total admission rates based on a linear programming approximation of the system.

The admissions control problem has additionally been studied for restaurant and health-care applications. Bertsimas and Shioda (2003) address the problem of allocating restaurant tables to parties and which party sizes to deny at different time intervals. Two papers analyze appointment scheduling for a hospital resource to balance low-priority demand that arrives far in advance and high-priority demand that has a shorter notice time. Gerchak et al. (1996) model a daily decision to decide how many waiting elective patients to admit to an emergency room, and Patrick et al. (2008) study a similar decision but with a more explicit schedule for a diagnostic resource.

Another tactical decision that bears similarity to the admission control decision is the lead-time quotation problem. Duenyas and Hopp (1995) model a production system as a G/G/1 queue with an increasing failure rate for the service time distribution. Modeling the state space based on the number of jobs waiting and the time-in-service of the active job, they show the optimality of a threshold policy for the acceptance decision based on the time-in-service of the active job. Savasaneril et al. (2010) modify this model to include inventory ordered using a base-stock policy and an M/M/1 production system. They show that the base stock policy has a threshold structure. Kapuscinski and Tayur (2007) consider a discrete-time model in which demand for two classes of customers arrives each period. Classes are distinguished by their margin and penalty costs, and the key decision is the lead times to quote to arriving demands. Their structural results rely on showing that the state space needed to describe the optimal policy can be reduced to variables representing the total "workload" and the latest due date.

The strategic question of whether to accept reservations in the context of restaurants was the subject of Alexandrov and Lariviere (2012). In their model, the restaurant has the same service capacity regardless of whether reservations are offered, but the number of customers served by a restaurant offering reservation is diminished by no-show customers.

Closest to our work is Papier and Thonemann (2010), who model the admission decision for a rental system with Poisson arrivals and exponential service times in which there are two classes of customers: higher value customers who make advance reservations with a constant notice time and lower value customers who request immediate service. All advance reservations must be accepted, but customers requesting immediate service may be turned away. They propose a heuristic based on approximating the probability of accepting demands in future decisions. The performance of this policy is compared to a policy that ignores reservations as a lower bound and to a policy with known demand arrivals by customer class. They position reservations for a rental system as a form of advance demand information to build on a line of research started by Hariharan and Zipkin (1995) that shows the value of advance demand information in conventional inventory settings.

3.3 Model Description

We study admission control policies for a single-class model in which each customer requests a reservation that must be immediately accepted or rejected. A rental firm begins a finite rental season over the horizon $[0, T]$ with y rental units. All customers request service a constant τ time units before they desire service to begin, and we refer to this time τ as the notice time. Demand is Poisson distributed with rate λ, and rental times (i.e., the duration of each rental) are exponentially distributed with rate μ. Without loss of generality, we let $\mu = 1$ and scale λ accordingly. We let $F(x) := 1 - e^{\mu x}$ represent the cumulative distribution function for the service time (with complement $\bar{F}(x) := e^{\mu x}$). Revenue per unit time is r, with the expected revenue of r/μ earned from satisfying a demand. Rejecting a demand carries with it a penalty of c_R. Not being able to fulfill an accepted reservation has a penalty of c_F. We assume that $c_R < c_F$; otherwise, the optimal policy is to accept all reservation requests.

Due to the complexities of this model's arrival process, we consider a finite horizon over which J reservation requests arrive, with each arrival epoch j referred to as period j for $j = 1, 2, \ldots, J$. Each request is associated with an acceptance decision $a_j(k, R) \in \{0, 1\}$ with $a_j(k, R) = 0$ representing a rejection and $a_j(k, R) = 1$ representing an acceptance. We let $k \in \{0, 1, \ldots, y\}$ represent the number of rental units that are in service (i.e., "busy") at the time of the decision. This decision takes place upon each customer j's arrival at time A_j. If the reservation request is accepted, the scheduled start time of the jth arrival is $S_j := A_j + \tau$. We denote the set of reservations that have been accepted but have not yet begun service upon the jth arrival as R. The vector $R := \{t_1, t_2, \ldots\}$ contains the time until future reservations are scheduled to start service; i.e., the next demand to begin service starts in t_1 time units. Accepting the jth reservation request at time A_j transforms R_j to $R_j^+ := R_j \cup \tau$.

We define two functions for the evolution of the reservation vector R over the time horizon. First, we define the function $U_j(R)$ to update the reservation vector from arrival epoch j to the next arrival epoch $j+1$ as

$$U_j(R) := \{t \in R : t > A_{j+1} - A_j\} - (A_{j+1} - A_j),$$

where $A_{J+1} = \infty$. We also note that $U_j(R) = \emptyset$ if $A_{j+1} - A_j > \tau$. Second, we use the function $\alpha_j(R)$ to choose a subset of reservations for which service begins before time A_{j+1} (i.e., those removed by the function $U_j(R)$), and define it as

$$\alpha_j(R) := \{t \in R : t < A_{j+1} - A_j\}.$$

This notation allows us to define $|\alpha_j(R \cup \tau)|$ stages between period j and $j+1$. The index i, $i = 1, 2, \ldots, |\alpha_j(R \cup \tau)|$, is used to refer to a specific stage. Each stage i, $i \leq \alpha_j(R)$, corresponds to either the start of service for an accepted reservation if a rental unit is available or a failure to fulfill the reservation if no rental unit is available. If $A_j + \tau < A_{j+1}$, stage $|R| + 1$ corresponds to the start time for the reservation under consideration for acceptance, and $t_{|R|+1} = A_j + \tau$.

Next, we define state probability variables to keep track of the number of busy units between periods j and $j+1$. For convenience, we denote the probability that a binomial random variable equals m with n trials and success probability p by the function $b_m(n,p)$. When $A_j + \tau < A_{j+1}$, we also use the abbreviated notation $b_m(n) := b_m\left(n, \bar{F}(A_{j+1} - A_j - \tau)\right)$ to represent the probability that m units are still in service at time A_{j+1} if n units are in service at time $A_j + \tau$. The probability that m rental units are busy immediately before the possible start of service at stage i after arrival epoch j, at which point k rental units are busy and the reservation vector is R, is defined as $p^j_{i,m}(k,R)$ and is calculated recursively with

$$p^j_{1,m}(k,R) := b_m(k, \bar{F}(t_1 - A_j))$$

and

$$p^j_{i,m}(k,R) := \sum_{n=m-1}^{y} b_m\left(\min\{y, n+1\}, \bar{F}(t_i - t_{i-1})\right) p^j_{i-1,n}(k,R), \qquad i = 2, 3, \ldots, |\alpha_j(R \cup \tau)|.$$

When $A_j + \tau < A_{j+1}$, the state probability $p^j_{|R|+1,m}(k,R)$ plays an important role in our analysis as it corresponds to the last time when the system state remains unaffected by $a_j(k,R)$. For convenience, we use the notation $\beta^j_m(k,R) := p^j_{|R|+1,m}(k,R)$, which may abbreviated to $\beta_m(k) := p^j_{|R|+1,m}(k,R)$ when the arrival epoch j is clear. To account for $a_j(k,R) = 1$ (i.e., when the admission decision is "yes"), we also define the state probability at time A_{j+1} as

$$\begin{aligned} p^j_{Y,m}(k,R) :&= \sum_{n=m-1}^{y} b_m\left(\min\{y, n+1\}\right) p^j_{|R|+1,n}(k,R) \\ &= \sum_{n=m-1}^{y} b_m\left(\min\{y, n+1\}\right) \beta_n(k). \end{aligned}$$

If $a_j = 0$ or $A_j + \tau \geq A_{j+1}$ (i.e., when the admission decision is "no" or not relevant to the system's evolution by time A_{j+1}), then the state probability at time A_{j+1} is instead

$$p^j_{N,m}(k,R) := \sum_{n=m}^{y} b_m\left(n, \bar{F}(A_{j+1} - t_{|\alpha_j(R)|})\right) p^j_{|\alpha_j(R)|,n}(k,R).$$

For convenience, we may refer to the state variables as $p_{i,m}(k)$ or $p_{i,m}$ when the meaning is clear as the state probability between some time A_j and A_{j+1} with reservation vector R at period j.

We can now define the maximum expected profit from decision j to the end of the horizon, which we denote by $\pi_j(k,R)$ where k is the number of rental units busy and R is the vector of accepted reservations at the time of decision a_j. The maximum expected profit-to-go for acceptance and rejection decisions for the jth demand are denoted by $\pi^1_j(k,R)$ and $\pi^0_j(k,R)$, respectively. Either the expected reward r or the penalty c_R is earned upon the admission decision, but the failure-to-serve penalty c_F (as well as a negation of the reward r) is assessed if applicable upon the start of service. Thus, the expected profit for an accepted reservation request is

$$\pi^1_j(k,R) := \begin{cases} r - (r + c_F)\sum_{i=1}^{|R|+1} p_{i,y} + \sum_{m=0}^{y} p_{Y,m}\pi_{j+1}(m,\emptyset) & \text{if } A_j + \tau < A_{j+1}, \\ r - (r + c_F)\sum_{i=1}^{|\alpha_j(R)|} p_{i,y} + \sum_{m=0}^{y} p_{N,m}\pi_{j+1}(m, U_j(R \cup \tau)) & \text{otherwise,} \end{cases}$$

and the expected profit for a rejected reservation request is

$$\pi^0_j(k,R) := -\, c_R - (r + c_F)\sum_{i=1}^{|\alpha_j(R)|} p_{i,y} + \sum_{m=0}^{y} p_{N,m}\pi_{j+1}(m, U_j(R)),$$

where $\pi_{J+1}(\cdot,\cdot) = 0$. We define the optimal acceptance policy $a^*_j(k,R) = 1$ if $\pi^1_j(k,R) \geq \pi^0_j(k,R)$ and $a^*_j(k,R) = 0$ otherwise.

We also define three comparison operators for use in our analysis. First, we use $\Lambda p_m(k) := p^j_{Y,m}(k,R) - p^j_{N,m}(k,R)$ to denote the change in the probability that the system has m busy units at time A_{j+1} if the jth reservation request is accepted rather than rejected when there are k busy units at time A_j. If $A_j + \tau < A_{j+1}$, then $\Lambda p_m(k) = 0$ because the accepted reservation has no effect on the system's state before time A_{j+1}. If $A_j + \tau \geq A_{j+1}$,

$$\begin{aligned} \Lambda p_m(k) :&= p^j_{Y,m}(k,R) - p^j_{N,m}(k,R) \\ &= \sum_{n=m-1}^{y} b_m(\min\{y, n+1\})\, p^j_{|R|+1,n}(k,R) - \sum_{n=m}^{y} b_m(n)\, p^j_{|R|+1,n}(k,R) \\ &= b_m(m) p^j_{|R|+1,m-1}(k,R) + \sum_{n=m}^{y-1} \left(b_m(n+1) - b_m(n)\right) p^j_{|R|+1,n}(k,R) \\ &= b_m(m)\beta_{m-1}(k) + \sum_{n=m}^{y-1} \left(b_m(n+1) - b_m(n)\right)\beta_n(k). \end{aligned}$$

Second, we also consider the effect of an additional busy rental unit at time A_j. If $A_j + \tau \geq A_{j+1}$, we define $\Delta p_m(k)$ as

$$\begin{aligned}\Delta p_m(k) :&= p^j_{N,m}(k+1,R) - p^j_{N,m}(k,R)\\ &= \sum_{n=m}^{y} \left(p^j_{|\alpha_j(R)|,n}(k+1,R) - p^j_{|\alpha_j(R)|,n}(k,R)\right) b_m(n).\end{aligned}$$

If $A_j + \tau < A_{j+1}$, we define $\Delta p^a_m(k)$ based on the value of $a = a_j(k,R)$,

$$\begin{aligned}\Delta p^0_m(k) :&= p^j_{N,m}(k+1,R) - p^j_{N,m}(k,R)\\ &= \sum_{n=m}^{y} \left(p^j_{|R|+1,n}(k+1,R) - p^j_{|R|+1,n}(k,R)\right) b_m(n)\\ &= \sum_{n=m}^{y} \left(\beta_n(k+1) - \beta_n(k)\right) b_m(n)\\ \Delta p^1_m(k) :&= p^j_{Y,m}(k+1,R) - p^j_{Y,m}(k,R)\\ &= \sum_{n=m-1}^{y-1} \left(p^j_{|R|+1,n}(k+1,R) - p^j_{|R|+1,n}(k,R)\right) b_m(n+1)\\ &\quad + \left(p^j_{|R|+1,y}(k+1,R) - p^j_{|R|+1,y}(k,R)\right) b_m(y)\\ &= \sum_{n=m-1}^{y-1} \left(\beta_n(k+1) - \beta_n(k)\right) b_m(n+1) + \left(\beta_y(k+1) - \beta_y(k)\right) b_m(y).\end{aligned}$$

Third, we define an operator Θ for the change in the expected profit-to-go in period j due to the acceptance of a reservation in period $j-1$ with reservation vector R in period $j-1$:

$$\Theta\pi_{j+1}(m,R) := \pi_{j+1}(m, U_j(R \cup \tau)) - \pi_{j+1}(m, U_j(R)).$$

We also use the operator Δ more generally to refer to the difference between some function $f(k+1)$ and $f(k)$; i.e., $\Delta f(k) := f(k+1) - f(k)$. For example, $\Delta\beta_m(k) := \beta_m(k+1) - \beta_m(k)$.

To complete our model definition, we describe the operation of the system on a sample path. A sample path $\boldsymbol{\xi}$ is comprised of information of information about arrival times and service durations. We define k^-_j and k^+_j as the number of busy units immediately before and after the possible service start time S_j of the jth arrival. A sample path $\boldsymbol{\xi}$ includes $\{A_1, A_2, \ldots, A_J\}$ and $\{s_{ij}\}$, which represents the remaining service duration of the ith busy rental unit at the time corresponding to the start of service of the jth arrival for $i = 1, 2, \ldots, k^+_j$. This implies that we resample the remaining service time for each busy rental unit, which is allowed due to the memorylessness of exponential service times. On a sample path, the number of busy units before and after each service start event has the following definition:

$$\begin{aligned}k^-_j &:= \sum_{i=1}^{k^+_{j-1}} \mathbf{1}\{s_{i,j-1} > S_j - S_{j-1}\},\\ k^+_j &:= \min\left\{y, a_j + k^-_j\right\},\end{aligned}$$

with $k_0^+ = 0$ and $\mathbf{1}\{\cdot\}$ as the indicator function. At any time t between the service start times of demands j and $j+1$, the number of busy rental units $k(t)$ is defined as

$$k(t) := \sum_{i=1}^{k_{j-1}^+} \mathbf{1}\{t > S_j - S_{j-1}\}, \qquad S_j < t < S_{j+1}.$$

The following statistics are used to evaluate the performance of some arbitrary policy γ on a sample path $\boldsymbol{\xi}$:

- $N_\gamma^Y := \sum_{j=1}^J \mathbf{1}\left\{a_j, k_j^- < y\right\}$, the number of accepted reservations that are successfully fulfilled to gain expected revenue r/μ;
- $N_\gamma^N := \sum_{j=1}^J (1 - a_j)$, the number of reservations that are not accepted, incurring a penalty c_R; and
- $N_\gamma^F := \sum_{j=1}^J \mathbf{1}\left\{a_j, k_j^- = y\right\}$, the number of reservations that are accepted but fail to be fulfilled, incurring a penalty c_F.

On any sample path, $N_\gamma^Y + N_\gamma^N + N_\gamma^F = J$. Taking the expectation over sample paths, we have $\mathbb{E}\left[N_\gamma^Y + N_\gamma^F \mathbb{E}\right] = \mathbb{E}_{\boldsymbol{\xi}}\left[\sum_{j=1}^J a_j(k)\right]$ and $\mathbb{E}\left[N_\gamma^N\right] = \mathbb{E}_{\boldsymbol{\xi}}\left[\sum_{j=1}^J (1 - a_j(k))\right]$. We use π^γ to refer to the expected profit over the entire horizon for some acceptance policy γ, which defines the whether to accept or reject a reservation request at any time t given the number of busy rental units k and reservation vector R. The optimal expected profit is defined as $\pi^* := \max_\gamma \pi^\gamma$.

3.4 Analytical Results

In this section, we first discuss properties related to how the system evolves over time. We then discuss the optimal reservation acceptance policy, which we conjecture takes the form of a threshold policy parameterized by the number of rental units that are busy given a reservation vector R. We describe how to compute numerical upper bounds for the expected profit on each realization of arrival epochs, and derive analytical upper and lower bounds for the expected profit based on its similarity to an Erlang loss system.

3.4.1 System Properties

In our analysis, we use a property that adding a busy rental unit at time A_j stochastically increases the system's state at every stage over the interval (A_j, A_{j+1}) given a decision a_j.

Lemma 3.1. *Given a_j, the probability that no more than m' rental units are busy immediately before each stage i, $i = 1, 2, \ldots, |\alpha_j(R \cup \tau)|$, is stochastically decreasing in k; i.e., $\sum_{m=0}^{m'} p_{i,m}^j(k, R) \geq \sum_{m=0}^{m'} p_{i,m}^j(k+1, R)$ for $m' = 0, 1, \ldots, y$ and $k = 0, 1, \ldots, y-1$.*

Proof. The proof is by induction. At stage $i = 1$, the property holds because $p_{1,m}^j(k, R)$ and $p_{1,m}^j(k+1, R)$ can be related through the properties of the binomial distribution as follows:

$$p^j_{1,m}(k+1,R) = p^j_{1,m-1}(k,R)\bar{F}(t_1 - A_j) + p^j_{1,m}(k,R)F(t_1 - A_j),$$

where $p^j_{1,m-1}(k,R) = 0$. The term $p^j_{1,m-1}(k,R)\bar{F}(t_1-A_j)$ represents the probability measure that the additional unit remains in service past time t_1 so that the state distribution with k in service at time A_j is shifted by one. The term $p^j_{1,m}(k,R)F(t_1 - A_j)$ represents the probability measure that the additional unit ends service before t_1 so that the additional unit has no effect on the state distribution. Therefore,

$$\sum_{m=0}^{m'} p^j_{i,m}(k+1,R) = \left(\sum_{m=0}^{m'} p^j_{i,m}(k,R)\right) - p^j_{1,m'}(k,R)\bar{F}(t_1 - A_j),$$

and the initial condition that the state distribution is stochastically decreasing in k at time t_1 holds as $p^j_{1,m'}(k,R)\bar{F}(t_1 - A_j) \geq 0$.

For any stage $i \geq 2$, we make a second argument over the possible states by induction within our first argument that makes an inductive argument over the stages. We begin by proving the initial condition for when $m' = y$: $p^j_{i,y}(k+1,R) \geq p^j_{i,y}(k,R)$, which is equivalent to

$$\begin{aligned} &\bar{F}^y(t_i - t_{i-1})\left(p^j_{i-1,y}(k,R) + p^j_{i-1,y-1}(k,R)\right) \\ &\leq \bar{F}^y(t_i - t_{i-1})\left(p^j_{i-1,y}(k+1,R) + p^j_{i-1,y-1}(k+1,R)\right), \end{aligned}$$

as the probability that y rental units remain in service at the next stage is $\bar{F}^y(t_i - t_{i-1})$. This inequality reduces to

$$p^j_{i-1,y}(k,R) + p^j_{i-1,y-1}(k,R) \leq p^j_{i-1,y}(k+1,R) + p^j_{i-1,y-1}(k+1,R),$$

which holds by the original inductive hypothesis. For any $m' = 1, 2, \ldots, y$, we must show that

$$\sum_{m=m'}^{y} p^j_{i,m}(k,R) \leq \sum_{m=m'}^{y} p^j_{i,m}(k+1,R),$$

which is equivalent to

$$\begin{aligned} &\sum_{m=m'}^{y}\left(\sum_{n=m-1}^{y} b_m\left(\min\{y,n+1\},\bar{F}(t_i - t_{i-1})\right)p^j_{i-1,n}(k,R)\right) \\ &\leq \sum_{m=m'}^{y}\left(\sum_{n=m-1}^{y} b_m\left(\min\{y,n+1\},\bar{F}(t_i - t_{i-1})\right)p^j_{i-1,n}(k+1,R)\right). \end{aligned}$$

Rearranging terms,

$$\begin{aligned} &\sum_{m=m'-1}^{y}\left(\sum_{n=m'}^{\min\{y,m+1\}} b_n\left(\min\{y,m+1\},\bar{F}(t_i - t_{i-1})\right)\right)p^j_{i-1,m}(k,R) \\ &\leq \sum_{m=m'-1}^{y}\left(\sum_{n=m'}^{\min\{y,m+1\}} b_n\left(\min\{y,m+1\},\bar{F}(t_i - t_{i-1})\right)\right)p^j_{i-1,m}(k+1,R), \end{aligned}$$

which holds because $\sum_{n=m'}^{\min\{y,m+1\}} b_n\left(\min\{y,m+1\},\bar{F}(t_i-t_{i-1})\right)$ is positive and non-decreasing in m and $\sum_{n=m}^{y} p_{i-1,n}^{j}(k,R) \leq \sum_{n=m}^{y} p_{i-1,n}^{j}(k+1,R)$ for any m by the original inductive hypothesis. Thus, we have shown our second inductive hypothesis to be true, which implies that the first inductive argument also holds.

3.4.2 Optimal Admission Policy

We conjecture that the optimal policy given a reservation vector R is a threshold policy in which the reservation is accepted if the number of busy units k is lower than some threshold value. This conjecture is not contradicted by any experiments of Section 3.6 for the optimal policy described in Section 3.4.3 when all arrival epochs are known.

The most direct way to prove the existence of a threshold policy is to prove that the profit is concave in the number of busy units using an inductive argument starting with the last period and working backwards. However, the modeling combination of lost sales and the reservation vector for our model produces complications that are avoided by Papier and Thonemann (2010) through only focusing on admitting customers whose service starts immediately. Specifically, complications arise as whether service for each existing reservations starts or fails must be tracked. That such difficulties occur fits in with other literature on lost sales models described by Levi et al. (2008), who note that future costs in lost sales models depend not on some summary statistic but on the actual sequence of orders in the pipeline.

Next, we demonstrate why the approach to proving concavity of the profit in k of Papier and Thonemann (2010), who use sample path coupling arguments, fails for our model.

Proposition 3.1. *The profit function $pi_J(k,R)$ is not necessarily concave in k for any given a reservation vector R.*

Proof. We provide a counterexample by showing that $\pi_J(k,R)$ is not necessarily concave in k. We focus on the last reservation acceptance decision $a_J(k,R)$ with $\tau=1$ and $R=\{0.9\}$ for a system with $y=2$ rental units. For cost and revenue parameters, we have expected revenue from service $r/\mu=1$, expected service time $1/\mu=2$, rejection cost $c_R=0.1$ and failure-to-serve cost $c_F=1$. The probability that the single existing reservation in R fails to be served is

$$\begin{aligned}
p_{1,2}^{J}(0,\{0.9\}) &= 0,\\
p_{1,2}^{J}(1,\{0.9\}) &= 0,\\
p_{1,2}^{J}(2,\{0.9\}) &= \left(\bar{F}(0.9)\right)^2 = 0.407.
\end{aligned}$$

If $a_J(k,\{0.9\})=1$, we also need to account for the probability that the Jth reservation request is not fulfilled:

$$\begin{aligned}
\beta_2^{J}(0,\{0.9\}) &= 0,\\
\beta_2^{J}(1,\{0.9\}) &= \bar{F}(1)\bar{F}(0.1) = 0.577,\\
\beta_2^{J}(2,\{0.9\}) &= \bar{F}(0.1)^2\left(1-F(0.9)^2\right) = 0.786.
\end{aligned}$$

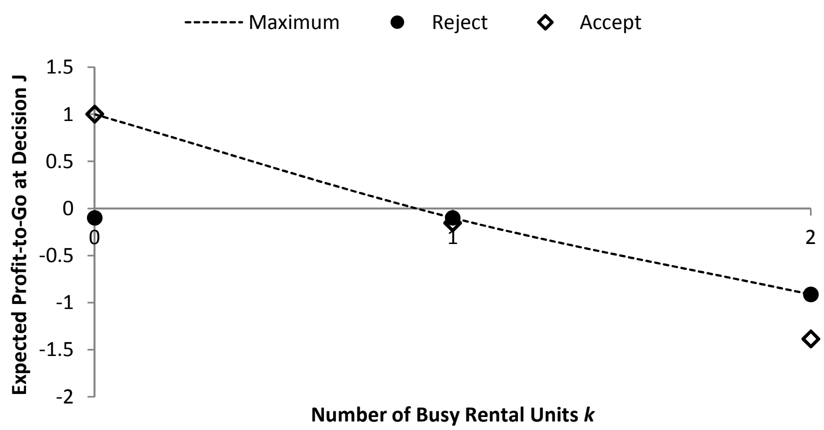


Figure 3.1: Expected profit-to-go violates concavity for the counterexample.

Rejecting reservation request J gives an expected profit of

$$\begin{aligned}
\pi_J^0(0, \{0.9\}) &= -c_R = -0.1, \\
\pi_J^0(1, \{0.9\}) &= -c_R = -0.1, \\
\pi_J^0(2, \{0.9\}) &= -c_R - (c_F + r/\mu)p_{1,2}^J(2, \{0.9\}) = -0.913,
\end{aligned}$$

while accepting reservation request J gives an expected profit of

$$\begin{aligned}
\pi_J^1(0, \{0.9\}) &= r/\mu = 1, \\
\pi_J^1(1, \{0.9\}) &= r/\mu - (c_F + r/\mu)\beta_2^J(1, \{0.9\}) = -0.154, \\
\pi_J^1(2, \{0.9\}) &= r/\mu - (c_F + r/\mu)\left(p_{1,2}^J(2, \{0.9\}) + \beta_2^J(2, \{0.9\})\right) = -1.385.
\end{aligned}$$

Thus,

$$\begin{aligned}
\pi_J(0, \{0.9\}) &= \max\{\pi_J^0(0, \{0.9\}), \pi_J^1(0, \{0.9\})\} = 1, \\
\pi_J(1, \{0.9\}) &= \max\{\pi_J^0(0, \{0.9\}), \pi_J^1(0, \{0.9\})\} = -0.1, \\
\pi_J(2, \{0.9\}) &= \max\{\pi_J^0(0, \{0.9\}), \pi_J^1(0, \{0.9\})\} = -0.913.
\end{aligned}$$

As shown in Figure 3.1, while $\pi_J^0(k, \{0.9\})$ and $\pi_J^1(k, \{0.9\})$ are concave in k, $\pi_J(0, \{0.9\})$ is not concave in k due to the transition from acceptance to rejection as the optimal decision when $k = 1$.

3.4.3 Upper Bound with Known Arrival Epochs

Due to the problem's definition in continuous time, the inclusion of the infinite-dimensional reservation vector in the state space makes the problem extremely difficult to solve as a continuous-time dynamic program. Attempting to solve the problem by discretizing the state space would still face the challenge of the number of possible combinations for the reservation vector. Thus, we compute an upper bound in which all arrival epochs are known.

If all arrival epochs $\boldsymbol{\xi}_A = \{A_1, A_2, \ldots, A_J\}$ are known, the optimal policy can be computed as a stochastic dynamic program through backwards induction. We label the value of the solution resulting from such as procedure as $OptUB$. Each J arrival events and J service start events constitute a stage of a finite-state discrete-event dynamic program in which the number of busy rental units at each arrival epoch is the state. The action space is $\{a_1, a_2, \ldots, a_J\} \in \{0,1\}^J$ and applies only to arrival events. We define E_j as the time of event j, $\theta_j \in \{A, S\}$ as the type of event (arrival or service), and $a_{(j)}$ as the acceptance decision corresponding to event j for $j = 1, 2, \ldots, 2J$. The state at the beginning of each stage j is k and all actions taken within τ time units prior to E_j.

For known $\boldsymbol{\xi}_A$, we let $\pi_j^{\boldsymbol{\xi}_A}(k)$ represent the expected profit function for the jth arrival, and define it as follows:

$$\pi_j^{\boldsymbol{\xi}_A}(k) = \max_{a\in\{0,1\}} \begin{cases} ra - c_R(1-a) + \sum_{k'=0}^{k} b_{k'}(k, \bar{F}(t_{j+1} - E_j))\pi_{j+1}^{\boldsymbol{\xi}_A}(k') & \text{if } \theta_j = A, \\ (-r - c_F)\mathbf{1}\{k = y, a_{(j)} = 1\} & \\ + \sum_{k'=0}^{k+a_{(j)}} b_{k'}(\min\{k + a_{(j)}, y\}, \bar{F}(t_{j+1} - E_j))\pi_{j+1}^{\boldsymbol{\xi}_A}(\min\{k', y\}) & \text{if } \theta_j = S. \end{cases}$$

We use backwards recursion to solve this system and determine the optimal policy for a system with sample path $\boldsymbol{\xi}_A$.

3.4.4 Erlang Loss Performance Bounds

As $\tau \to 0$ and $T \to \infty$, the system behaves as an M/M/c/c queue with a blocking probability P_b given by the Erlang loss formula (Erlang 1917). That is, for a system with mean service time $1/\mu$ and y rental units, the steady-state probability that all units are busy is

$$P_b := \frac{\frac{\left(\frac{\lambda}{\mu}\right)^y}{y!}}{\sum_{i=0}^{k} \frac{\left(\frac{\lambda}{\mu}\right)^y}{i!}}, \qquad k = 0, 1, \ldots, y.$$

By using this formula to estimate the service or rejection rate, we derive an upper and lower bound on the system's profit. As an upper bound, we consider a system in which an omniscient decision maker accepts a reservation if and only if it will be successfully fulfilled so that no failures occur. As a lower bound, we consider a policy in which all reservations are accepted.

We first define the upper bound $ErlUB$ for the profit over the rental horizon based on an omniscient decision maker:

Proposition 3.2. *The steady-state probability of the M/M/c/c queue over the horizon with service at the end of the rental horizon provides an upper bound on the service rate; i.e.,*

$$\frac{\min\left\{\lambda T, \left(T + \frac{1}{\mu}\right)(1 - P_b)\left(\frac{\lambda}{\mu}\right)\right\}}{\lambda T} \geq \frac{\mathbb{E}\left[N^Y\right]}{\lambda T}.$$

Furthermore, the expected profit from the M/M/c/c queue with perfect admission decisions provides the following upper bound on the expected profit:

$$\pi^{ErlUB} = \frac{r}{\mu}\min\left\{\lambda T, \left(T + \frac{1}{\mu}\right)(1 - P_b)\lambda\right\} - c_R \max\left\{0, \lambda T - \left(T + \frac{1}{\mu}\right)(1 - P_b)\lambda\right\} \geq \pi^*$$

Proof. The proof relies on a sample path argument in which the remaining service times of any busy rental units are resampled upon each arrival event. The key difference between the M/M/c/c queue and our rental model is the distribution of the state of the system at time 0. We define $k^{ErlUB}(t)$ and $k(t)$ as the number busy for the M/M/c/c system and our model, respectively. For the M/M/c/c queue, it is the queue's stationary distribution; i.e., $k^{ErlUB}(0) \geq 0$. For our model, it begins with $k(0) = 0$ busy rental units, according to how we have defined our model. By Lemma 3.1, it is clear to see that $k^{ErlUB}(t) \geq k(t)$ for $t \geq 0$. This implies that $k^{ErlUB}(T+\tau) \geq k(T+\tau)$, and the steady-state probabilities also provide an upper bound on the number of rentals that receive service at the end of the horizon.

We note that we cannot simply use $\lambda T P_b$ as an upper bound on the number of accepted reservations, as the system starts with 0 rental units in service rather than at steady-state. In particular, the acceptance rate for the real system would be higher than $(1-P_b)$ early in the horizon. Thus, we must use the total amount of service provided over the rental horizon for our upper bound. Over the interval $[\tau, T+\tau]$, the system is in steady state, which corresponds to $(1-P_b)\lambda/\mu$ servers busy in expectation for T time units. At time $T+\tau$, there are again $(1-P_b)\lambda/\mu$ servers busy in expectation, and each busy server has a remaining service time of $1/\mu$. Therefore, for a system that accepts reservations over a horizon of T time units and allows any rental units in service at time $T+\tau$ to complete service, we can compute an upper bound on the expected total service provided,

$$\left(T + \frac{1}{\mu}\right)(1-P_b)\left(\frac{\lambda}{\mu}\right) \geq \frac{\mathbb{E}\left[N^Y\right]}{\mu}$$

which gives an upper bound on the expected number of rental units served by dividing by the mean service time $1/\mu$,

$$\left(T + \frac{1}{\mu}\right)(1-P_b)\lambda \geq \mathbb{E}\left[N^Y\right].$$

Because the total expected number of arrivals over the horizon is λT, we can restrict this expression to

$$\min\left\{\lambda T, \left(T + \frac{1}{\mu}\right)(1-P_b)\lambda\right\} \geq \mathbb{E}\left[N^Y\right].$$

By assuming that any reservation requests that are not accepted are rejected, we use our upper bound on the number served to get a lower bound on the total number of rental units that cannot be served,

$$\max\left\{0, \lambda T - \left(T + \frac{1}{\mu}\right)(1-P_b)\lambda\right\} \leq \mathbb{E}\left[N^N + N^F\right] = \mathbb{E}\left[N^N\right],$$

because $N^F = 0$ for the $ErlUB$ policy. As every reservation request that is not accepted and successfully fulfilled is either accepted and unsuccessfully fulfilled or rejected, we assume that all such demand is rejected to incur cost $c_R < c_F$. Thus, the expected number of accepted reservations is maximized and the expected penalty of all unserved demand is minimized, and $\pi^{ErlUB} \geq \pi^*$.

The M/M/c/c queueing model also allows us to provide a lower bound for the performance of a policy *ErlLB* in which all reservation requests are accepted using an upper bound on the number of reservations for which service fails.

Proposition 3.3. *The steady-state probability of the M/M/c/c queue over the horizon with service at the end of the rental horizon provides an upper bound on the rate at which accepted reservations cannot be fulfilled; i.e.,*

$$P_b \geq \frac{\mathbb{E}\left[N^N\right]}{\lambda T}.$$

Furthermore, the expected profit from the M/M/c/c queue provides the following lower bound on the expected profit using a policy in which all reservations are accepted:

$$\pi^{ErlLB} = \frac{r}{\mu}\lambda T(1-P_b) - c_F \lambda T P_b \leq \pi^*.$$

Proof. The proof of this proposition also relies on a sample path argument and is similar to that of Proposition 3.2. As before, the key difference between the M/M/c/c queue and our model is the distribution of the state of the system at time 0. Using the same sample path argument as in Proposition 3.2, we observe that $k^{ErlLB}(t) \geq k(t)$ for $t \geq 0$. Because the state for the M/M/c/c queue is stochastically greater than the state of our rental system at every point in time, the rate at which arrivals are blocked for the M/M/c/c queue exceeds the rate of failed reservations when all are accepted in our rental model. In particular,

$$\lambda T P_b \geq \mathbb{E}\left[N^F\right],$$

which gives a lower bound for the service rate of

$$\lambda T(1-P_b) \leq \mathbb{E}\left[N^Y\right].$$

Because $\mathbb{E}\left[N^Y\right] = 0$ in a policy in which all reservations are accepted, we combine the bounds on the failure rate and service rate to get a lower bound on the expected profit under a policy in which all reservations are accepted,

$$\pi^{ErlLB} = \frac{r}{\mu}\lambda T(1-P_b) - c_F \lambda T P_b \leq \pi^*.$$

3.5 Heuristics

We present simple heuristics for the acceptance decision in this section. We begin by assuming deterministic service times and then provide a newsvendor-style heuristic that considers whether a reservation is able to be successfully fulfilled as a random variable.

3.5.1 Policies Assuming Deterministic Service Times

We first introduce three policies in which service times are considered to be known, a common modeling assumption in lead-time quotation literature such as Kapuscinski and

Tayur (2007). These policies allow us to investigate the value of accounting for service time variability.

First, in the Med policy, the remaining service times of all busy rental units at the decision time t and the service times of all future reservations are considered to be the median service time $\ln(2)/\mu$. Given k busy rental units at the time of the decision and a reservation vector R, we keep track of whether each reservation i, $i = 1, 2, \ldots, |R|$, is successfully fulfilled with the variable $u_i \in \{0, 1\}$. If reservation i is successfully fulfilled, $u_i = 1$; otherwise, $u_i = 0$. Thus, we can define u_i recursively,

$$u_i^{Med} = \begin{cases} 1 & \text{if } k\mathbf{1}\left\{t_i < t + \frac{\ln(2)}{\mu}\right\} + \sum_{i'=1}^{i-1} u_{i'}^{Med}\mathbf{1}\left\{t_i < t_{i'} + \frac{\ln(2)}{\mu}\right\} < y, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{1}(\cdot)$ is the indicator function. The decision to accept or reject a reservation request is then

$$a^{Med}(k, R) = \begin{cases} 1 & \text{if } k\mathbf{1}\{t_i < t + \frac{\ln(2)}{\mu}\} + \sum_{i=1}^{|R|} u_i^{Med}\mathbf{1}\{t_i < t_i + \frac{\ln(2)}{\mu}\} < y, \\ 0 & \text{otherwise.} \end{cases}$$

Next, we describe a policy $Mean$ that uses the mean service time $1/\mu$ rather than the median service time $\ln(2)/\mu$. The decision to accept or reject a reservation request is

$$a^{Mean}(k, R) = \begin{cases} 1 & \text{if } k\mathbf{1}\{t_i < t + \frac{1}{\mu}\} + \sum_{i=1}^{|R|} u_i^{Med}\mathbf{1}\{t_i < t_i + \frac{1}{\mu}\} < y, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we consider the $Quant$ policy in which the remaining lifetimes are assumed to be constant but differ from the median as they are spread evenly over the quantiles of the exponential distribution. Specifically, each busy unit k', $k' = 1, 2, \ldots, k$, has a remaining duration of $-\ln(1 - k'/(k+1))/\mu$. Any reservations that start service after time t are assumed to have service time $\ln(2)/\mu$ as in the Med policy. The successful service variable is defined as follows:

$$u_i^{Quant} = \begin{cases} 1 & \text{if } \sum_{k'=1}^{k} \mathbf{1}\left\{t_i < t - \ln(1 - k'/(k+1))/\mu\right\} + \sum_{i'=1}^{i-1} u_{i'}^{Quant}\mathbf{1}\left\{t_i < t_{i'} + \frac{\ln(2)}{\mu}\right\} < y, \\ 0 & \text{otherwise.} \end{cases}$$

The acceptance decision for the $Quant$ policy is then

$$a^{Quant}(k, R) = \begin{cases} 1 & \text{if } \sum_{k'=1}^{k} \mathbf{1}\left\{t_i < t - \ln(1 - k'/(k+1))/\mu\right\} + \sum_{i=1}^{|R|} u_i^{Quant}\mathbf{1}\{t_i < t_i + \frac{\ln(2)}{\mu}\} < y, \\ 0 & \text{otherwise.} \end{cases}$$

3.5.2 A Stochastic Availability Heuristic

We now present a heuristic $Avail$ for determining whether to accept a reservation at some time t based on the service probability probability when the reservation would begin service. Specifically, the state probability distribution is tracked over the upcoming τ time units to determine the probability that at least one rental unit would be available for the reservation under consideration. The system state evolves as over $|R|+2$ stages: stage 0 at time t, stage

i corresponding to each reservation beginning service at time $t + t_i$ for $i = 1, 2, \ldots, |R|$, and stage $|R| + 1$ at time $t_{|R|+1} := t + \tau$ when the reservation under consideration would begin service. We let p_{in} represent the probability that the system is in state n at the beginning of stage i.

The starting state is based on the current number k of busy rental units, and is defined as

$$p_{0,n} = \mathbf{1}\{k = n\}.$$

As the event occurring at time t is merely a decision and not the start of service, the state probability distribution immediately before event 1 depends only on the number of reservations that complete service:

$$p_{i,n} = \sum_{n'=0}^{y} p_{0,n'} b_n(n', \bar{F}(t_1 - t_0)),$$

recalling that $b_n(n', p)$ is the probability that there are n successes for a binomial random variable with n' trials, each of which has a success probability p. For the remainder of the events, the state probability distribution at the beginning of each event can be defined recursively as follows:

$$p_{i,n} = \sum_{n'=0}^{y} p_{(i-1),n'} b_n(\min\{n' + 1, y\}, \bar{F}(t_1 - t_0)), \qquad i = 2, 3, \ldots, |R| + 1.$$

Thus, the probability that the system will have n busy rental units when the reservation under consideration would start service is $p_{|R|+1,n}$. We are particularly concerned with $p_{|R|+1,y}$, which is the probability that the reservation commitment is not able to be fulfilled. The *Avail* heuristic can then be defined as follows:

$$a^{Avail}(k, R) = \begin{cases} 1 & \text{if } (1 - p_{|R|+1,y})r + c_R \geq p_{|R|+1,y} c_F, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

This policy can be thought of as similar to a newsvendor policy for weighing the trade-offs of accepting and rejecting the reservation given that the reservation is the last reservation to arrive — i.e., it is the last acceptance decision over the horizon. We make a preliminary observation about this policy in comparison to the optimal policy. While it is identical to the optimal policy for the last arrival — assuming the last arrival is known — the heuristic is more likely to accept a reservation under the same circumstances at the time of a decision as the optimal policy.

Proposition 3.4. *If the optimal policy $a_j^*(k, R) = 1$ is to accept the reservation, then the policy from the Avail heuristic $a_j^{Avail}(k, R) = 1$ is to accept the reservation.*

Proof. We show that $a_j^{Avail}(k, R) = 0$ implies that $a_j^*(k, R) = 0$ for any $j = 1, 2, \ldots, y$. We write the optimal policy for any decision j' after arrival j on any sample path as $a_{j'}^*(\boldsymbol{\xi}|a_j)$. By Lemma 3.1 using any set of actions $a_{j+1}, a_{j+2}, \ldots, a_J$, the state of the system for $a_j = 1$ is stochastically greater than that of $a_j = 0$. Thus, we take the decision sequence

Table 3.1: Simulation parameters.

Parameter	Rent the Runway	Redbox	Railroad
Time Horizon T (days)	183	60	365
Arrival Rate λ (per week)	5	14	2.7
Mean Service Time $1/\mu$ (days)	10	2	11.2
Notice Time τ (days)	7	2	7
Expected Revenue per Rental r/μ (\$)	59	2.4	350
Rejection Penalty c_R (\$)	5	1	0
Service Failure Penalty c_F (\$)	30 or 118	20	850

$a^*_{j+1}(\boldsymbol{\xi}|a_j = 1), a^*_{j+2}(\boldsymbol{\xi}|a_j = 1), \ldots, a^*_J(\boldsymbol{\xi}|a_j = 1)$ corresponding to the optimal policy on any sample path with $a_j = 1$, and apply it to a system with $a_j = 0$. By Lemma 3.1, the profit from time $A_{j+1} + \tau$ onward can only increase as the expected number of accepted reservations (i.e., $\mathbb{E}\left[\sum_{j'=j+1}^{J} a^*_{j'}(\boldsymbol{\xi}|a_j = 1)\right]$ stays the same and the expected number of failed reservations can only stay the same of decrease. Furthermore, relaxing the constraint of using the decision sequence $a^*_{j+1}(\boldsymbol{\xi}|a_j = 1), a^*_{j+2}(\boldsymbol{\xi}|a_j = 1), \ldots, a^*_J(\boldsymbol{\xi}|a_j = 1)$ can only increase the value of the optimal policy.

We now consider the profit corresponding to the service of the jth arrival. Going from $a_j = 1$ to $a_j = 0$ will increase the profit corresponding to the jth reservation, as $a_j^{Avail}(k, R) = 0$ implies that $(1 - p_{|R|+1,y})r + c_R < p_{|R|+1,y}c_F$. Therefore, the expected profit of accepting the jth reservation is negative by the system state equations. Because rejecting the jth reservation improves the optimal expected profit corresponding to customers $j, j + 1, \ldots, J$, $a_j^{Avail}(k, R) = 0$ implies that $a^*_j(k, R) = 0$.

3.6 Numerical Results

In this section, we study the performance of the various heuristics in numerical experiments, and use secondary system metrics to explain the difference in performance. These insights lead to more advanced heuristics in Section 3.7. We also use the numerical results to develop business insights about reservation policies; in particular, we show that the expected profit is decreasing in the notice time, a relationship that runs counter to traditional insights about advance demand information in non-rentals inventory settings.

Our experiments are primarily motivated by Rent the Runway, and we also use cases motivated by the railcar rental business discussed in Papier and Thonemann (2010) and Redbox, which rents movies and games through kiosks. We evaluate the impact of the demand notice time, the number of rental units, and the penalty for failing to serve an accepted reservation on various system metric. Key metrics include the total revenue, total penalties, total profit, the rate at which reservations are accepted and successfully fulfilled, the rate at which reservations are accepted but not fulfilled, and the rate at which reservations are rejected.

Parameters for the experiments are presented in Table 3.1. The parameters used for Rent the Runway are similar to that of the base model presented in the previous chapter. For the mean rental time, we use a four-day rental with a conservative estimate of three days for shipping and cleaning before after the rental. The exponential distribution with a mean of ten days provides a rough estimate for the rental duration when early returns, customers who choose the eight-day rental option, customer return time variability, shipping variability, and time required for repairs are considered. For the penalty of failing to serve an accepted reservation, we use values equal to one-half and twice the expected net revenue per rental.

We also evaluate our reservation acceptance heuristics in a scenario motivated by Redbox. Based on an average transaction value of $2.49 reported in its 2013 annual report and a $1.20 per night movie rental fee, we estimate the rental duration as exponentially distributed with mean of two days. We set the notice time to two days. Redbox currently accepts reservations on the day on which the rental begins if there is a movie already in inventory that can be reserved.

Finally, we conduct an experiment using parameters from the "Type 1" model of the railcar business described in Papier and Thonemann (2010). We limit the demand to the customers who make reservations one week in advance, which comprise 30% of all customers.

In our simulations, generating the optimal policy for a sample path of arrivals is the most time-intensive process. However, since we can apply the optimal policy for multiple service time sample paths, we run multiple simulations with random service times for each realization of demand arrival times. Specifically, we use 75 sets of replications so that the width of the 95% confidence interval for the expected profit is within 2% of the expected profit of the $OptUB$ policy. In each set of replications, we generate all arrival times over the horizon and then generate 200 replications for the service times. This strategy allows us to reduce the number of times that we have to compute the upper bound policy. Reservations accepted before the end of the horizon are served to completion without any reservations being accepted after the end of the horizon.

We compare the performance of the following bounds and heuristic policies:

- $OptUB$. The upper bound provided by a dynamic program for a given set of arrival times, as described in Section 3.4.3.

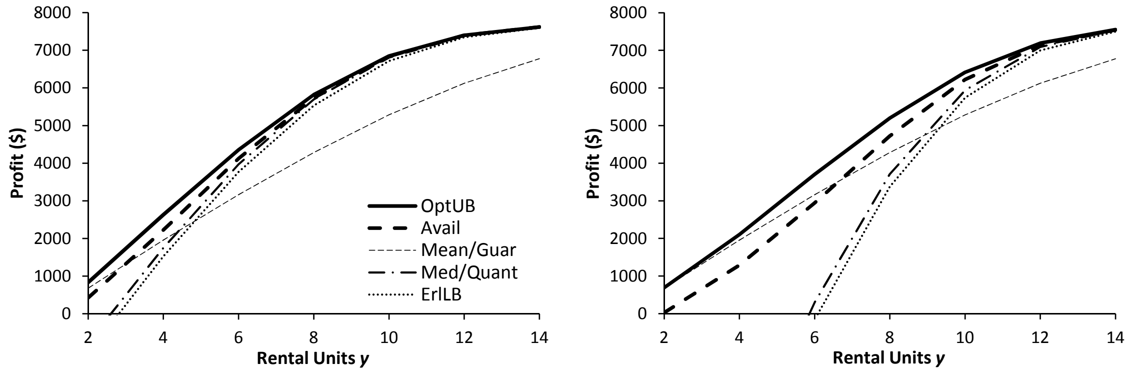
- $Avail$. The performance of a heuristic based on the probability that a rental unit will be available at the time the reservation begins, as described in Section 3.5.2.

- Med. The performance of a heuristic that accepts reservations assuming that all remaining service has a duration equivalent to the median service time, as described in Section 3.5.1.

- $Mean$. The performance of a heuristic that accepts reservations assuming that all remaining service has a duration equivalent to the mean service time, as described in Section 3.5.1.

- $Quant$. The performance of a heuristic that accepts reservations assuming that remaining service times for busy rental units are evenly spread across the quantiles of the service time distribution, as described in Section 3.5.1.

- All. The performance of a heuristic that accepts all reservations.



(a) Low failure penalty ($c_F = r/2\mu$) (b) High failure penalty ($c_F = 2r/\mu$)

Figure 3.2: Expected profit for Rent the Runway test cases as a function of the number of rental units.

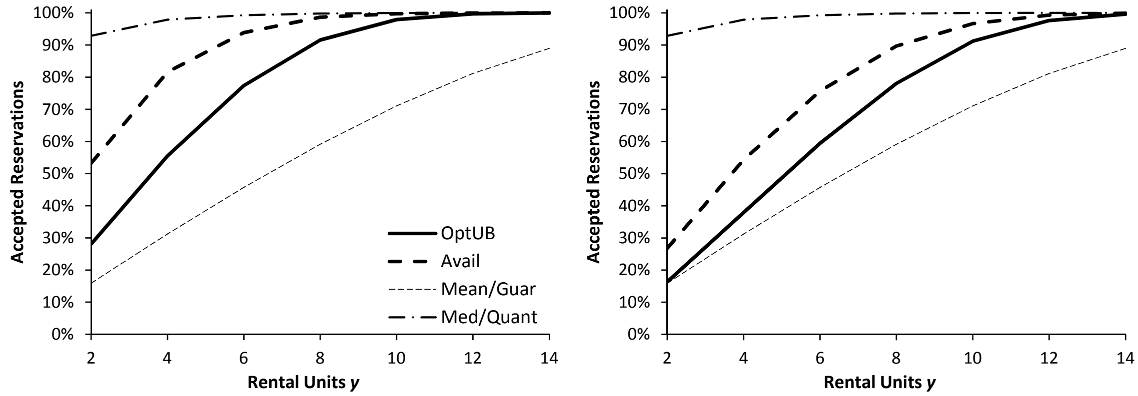
- *Guar*. The performance of a heuristic in which a reservation is accepted only if a rental unit is guaranteed to be available; i.e., $a_j^{Guar}(k, R) = \mathbf{1}\{k + |R| < y\}$.
- *ErlUB*. The analytical upper bound provided by the M/M/c/c queueing model discussed in Section 3.2.
- *ErlLB*. The analytical lower bound provided by the M/M/c/c queueing model discussed in Section 3.2.

3.6.1 Rent the Runway Test Cases

In Figure 3.2, we show the expected profit for the various bounds and heuristics using different rental inventory levels for low and high failure-to-serve costs. To help explain the differences among the policies, Figure 3.3 displays the corresponding acceptance rate to evaluate how aggressive or conservative a policy is. Figure 3.4 shows the percentage of accepted reservations that cannot be served, thus incurring a penalty.

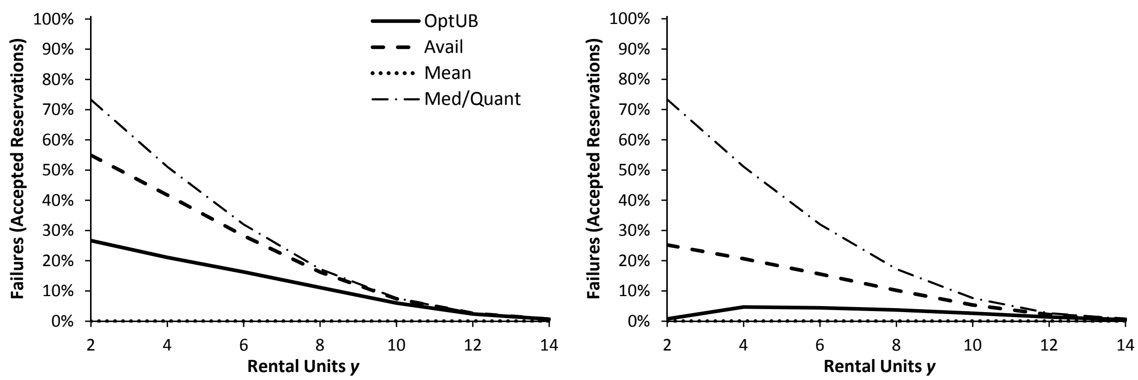
We first comment on the difference between the two upper bounds, *ErlUB* and *OptUB*. For between 2 and 10 rental units with $c_F = r/2\mu$, the *ErlUB* bound exceeds the *OptUB* bound by between \$550 and \$850. This reflects the risk of failing to serve an accepted reservation for the *OptUB*, as it is assumed that any reservations that are not served were rejected in the *ErlUB* policy. As the number of rental units increases, the two policies appear to converge to the same value as the reservation acceptance rate for the *OptUB* converges to 100%. Because the *OptUB* bound dominates the *ErlUB* bound, we omit showing it in the figures.

The expected profit for the *OptUB* bound exceeds that of the *Avail* heuristic by no more than \$850 for either penalty cost value. For $y \geq 6$ when $c_F = r/2\mu$ and $y \geq 10$ when $c_F = 2r/\mu$, the *Avail* heuristic achieves at least 95% of the expected profit. In all cases, the *Avail* heuristic accepts a higher percentage of the reservation requests than the *OptUB* heuristic, an observation that is not surprising given Proposition 3.4 for the relationship between the *Avail* heuristic and the optimal policy. Naturally, it also is unable to fulfill a higher percentage of accepted reservations. For example, when $y = 8$ when $c_F = 2r/\mu$, the



(a) Low failure penalty ($c_F = r/2\mu$) (b) High failure penalty ($c_F = 2r/\mu$)

Figure 3.3: Expected acceptance rate for Rent the Runway test cases as a function of the number of rental units.



(a) Low failure penalty ($c_F = r/2\mu$) (b) High failure penalty ($c_F = 2r/\mu$)

Figure 3.4: Expected failure-to-serve rate among accepted reservations for Rent the Runway test cases as a function of the number of rental units.

Avail heuristic accepts 89.7% of reservations but is unable serve 10.2% of the reservations. By contrast, the *OptUB* bound policy only accepts 78.1% of reservations, and is unable to serve 3.7% of the accepted reservations.

For these scenarios, the *Med* and *Quant* policies both perform almost identically to the *All* policy in which all reservations are accepted. Noting that these two policies do not change with the failure penalty c_F, at least 99% of reservations are accepted for $y \geq 6$. Because $\tau > \ln(2)/\mu$, the *Med* heuristic assumes that no rental units that are busy remain busy when they would be needed for the reservation under consideration. The additional spread in the remaining service times for busy rental units from the *Quant* policy does not make a significant difference in the acceptance rate.

Compared the *Med* policy, the *Mean* policy has the opposite effect because $\tau < 1/\mu$. It assumes that any busy rental units will remain busy in addition to any reservations that begin before the reservation under consideration. Thus, it operates as a holdback policy in which a reservation is only accepted if there is a rental unit whose availability over the entire notice interval is insured; i.e., the *Guar* policy. For a low number of rental units, the profit from the *Mean* policy exceeds that of the other heuristics. When $y = 2$ when $c_F = r/2\mu$, the *Mean* heuristic only accepts 15.9% of reservations while 28.1% are accepted in the *OptUB* policy. All accepted reservations are successfully fulfilled under the *Mean* heuristic compared to 73.3% for *OptUB*. However, as the number of rental units increases, the *Mean* heuristic's relative performance decreases. It performs worse than the heuristics that accept all reservations for $y \geq 6$ when the failure penalty is low and for $y \geq 10$ when the failure penalty is high. The poor performance of the *Mean* policy also indicates the value of policies more aggressive than a holdback policy. As the number of rental units increases, there is significant value in accepting reservations even when there is not a rental unit guaranteed to be available.

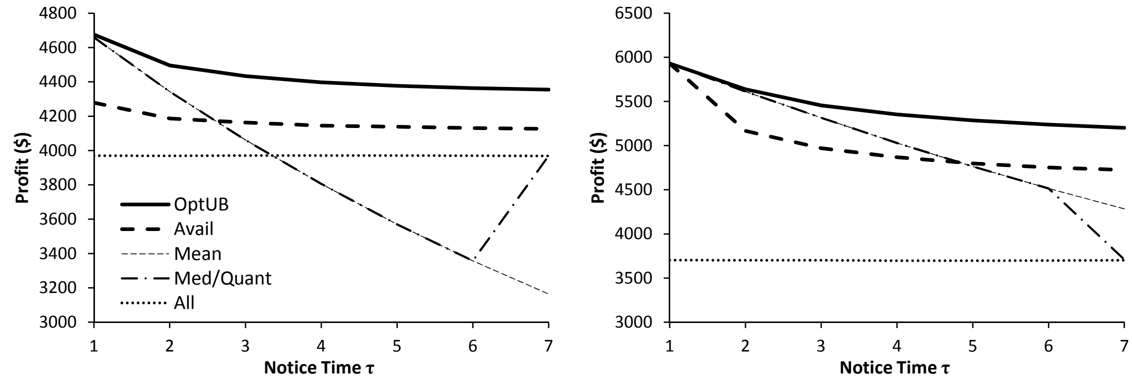
We also study the effect of reducing the notice time from seven days to one day, as shown in Figure 3.5. As expected, the upper bound on the expected profit increases as the notice time decreases by approximately \$200 for low c_F or \$500 for high c_F. This reflects the decreased uncertainty about whether a rental unit is available to serve a reservation. For the *Avail* policy, the profit is within \$500 of the upper bound regardless of notice time. As the notice time decreases, the *Avail*, *Med*, and *Mean* policies converge to the optimal policy, which is to only accept reservations when there is a rental unit available as τ approaches zero. Both *Med* and *Mean* have the same performance for $\tau \leq 6$. For $\tau = 7$, the *Med* policy admits all reservations, which increases the expected profit for low c_F and decreases the expected profit for high c_F.

Figure 3.5 indicates that the optimal profit is decreasing in the notice time, an insight that the value of advance demand information is limited by how far in advance the information is received. This relationship distinguishes the operation of a rental system from a more conventional inventory system described in Hariharan and Zipkin (1995). Assuming demand and prices do not change (or would change favorably) with the notice time, a rental firm may benefit by reducing the notice time. In effect, this is similar to a strategy used by some rental businesses (e.g., Redbox) of only accepting reservations up to one day in advance.

3.6.2 Railcar and Redbox Test Cases

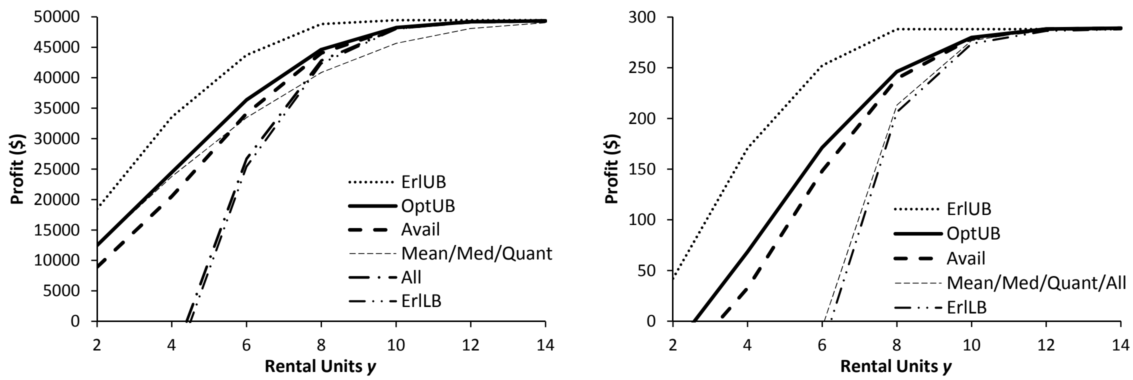
We next analyze the two additional test cases described above focusing on a railcar rental business and the movie rental business Redbox. The railcar test case involves less frequent arrivals and a high penalty for a service failure compared to Rent the Runway, and reinforces the trends observed for the Rent the Runway case. The *Avail* policy achieves 83.9% of the expected profit of the *OptUB* bound for $y = 4$, 98.6% for $y = 8$ and exceeds 99.9% for $y \geq 12$. The *Mean*, *Med*, and *Quant* heuristic policies play a similar role as $\tau > \ln(2)/\mu$ and $\tau < 1/\mu$ again. The *Mean* heuristic — a holdback policy — outperforms the other heuristics for $y \leq 4$ and is outperformed by all other heuristics for $y \geq 4$. Due to the relatively high cost of a service failure, it is not surprising that the conservative *Mean*, *Med*, and *Quant* policies, which act as holdback policies, perform well for $y \leq 6$.

However, the Redbox test case shows the sensitivity of the *Mean*, *Med*, and *Quant* heuristics to the notice time and service rate parameters. In this case, both $\tau > \ln(2)/\mu$



(a) Low failure penalty ($c_F = r/2\mu$) with $y = 6$ (b) High failure penalty ($c_F = 2r/\mu$) with $y = 8$

Figure 3.5: Expected profit for Rent the Runway test cases as a function of the notice time τ.



(a) Railcar Test Case (b) Redbox Test Case

Figure 3.6: Expected profit as a function of the number of rental units for additional test cases.

and $\tau > 1/\mu$, which means that these three heuristics operate approximately as the *All* policy in which all reservations are accepted. Because both the ratios of the arrival rate to the service rate and the service failure penalty to the revenue per rental are high relative to Rent the Runway, the results match our expectation that an aggressive policy would perform poorly for low values of y. We also note that the *Avail* heuristic outperforms all other heuristics for any number of rental units. Its expected profit is within $40 of that of the $OptUB$ policy, and it achieves at least 97.0% of the optimal upper bound for $y \geq 8$.

3.7 Improvements to the *Avail* Heuristic

Noting that the *Avail* heuristic is always more aggressive than the optimal policy, we introduce two additional heuristics that are more conservative than the *Avail* heuristic. Both heuristics compare the decision to accept the jth reservation to the decision to accept the $(j+1)$st reservation. If accepting the $(j+1)$st reservation is more favorable than accepting the jth reservation, the policy is to wait and make the same comparison upon the next arrival. We note this does not necessarily mean that the the $(j+1)$st reservation will be accepted, as the same comparison must be made between the $(j+1)$st and $(j+2)$nd reservations.

Because the failure-to-serve probability will always be lower for the $(j+1)$st reservation, we modify the value of serving each reservation so that the heuristic does not always choose serving the $(j+1)$st reservation over the jth reservation. Specifically, we assume that the both the jth reservation and the $(j+1)$st reservation complete service by some future time. Thus, the expected revenue from serving the jth reservation is always greater than that from serving the $(j+1)$st reservation, but will be less than or equal to the expected revenue r/μ without the completion time assumption. Using $p'_{|R|+1,y}$ to denote the probability that y rental units are busy at time $A_j + 1/\lambda + \tau$ after the jth reservation was rejected, we define two heuristics corresponding to different versions of this last completion time:

- *AvailMean*, which uses a last completion time of $A_j + \tau + 1/\lambda + 1/\mu$ so that

$$a^{AvailMean}(k,R) = \begin{cases} 1 & \text{if } p_{|R|+1,y}c_F + (1-p_{|R|+1,y})\,(1-\exp(-\mu(1/\mu+1/\lambda)))\,r \\ & \qquad \geq p'_{|R|+1,y}c_F + (1-p'_{|R|+1,y})\,(1-\exp(-1))\,r, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

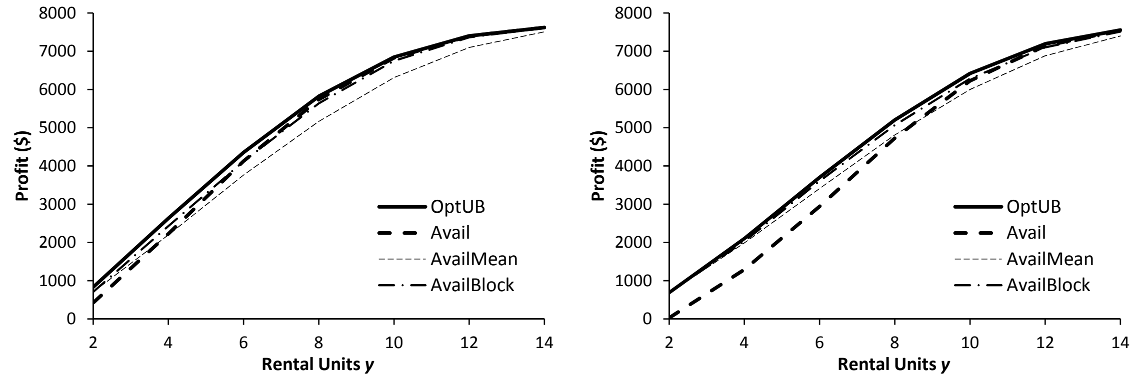
  In this comparison, the $p_{|R|+1,y}c_F$ term represents the failure probability and cost. The other term represents the probability of successfully serving the reservation and the expected service time conditional on service being less than or equal to some time duration. For the jth reservation, this duration is $1/\mu + 1/\lambda$ time units; for the $(j+1)st$ reservation, this duration is $1/\mu$ time units.

- *AvailBlock*, which uses a last completion time of $A_j + \tau + 1/\lambda + P_b/\mu$ and is expressed as

$$a^{AvailBlock}(k,R) = \begin{cases} 1 & \text{if } p_{|R|+1,y}c_F + (1-p_{|R|+1,y})\,(1-\exp(-P_b\mu(1/\mu+1/\lambda)))\,r \\ & \qquad \geq p'_{|R|+1,y}c_F + (1-p'_{|R|+1,y})\,(1-\exp(-P_b))\,r, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

  The inequality builds upon that of *AvailMean* by introducing the blocking probability. Specifically, the duration by which service is assumed to have been completed is multiplied by the blocking probability. Incorporating the blocking probability serves to give higher weight to choosing to accept the jth reservation when the system is generally less busy. As the blocking probability increases to 1, the *AvailBlock* heuristic becomes the same as the *AvailMean* heuristic.

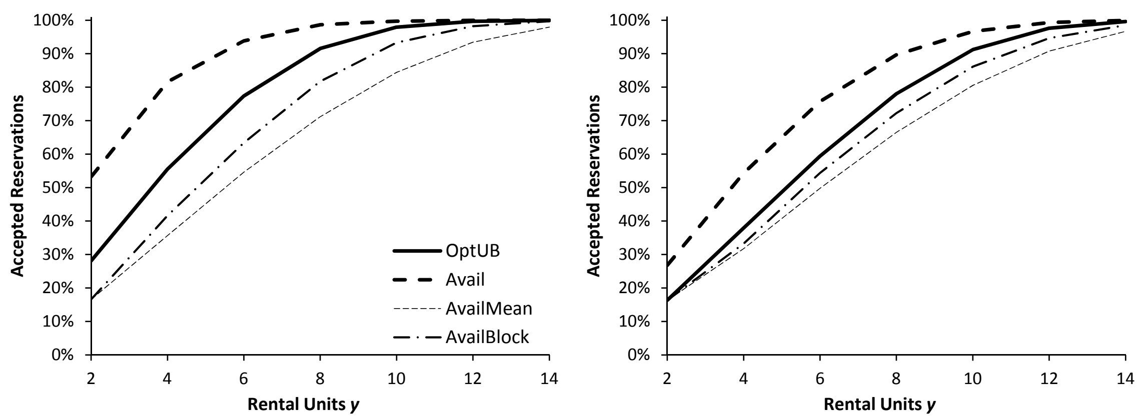
We now extend our numerical testing from Section 3.6.1 to include the *AvailMean* and *AvailBlock* heuristics, and see that the *AvailBlock* heuristic performs well compared to the optimal solution. We also note that *AvailBlock* outperforms *AvailMean* under every scenario, as depicted in Figure 3.7. As evidenced by the acceptance rates shown in



(a) Low failure penalty ($c_F = r/2\mu$)

(b) High failure penalty ($c_F = 2r/\mu$)

Figure 3.7: Expected profit as a function of the number of rental units for $AvailMean$ and $AvailBlock$ heuristics.



(a) Low failure penalty ($c_F = r/2\mu$)

(b) High failure penalty ($c_F = 2r/\mu$)

Figure 3.8: Expected acceptance rate as a function of the number of rental units for $AvailMean$ and $AvailBlock$ heuristics.

Figure 3.8, we observe that being more aggressive than $AvailMean$ improves the expected profit. As expected, the $AvailMean$ and $AvailBlock$ heuristics have a lower acceptance rate than the $Avail$ heuristic, and $AvailBlock$ is more aggressive than $AvailMean$. We can also see that the acceptance rate is lower than that of $OptUB$, which is also to be expected as $OptUB$ can be more aggressive given its perfect knowledge of the next arrival epoch.

The $AvailBlock$ heuristic outperforms $Avail$ when the number of rental units is low, but the $Avail$ heuristic improves upon the $AvailBlock$ heuristic when the number of rental units is sufficiently high. However, if $Avail$ outperforms $AvailBlock$, the expected profit from the $AvailBlock$ heuristic still achieves at least 98.7% of the upper bound on the optimal expected profit when the failure penalty is low and at least 94.6% of the upper bound on the optimal expected profit when the failure penalty is high.

3.8 Conclusions

In this chapter, we studied the reservation acceptance decision for a stochastic model of a rental system. Rental businesses, such as Rent the Runway, in which customers need to rent an item starting on a specific date and in which there is significant variability in the return date provided the primary motivation for our model. We proposed a new model in which the decision is whether to accept customers' reservation requests in advance of their rental start time based on the current number of busy rental units and the existing set of accepted reservations. We discussed the challenges in proving that the optimal reservation acceptance policy given a set of accepted reservations is a threshold policy in which it is optimal to accept the reservation if the number of busy rental units is less than or equal to some threshold value.

We then developed a heuristic newsvendor-style policy and studied its performance on test cases related to Rent the Runway, a railcar rental business, and a kiosk-based movie rental business. We compared it to lower and upper bounds based on the Erlang loss formula and an upper bound from the optimal policy for system with known arrival epochs, and found it to perform close to optimal except for when there are very few rental units. Through this heuristic, we showed the value of accepting reservations even if successfully fulfilling the reservations is not guaranteed. The simple heuristic provides decision makers with a simple algorithm to guide reservation acceptance decisions in online setting, and provides a starting point for providing an availability calendar.

In future work, we hope to attempt an alternate modeling framework and generalize the model to account for other model elements. In particular, a discrete-time model similar to Kapuscinski and Tayur (2007) with simpler (e.g., Bernoulli) demand or rental duration distributions could allow for structural results about the optimal policy. For additional model elements, usage-based loss as defined in Chapter 2 provides an interesting and important business feature to study. Similarly, customer no-shows or cancellations may affect the system's performance and could be included in the model. Finally, stochastic notice times may be important to study for many applications. For these three possible future paths, our heuristic policy could be easily extended, and we suspect that similar structural results can be developed.

Chapter 4

The Pennsylvania Adoption Exchange Improves Its Matching Process

With Mustafa Akan, Onur Kesten, and M. Utku Ünver

4.1 Introduction

With a goal of minimizing the number of children who "age out" of the foster care system, state governments, county agencies, and non-profit organizations have devoted significant resources to providing children in foster care with permanent placements in a timely manner. Federal legislation such as the Fostering Connections to Success and Increasing Adoptions Act of 2008 has mandated and reinforced these efforts. The state of Pennsylvania funds the Pennsylvania Adoption Exchange (PAE), which was established in 1979 to support county and nonprofit agencies as they attempt to find adoptive families for children who are difficult to place due to attributes such as age or special needs. In addition to listing children on a website and hosting in-person matching events, PAE maintains detailed data on children and the preferences of families. PAE is mandated by the state to recommend matches between families and children. We collaborated with managers at the Pennsylvania Statewide Adoption and Permanency Network (SWAN), a non-profit organization responsible for administering PAE on behalf of Pennsylvania, to redesign the match recommendation process.

PAE's match recommendation function has two main goals. First, it helps overcome geographical and institutional barriers in the adoption search process, given Pennsylvania's 67 counties that are supported by 82 non-profit organizations. Second, the match recommendation system helps social workers search through extensive data on child characteristics and family preferences. Furthermore, PAE managers believe that case workers sometimes have excessively high expectations — i.e., they are waiting for the "perfect" family — so the process has an additional goal of promoting a decision-making structure.

In this chapter, we investigate PAE to help increase the success of match attempts. Our project contributes to an interesting and important public policy area and non-profit application of market design. We focus on simple changes that address PAE's most significant challenges, and identify key elements of the child adoption market. We have worked with PAE to collect additional information from families and children and create a spreadsheet matching tool that is now used by PAE staff to recommend families. While the idea of a computerized matching tool for ranking families is not new, we believe that we are the first to link its effectiveness to an increased rate of successful adoption through a discrete-event simulation of the adoption network. We are also among the first to note that the matching process may distort incentives for families in revealing their preferences truthfully and propose simple remedies.

The remainder of our chapter is organized as follows. First, we provide context for the problem of a match recommendation system as it relates to research on the design of matching markets. We then characterize the challenges that PAE faces and assess the current system through case worker surveys and a regression analysis using child outcome data from 2005 to 2013. We describe how PAE recommends prospective families for children by comparing children's needs with family preferences on a set of approximately 100 attributes using a spreadsheet tool. Based on this understanding of PAE's role in the matching process, we analyze the value of the network and the information available to the network through a discrete-event simulation of PAE's operation. We then discuss our recommendations for the information that PAE collects, the decision rules for match recommendations, and the interaction with system participants. Finally, we conclude by summarizing the implemented improvements and possible future improvements in adoption and other similar domains.

4.2 Design of Adoption Markets

We view PAE as a two-sided matching market and rely on the market design literature to frame our approach to the problem while using operations research and economics techniques to improve the current recommendation practice. Early market design work focused on understanding and improving centralized clearinghouses that operate in the absence of prices and face institutional and ethical constraints. The seminal work of Gale and Shapley (1962) introduced the formal two-sided matching framework. This theory was subsequently advanced and adapted to important applications such as the design of the National Residency Matching Program in the US for matching medical school graduates to internships, residencies, and fellowships at hospitals, as described by Roth and Peranson (1999). This approach was also adapted for other applications such as the assignment of students to public schools (Abdulkadiroğlu and Sönmez 2003) and kidney exchange (Roth et al. 2005).

The inner workings of online dating and job search markets have been documented in more detail relative to adoptions. For example, see Lee (2007) and Hitsch et al. (2010) for empirical assessment of sorting in online dating, and Niederle and Roth (2003) for empirical assessment of the scope of job market. A common theme of these papers is that a centralized platform; i.e., frictionless matching leads to better sorting and higher scope with respect to traditional decentralized search with friction. Although the adoption of children in state custody cannot be characterized as a purely centralized matching system, our current approach has much in common with this line of research in terms of identifying the deficiencies of an existing matching system and recommending improvements both in elicitation of preferences and implementing the match recommendations. Similarly, more recent literature on market design in live-donor organ exchange has focused on introducing new additional tools to increase the number of transplants rather than new matching algorithms, such as the introduction of non-directed donor chains (see Roth et al. (2006), and Rees et al. (2009)).

A recent strand of literature in matching market design has focused on introducing new ideas to improve the functioning of a centralized or decentralized matching market rather than designing new clearinghouses to conduct the matching. For example, Coles et al. (2010) report the introduction of a signaling tool for the academic job market for new economics PhDs. Lee et al. (2011) report an experiment to measure the effects of the use of signaling devices on online dating. Also, a new company, OrganJet, uses private jets to transport patients in order to overcome the inefficiencies of regionally isolated organ donation networks (Ata et al. 2012). A related study by Arikan et al. (2012) shows that broader sharing of the bottom 15% of kidneys (in terms of quality) from deceased donors leads to significantly increased procurement rates for those organs.

4.3 Child Adoptions in Pennsylvania

The statewide network's primary goal is to help find permanent families for children in state custody. Children who fail to be placed upon initial attempts at the county level are provided with extra services at the state level, including match recommendations from PAE. County child welfare appropriations in Pennsylvania exceeded $1.5 billion in fiscal year 2014-2015 (PA Department of Human Services 2015), and are largely used to support approximately 15,000 children in foster care and 2,000 children who are classified as waiting for adoption (Children's Bureau, U.S. Department of Health and Human Services 2014). Between 2007 and 2012, an average of 239 children per year registered to receive match recommendations from PAE. The Office of Children, Youth and Families (OCYF) of Pennsylvania's Department of Human Services mandates that children without an identified adoptive family must be registered with PAE within ninety days of termination of parental rights (TPR).

PAE managers believe that the best scenario for children is to be placed with a permanent family, even for a child who is about to reach the age of majority. For children who are placed, the time during which the child is a legal orphan should be minimized. Furthermore, the suitability of the family for the child is also important. In particular, some families are better prepared than others to handle children with certain needs, whether medical, behavioral, or psychological.

Prior to TPR, a county case worker seeks a suitable family for the child. If there is no clear kinship adoption possibility or potential family within the agency's local network, the CYS worker may contact SWAN and request match suggestions from PAE. The case worker must register the child with PAE in the ninety days after TPR if the worker has not identified a resource for a child. Through working with PAE coordinators, the CYS worker then receives between five and ten families to consider and pursue. After identifying interested families, the worker then arranges an interview and consults with a committee comprised of social workers and other professionals with diverse expertise to choose a family with which to place the child. After a series of successful visits of increasing duration and decreasing supervision, the child is placed with the family. According to Pennsylvania law, an adoption can then be finalized after six months.

Investigations into whether a family and child could be matched require a substantial time investment by case workers representing the children, social workers representing the families, and even the children and families themselves. Overburdened case workers have limited time to review and process lengthy family profiles, and the human element of the matching process often necessitates face-to-face meetings before a decision can be made whether or not to pursue an adoption. Furthermore, PAE stores limited information for use in understanding families' preferences, and discrepancies between families' stated and revealed preferences can cause difficulties for selecting the best families.

4.4 Survey of Case Workers

To characterize the challenges facing PAE, we worked with a PAE manager to perform a survey of case workers for all active children in spring 2011 to gain a broader picture of feelings about PAE across the state. Survey recipients included both county case workers and social workers known as "child-specific recruiters" who work for private non-profit agencies and serve as an additional resource for county case workers in finding families for hard-to-place children. Seventy-seven completed responses were received — forty-three (56%) from public case workers and thirty-four (44%) from private social workers. The survey was divided into three sections: the first set of questions was general, the second set tried to elicit opinions about the current system, and the third aimed to assess attitudes regarding possible changes to PAE.

The case workers were first asked to indicate the helpfulness of various avenues of finding families: the centralized match recommendation system (known as "electronic match recommendations"), inquiries from a public website (www.adoptpakids.org) that PAE uses to promote children seeking adoptive families, decentralized lists of prospective families who registered with non-profit organizations ("affiliates' lists of waiting families") with which the worker had some connection, and in-person matching events sponsored by SWAN or held locally. Least helpful was PAE's centralized matching system: 65% of respondents said that this never or rarely served as the initial source of prospective families for children who are successfully placed. The survey's second part gauged case workers' satisfaction with the centralized system. The design of this part of the survey was assisted by the information gathered through family registration forms (CY 131) and child registration forms (CY 130). The first set of questions in this section addressed the Resource Family Registry, which is the database of prospective families and children seeking adoptions. Furthermore, no respondents strongly agreed and only 32% agreed with the statement that "PAE does a good job of recommending the most suitable families via electronic matches from the Resource Family Registry for each child." The case workers also testified to the difficulty of making placement decisions and case worker bias. More respondents agreed or strongly agreed (53%) than disagreed or strongly disagreed (37%) with the statement that they "know of case workers who struggle to make placement decisions for children because of emotional attachments to those children." Even more respondents agreed (65%) than disagreed (23%) with the statement that they "know of case workers whose personal preferences lead to negative perceptions towards some families."

Survey responses demonstrate both the ineffectiveness of the current matching system and the case workers' mistrust of match recommendations from PAE. However, case workers expressed more positive views about possible helpfulness of the registration data, with over 60% of respondents agreeing or strongly agreeing that family data is "helpful" for screening and child data is "accurate." This indicates the potential value of a statewide matching network, and motivates efforts to improve PAE's ability to help case workers find families for children in state custody.

4.5 Analysis of Child Outcomes

We reviewed registration and outcome information about children served by PAE to better understand adoption trends in Pennsylvania and the varying levels of difficulty in trying to find adoptive placements. PAE managers overcame significant challenges related to the decentralized nature of the adoption process in Pennsylvania to prepare this data set for our use. We are the first to analyze the relationship between child outcomes and child attributes upon registration in Pennsylvania, and the results of the analysis have provided insights about what children might require additional adoption resources and information to share with case workers as part of training on best practices.

Between 2005 and 2013, PAE assisted in the family-finding process for 1,853 children seeking adoptive families. This set of children was a subset of children in state custody with the goal of adoption; only when the matching process encounters difficulties at the county level does the search process shift to the state level. The mean age of a child upon registration with PAE was 9.41 years, and the median age was 9.63 years. Boys comprised 57.8% of all PAE registrants. Of these 1,853 children, outcomes were known for 1,514 children, as 283 were still active upon the creation of the data set in May 2013, and outcomes were missing for 56 children. Otherwise, child outcomes are known and grouped into categories, each of which has a value between 0 and 1 that is given by PAE managers. The most desirable outcome, a finalized adoption, has a value of 1. Emancipation, which can be referred to as "aging out" of the system, is the least desirable outcome and has a value of 0. Other positive outcomes include permanent guardianship arrangement (0.8) and living with a relative (0.7), among other scenarios. An outcome of "hold" with a foster care arrangement is considered a neutral outcome and has value of 0.5. Other negative outcomes include placement in a residential facility (0.2) and a goal change so that the child's case worker is no longer seeking an adoptive placement for the child (0.1).

Of the children for whom outcome data are known, the county case worker succeeded in finding a finalized adoptive placement for 41.4% (627 of 1,514) of children. Another 19.1% of children have lesser positive outcomes with values of 0.7 or 0.8. Negative outcomes with values less than or equal to 0.2 are experienced by 26.2% of children, with 12.4% of children aging out of the system. The remaining portion (13.3%) of children have neutral outcomes. Using the values given by PAE managers, the expected outcome value for a child in the data set is 0.64.

We developed a linear regression model and a logistic regression model to analyze the relationship between children's attributes when they were registered with PAE and their outcomes. Eighty-eight factors were created from registration data, and the outcome was used as the dependent variable. Specifically, the outcome value was used for the linear model and a binary variable with positive outcomes having value one and neutral and negative outcomes having value zero for the logit model. The square of the child's age upon registration was included to represent the increasing importance of age for older children. The date of registration was expressed in the fractional number of years after January 1, 2005. Gender was represented by a binary variable, as were whether the child had a designation of being African-American and/or Hispanic. Another binary variable represented whether the child had more than one race designation. Eighteen binary variables represented the items under the sections labeled "Educational Status" and "Special Needs" on the CY 130 forms, and an additional variable counted the number of binary variables with positive responses. The last category of variables was 61 questions in the CY 130 section labeled "Characteristics of Child," with a binary variable for each question representing a "Yes" answer. A variable that counted the number of positive responses in the section was also used.

Starting with these 88 variables, we performed a backward stepwise procedure using the Akaike information criterion for both the linear and logit models to select which variables to include in the model. Linear and logit regressions were performed on the union of the variables from these two models (Table 4.1). To roughly assess model performance for the linear model, rounding the predicted outcome to the nearest outcome value and classifying it as positive, negative, or neutral, the model classifies 60.0% into the correct one of three categories. Only 8.1% of children had a positive or negative outcome that was errantly predicted as the opposite outcome. All other prediction errors involved the neutral outcome. When binary child outcomes are considered for the logistic models, 75% of outcomes are correctly predicted. For children with negative/neutral outcomes, 59% were correctly predicted, and 85% were correctly predicted for children with positive outcomes.

The child's age upon registration plays a significant role in the regression model and confirms that older children become increasingly difficult to place. For the linear model, the predicted likelihood of success decreases by 0.034 per year at age eight and by 0.087 per year at age 16. The accumulative decrease in the expected outcome compared to a newborn child is then 0.054 for an eight year-old child and 0.535 for a 16 year-old child. Although PAE managers anticipated this general trend, they found the quantification of this relationship to be very helpful as they instruct case workers around the state on best practices. Specifically, they can encourage case workers to register children with the PAE as soon as possible in the adoption process and perform a search through PAE in parallel with family reunification or other placement efforts.

Mental retardation, which had a coefficient of -0.109 for the linear model, was the most negative of the significant special needs factors from the "Child's Statuses" section of the CY 130 form for both models. Two special needs factors — having siblings and attending school in a general education setting — had statistically significant positive coefficients for both models. For the "Characteristics of Child" attributes, the result that surprised PAE managers was a linear regression coefficient of -0.118 for children who use foul or bad language, which was also significantly negative for the logit model. PAE managers found this information valuable to share with case workers as part of training on how to identify challenges to a successful placement. Other factors that were significantly negative with 95% confidence for both models were a difficulty accepting and obeying rules, a desired contact with siblings, contact with birth parents, and contact with the former foster family. The child's gender was revealed not to have a significant effect in either model, but outcome

Table 4.1: We choose 28 factors from the available 88 factors to model child outcomes using ordinary least squares and logistic regression methods. Age upon registration, which is a negative factor for children six years of age and older, was the most important factor for predicting outcomes.

	Dependent variable:							
	Outcome Value			Outcome (Binary)				
	Ordinary Least Squares			*Logistic*			Freq.	Import.
Constant	0.794	***	(0.046)	1.516	***	(0.372)		
Age upon Registration (years)	0.020	**	(0.009)	0.102		(0.075)		High
(Age upon Registration)^2	-0.003	***	(0.0005)	-0.017	***	(0.004)		High
Registration Year (after 2005)	-0.009	**	(0.004)	-0.059	*	(0.031)		
Male	0.019		(0.017)	0.100		(0.128)	57.1%	High
African-American	-0.034	**	(0.017)	-0.198		(0.132)	42.5%	High
Hispanic	-0.051	**	(0.024)	-0.303	*	(0.179)	14.1%	High
SPECIAL NEEDS								
Mental Retardation Diagnosis	-0.109	***	(0.031)	-0.562	**	(0.230)	9.0%	High
Multiple Placement History	-0.035	*	(0.018)	-0.189		(0.137)	45.6%	Medium
Drug Exposed Infant	-0.020		(0.026)	-0.100		(0.202)	11.6%	Medium
Emotional Disability	-0.019		(0.022)	-0.071		(0.162)	20.2%	Medium
General Education	0.064	***	(0.019)	0.353	**	(0.146)	37.1%	
Siblings	0.085	***	(0.019)	0.465	***	(0.143)	47.3%	High
CHILD CHARACTERISTICS								
Blind	-0.164	*	(0.085)	-0.899		(0.611)	1.0%	Medium
Uses Foul or Bad Language	-0.118	***	(0.027)	-0.613	***	(0.194)	15.0%	Medium
History of Running Away	-0.086	**	(0.043)	-0.443		(0.321)	4.2%	High
Desires Contact with Siblings	-0.079	***	(0.020)	-0.443	***	(0.152)	59.4%	Low
In Contact with Former Foster Family	-0.064	***	(0.022)	-0.353	**	(0.162)	18.8%	Low
Rejects Father Figures	-0.061	**	(0.031)	-0.345		(0.230)	8.5%	Low
Difficulty Accepting and Obeying Rules	-0.061	***	(0.022)	-0.337	**	(0.160)	36.9%	Low
In Contact with Birth Parents	-0.058	***	(0.020)	-0.327	**	(0.153)	26.0%	Low
Num. of Characteristics Present	0.007	***	(0.003)	0.034	*	(0.020)		
Parent(s) with Criminal Record	0.017		(0.018)	0.087		(0.138)	51.6%	Low
Difficulty Relating to Others	0.018		(0.022)	0.101		(0.168)	31.0%	Low
Speech Problems	0.024		(0.024)	0.176		(0.191)	18.4%	Low
Previous Adoption or Disruption	0.038	*	(0.021)	0.220		(0.155)	24.1%	Low
Strong Ties to Foster Family	0.041	**	(0.018)	0.226	*	(0.134)	54.2%	Low
Vision Problems	0.042	*	(0.023)	0.224		(0.175)	17.1%	Low
High Achiever	0.054	**	(0.025)	0.283		(0.190)	13.1%	Low
Observations		1,514			1,514			
R^2		0.345						
Adjusted R^2		0.333						
Akaike Inf. Crit.			1,697					

Note: The values in parentheses indicate the standard deviation. Italicized variable names refer to "Characteristics of Child" questions on CY 130 form. Also, *p<0.1; **p<0.05; ***p<0.01. "Freq." refers to an attribute's prevalence among the observation. "Import." refers to the existing default weight given to the attribute before our analysis.

value decreases of 0.034 and 0.051 (for the linear model) were expected for children with African-American and Hispanic designations, respectively, although the logit model did not determine designation as African-American to have a significant effect.

To understand how the regression results compare to managerial intuition about the different factors' relative importance, we compare them to PAE managers' existing classification the factors. As part of a previous attempt to create a family ranking tool that had encountered difficulties, managers divided the factors from the CY 130 form based on their perceived importance into groups with 15 factors as "high," 18 factors as "medium," and

41 factors as "low." Of the ten significant factors with the most negative OLS coefficient, managers had assessed three of the factors (Hispanic, mental retardation diagnosis, and history of running away) as of high importance and two factors (blind and uses foul or bad language) as of medium importance. They classified the remaining five factors as of low importance. These five factors, which include two behavioral traits and three related to a child's social connections, merit closer attention and a higher weight in the matching process to help identify families more suited to a child's needs. As a result of this analysis, they ultimately decided to reclassify almost all significant ($p < 0.1$) factors with a negative OLS coefficient as of high importance in the match assessment tool, which we will further discuss later in the chapter. The lone exception to this rule was multiple placement history, which remained at medium priority due to the low magnitude of its coefficient. PAE managers particularly appreciated the suggestion to increase the importance of a factors related to a child's social connections, as they identified those connections as a frequent obstacles to adoption.

PAE managers identified four other characteristics, which represent some of the most severe behavioral needs on the CY 130 form (e.g., abusing animals), as of high importance that did not appear as significant factors. However, increased managerial attention to finding families related to these children's needs might have led to their exclusion from the final model for predicting child outcomes, and we do not make a recommendation about whether to reduce the emphasis on these child characteristics.

4.6 Match Assessment Tool

PAE is only one of several governmental and non-profit institutions that have developed tools to assess a possible match between a child and family. Hanna and McRoy (2011) describe the practice of matching in adoption as a means of finding families that have the right capabilities for handling a child's special needs and identifying gaps in a family's capabilities. They emphasize the need for standardization and data collection, point to match assessment tools as an important part of the family finding process, and review seven different tools used in practice by public and private agencies. Although some tools use more attributes (up to 277) than PAE, the design of PAE's existing matching tool surpasses all seven tools in terms of nuance in how attributes are weighted and its ambitions for helping to find families within a statewide network.

According to its intended design — which we formally express in Algorithm 1 — PAE's match assessment tool computes a family's score between 0% and 100% for a child based on 78 pairs of child attribute values and family preferences from CY 130 and CY 131 registration forms (Table 4.2). PAE managers had assigned a number of possible points to each of these pairs, which is 100 for items of high importance, 10 for items of medium importance, and 1 for items of low importance. When a special need or other attribute is not applicable or of an unknown state for a child, no points are either eligible or awarded to each family for that attribute. We note that PAE also does not give preference to families who say that a special need is undesirable when matching a child who does not have that special need. This practice could create incentives for families to hide special needs that they can accommodate if the families want to be considered for children without those special needs. For attributes that are applicable for a child, a positive family response — a child's

Algorithm 1 Design of existing algorithm for computing scores for each family for a given child.

Inputs:

1) Attribute value c_k for each attribute $k = 1, 2, \ldots, K$ corresponding to Table 4.2 for a single child

2) Preference value f_{jk} for each family $j = 1, 2, \ldots, F$ and each attribute $k = 1, 2, \ldots, K$

3) Weight $w_k \geq 0$ for each attribute $k = 1, 2, \ldots, K$

Output: Set of family matching scores $\{y_1, y_2, \ldots, y_F\}$

for $j = 1$ **to** F **do**
    Points possible $x_j^{TOT} \leftarrow 0$
    Points earned $x_j \leftarrow 0$
    for $k = 1$ **to** K **do**
        if $c_k \notin \{$"not applicable", "unknown", "no"$\}$ **then**
            $x_j^{TOT} \leftarrow x_j^{TOT} + w_k$
            if $f_{jk} =$ "will consider" **then**
                $x_j \leftarrow x_j + w_k/2$
            else
                if f_{jk} is compatible with c_k (see Table 4.2) **then**
                    $x_j \leftarrow x_j + w_k$.
                end if
            end if
        end if
    end for
    $y_j \leftarrow x_j / x_j^{TOT}$
end for
return $\{y_1, y_2, \ldots, y_F\}$

Table 4.2: PAE managers use data on 76 attributes to recommend families for children. Weights displayed represent weights used in the existing algorithm.

	Number of Attributes (Scoring Weight)			
Category	High (100)	Medium (10)	Low (1)	Child Attribute Values (Family Preference Values)
Child Demographics				
Age	1	0	0	Current age (Max/min age)
Race/Ethnicity	6	0	0	Applicable/*not applicable* (Preferred/not preferred)
Gender	1	0	0	Male/female (Male/female/either)
Child Status				
Educational Status	0	1	0	Applicable/*not applicable* (Approved/not approved)
Special Needs	6	7	0	Applicable/*not applicable* (Approved/not approved)
Characteristics of Child				
Health	0	3	7	Yes/*no*/*unknown* (Acceptable/will consider/unacceptable)
Education	0	1	7	Yes/*no*/*unknown* (Acceptable/will consider/unacceptable)
Characteristics and Behaviors	5	5	11	Yes/*no*/*unknown* (Acceptable/will consider/unacceptable)
Connections and History	0	0	14	Yes/*no*/*unknown* (Acceptable/will consider/unacceptable)
Contact with Birth Family	0	0	1	Yes/*no*/*unknown* (Acceptable/will consider/unacceptable)

Note: Items in italics indicate child attribute values for which the attribute does not count as part of the total matching score.

age within the family's range, a matching gender, and answers of preferred, approved, or acceptable — receives all possible points for the item. An answer of "will consider" receives 50% of the possible points. Otherwise, the family's answer receives no points. For a specific child, the family's score is simply the sum of points received for their answers divided by the sum of possible points for the child. Appendix B provides an example match score for a child and family.

In recent years, PAE has struggled to make match recommendations that help case workers to find and assess families. Specifically, coordinating algorithm design with an information technology contractor and managing data collected over time across Pennsylvania's 67 counties proved difficult for PAE, and child case workers received unhelpful or illogical match recommendations as a result of a flawed implementation of Algorithm 1. The automated match recommendations were even abandoned for a period of over two years during which PAE coordinators manually searched through CY 130 and CY 131 forms to provide match suggestions.

Recognizing the shortcomings of the current rules for choosing matches, we worked with the SWAN managers to redesign the matching tool. This resulted in a spreadsheet-based algorithm that use PAE data about families and children to select matches. For ease of implementation, we focused on policies that had the same form as the PAE's match assessment tool. In particular, we considered policies that are based on a point system and ranked families according to some compatibility criteria. Rather than making specific assumptions about the relative importance of each criteria, our method offers the PAE managers the flexibility to select their desired weights and any other geographical constraints, as shown in Figure B.1. To find select prospective families for a child, the spreadsheet tool computes a ranked list of families for a child using CY 130 and CY 131 information stored in tables elsewhere in the matching tool.

The matching algorithm that we implemented differs in two aspects from those that have been applied in prominent centralized two-sided matching applications. First, the algorithms studied in that literature, such as Gale and Shapley's deferred acceptance algorithm, produce a batch of "final" matches that are concurrently implemented. However, our algorithm generates a list of mere recommendations, which may be implemented in conjunction with the judgment of the professionals. In this sense, our approach is closer to that adopted by the literature on semi-decentralized matching platforms such as those for online dating and job matching. The literature on these markets almost exclusively focuses on estimating participants preferences as opposed to increasing match success rate and quality, which we pursue here. The second, more subtle difference is that in centralized matching, participants are required to rank-list all available options, whereas here such ranking information is impossible to directly elicit because of limitations such as market size and informational asymmetry. One main function of the above algorithm can be seen as constructing such preferences from given pieces of information in the CY 130 and CY 131 forms and using them as the basis of recommendations.

4.7 Simulation of the Pennsylvania Adoption Exchange

To examine the impact of a simple matching tool's effective use on the network's overall adoption rate, we represent PAE's matching process as a discrete-event simulation. We show the value of a statewide pool of families compared to a decentralized search process, and analyze how the ability of PAE to predict a match's success increases the number of matches. Specifically, we model how different levels of information about child attributes and family preferences affect the number of matches and the number of attempts before a successful match. We rely on the results of the regression analysis from the previous section to identify the most important child attributes for the simulation and additionally introduce relevant family registration data to model family preferences.

As an alternative to conventional techniques such as clinical trials, which would require many years to evaluate, discrete-event simulation has long been used to estimate the effects of policy changes, especially organ allocation policies (cf. Ata et al. (2012)). Similarly, our discrete-event simulation model of PAE's operation is modular and based on input parameters estimated with real data. Some of the simulation studies on organ allocation use a finite-horizon model, which is necessary since the data is highly time-dependent and reaching steady-state is very unlikely unless there is an alternative therapy for transplant patients. However, we assume stationary parameters in our simulation model, which is

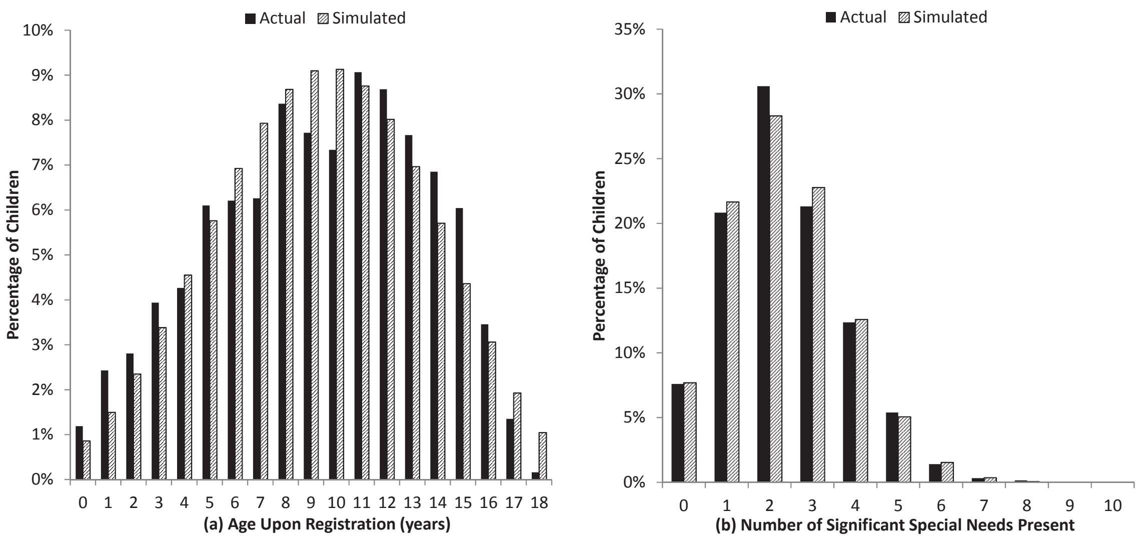


Figure 4.1: Based on age upon registration data for 1,853 children, we simulate the child's age using a beta distribution and use a binomial random variable to simulate the number of significant special needs present for each child.

justified in the context of child adoption as the population characteristics of children and families in the system do not change dramatically over time.

The adoption network is divided into regions that constitute separate adoption networks defined by geographical and/or institutional barriers. Children may only be adopted by families that reside within the same region. Decreasing the number of regions to increase the size of each region provides each separate network with more prospective families and more children to match. We model the state of the PAE before our project as 20 separate regions due to the ineffectiveness of the central matchmaker. In that case, each region may correspond to a large county or a coalition of smaller counties in Pennsylvania. Because county case workers do face geographical limitations in matching, we do not expect a perfect centralized matchmaker to operate as a single region. Instead, managerial insights are primarily motivated by two cases: doubling the region size — i.e., dividing the state into 10 regions instead of 20 regions — and a system with four regions corresponding to PAE coordinators who provide match recommendations to county case workers.

4.7.1 Children

Children are defined by their age, number of special needs, and region in which they reside. A younger child is generally preferred to an older child, and a child with fewer special needs is generally preferable to a child with more special needs. The age attribute corresponds to the child's age upon registration with PAE. Using available data for the 1,853 children who have been registered with PAE, we fit the data using the input analyzer tool of @Risk and compared alternatives using a q-q chart. A beta distribution with shape parameters $\hat{\alpha} = 5.7736$ and $\hat{\beta} = 4.8877$ and scaled to be within the interval $[-5.4648, 22.738]$ was determined to produce the best fit (Figure 4.1a). In the simulation, any age values outside

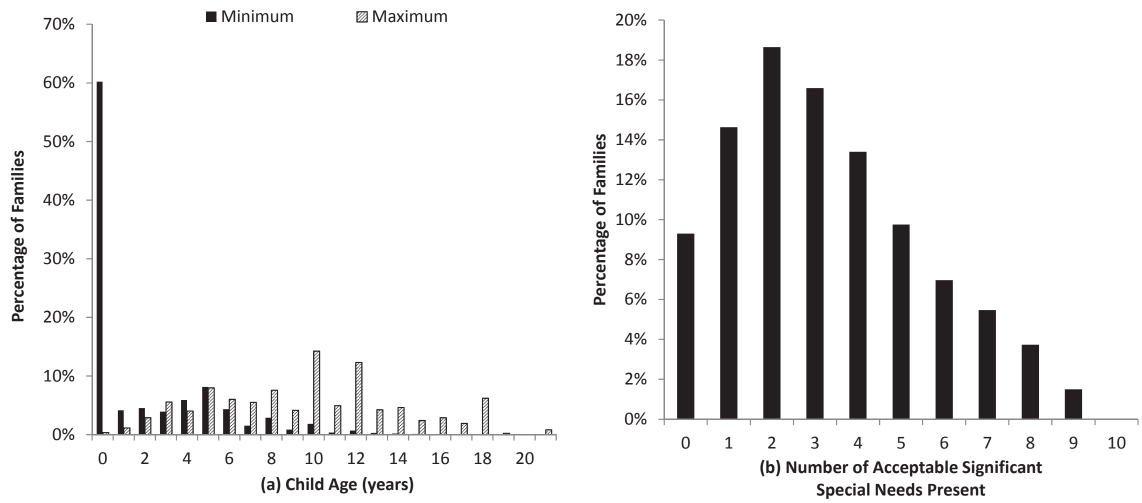


Figure 4.2: To simulate family preferences, we sample actual preferences from 2,194 registered families.

the interval $[0.0, 19.0]$ were discarded and resampled.

The special needs attribute corresponds to a count of the presenting attributes out of the ten child attributes that had a significant negative coefficient with a value less than -0.05 in the OLS regression analysis. This cutoff is arbitrary and used only in the simulation analysis to designate attributes of high importance. As with age, we fit the data using @Risk's input analyzer, and modeled the number of special needs present as a binomial random variable with parameters $n = 14$ and $p = 0.16741$ (Figure 4.1b). Any special needs values greater than 10 were discarded and resampled.

The registration age and number of significant negative special needs are positively correlated with a correlation coefficient of 0.230. Therefore, we used a Normal-To-Anything (NORTA) process with two base vectors that have a correlation coefficient of 0.239, which was obtained via a simulation approach. The standard multivariate normal vectors that follow a NORTA distribution are transformed to the age and special needs distributions using the method as described by Biller and Ghosh (2006).

The value of the child's region attribute is a random variable uniformly distributed over all regions. Children arrive in the system as a Poisson process with a rate of 239 per year, which is the average number of children who were fully registered with PAE annually between 2007 and 2012.

4.7.2 Families

Families are defined by their region of residence and their preferences for an adoptive child's age and maximum number of special needs, as well as the relative weight of the age preference compared to the special needs preference. They arrive to the matching system as a Poisson process with rate 282 per year, which is the mean number of families to fully register with PAE as approved adoption resources each year between 2007 and 2012. Managers estimate that there are in the order of 1,000 prospective families available at any point in time, which implies that the expected time in system is 3.55 years by Little's Law. Thus, we model the family's time in the system as an exponential random variable with a mean of 3.55 years, as the distribution of families' time in system is not tracked by PAE. We note that higher number of families compared to the number of children in the system creates a disparity in the distribution of children available to adopt and the distribution of family preferences, which is reflected in the PAE system through children who "age out" of the system without an adoptive placement.

We model the families' behavior as myopic, accepting the first child that they are offered for which their utility of a match with the child is positive. The model behind the family acceptance decision is discussed in Appendix C. Once a family accepts a child, the family departs from the system. The values for a child's minimum age, maximum age, and number of acceptable special needs are sampled together from the data on 2,194 families (Figure 4.2). As with children, families are uniformly distributed over the regions.

4.7.3 Matching Process

We represent the matching process as a series of events that take place upon the arrival of a child in the system, which is viewed as the driver of the matching process by PAE managers. Matches are offered sequentially to families within the child's region. In each system, families are sorted according to criteria that correspond to PAE's operation with different levels of information. The highest-ranking family is selected and offered the child as a match. If a family accepts a match (i.e., its utility for the match is positive), both the family and child depart from the system. Whether a family's utility is positive depends on the child's characteristics, the family's preferences, and a random term to represent the uncertainty of attraction. Because data is not available to estimate the randomness of this process, we tested two values of the variability of the error term that we label as low attraction variability and high attraction variability. If the family rejects the match, the family remains in the system and another match — up to ten total match attempts — is attempted for the child. For simplicity, we model the matching process as an instantaneous event, although in practice time elapses between sequential matching attempts. If no match is found for the child, the child departs from the system. A flowchart representing this process is provided as Figure C.1 in Appendix C.

We present three methods for ranking families to investigate the value of information. The three methods and their interpretations are as follows:

1. **Critical Attribute (CA)** represents a system in which case workers can search for families based on either the age or special needs attribute due to constraints on their search time and effort. We represent the matching process before our collaboration with PAE as following the CA policy.

2. **Unknown Weight (UW)** represents a simple version of a centralized matching system that is limited in its ability to properly incorporate family preference information. Age and special needs attributes are given equal weight in this model.

3. **Full Information (FI)** represents an improved version of PAE's centralized matching system with families sorted based on a known age preference, special needs preference, and preference weighting term.

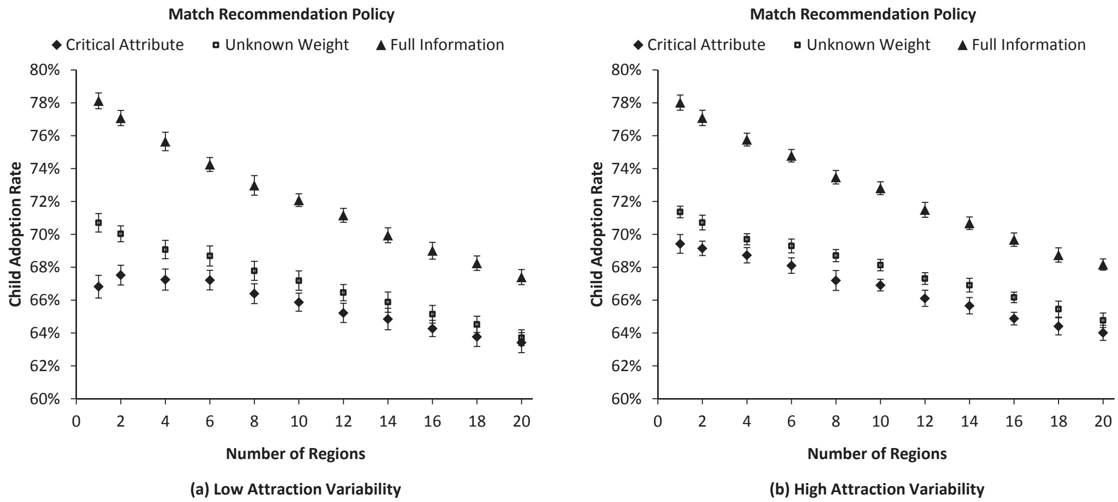


Figure 4.3: The child adoption rate increases with the quality of information available for matching and decreases with the number of regions (i.e., the segmentation of the network). Bands represent 95% confidence intervals.

4.7.4 Simulation Results

We first compare the simulated mean percentage of children matched over the five-year horizon based on the attraction variability and number of regions for CA, UW, and FI decision rules. The adoption rate increases with the amount of information about the families' preferences utilized in the match recommendation process (Figure 4.3); i.e., CA exhibits a lower adoption rate than UW, which in turn has a rate lower than that of FI. The UW policy only slightly improves upon the CA policy with a maximum increase of 3.9 percentage points. However, the FI policy improves on the CA policy by between 4.0 and 11.3 percentage points in the child adoption rate. Whether the attraction variability is low or high appears to have minimal impact on the performance of the policies in terms of the overall adoption rate.

The mean adoption rate always either increases as the number of regions decreases or shows a statistically insignificant decrease. A completely centralized system (i.e., one region) results in an adoption rate that is between 3.4 and 10.7 percentage points higher — depending on the attraction variability and the recommendation rule — than the completely decentralized case with 20 regions. This validates the role of a statewide network. With a larger pool of families, it is more likely for a family to exist that seeks the type of child being matched or can accommodate the child's special needs.

In addition to an increase in the adoption rate for the UW and FI policies compared to the CA policy, the better use of information also corresponds to a decrease in the mean number of match attempts until a successful adoption for children who are successfully adopted (Figure 4.4). We study this metric as a proxy for two important secondary measures of success for the adoption network: the workload for case workers and time in system for the child. Fewer attempts until success means less work for overburdened case workers and a shorter time in state custody for a child. Depending on the number of regions, the UW

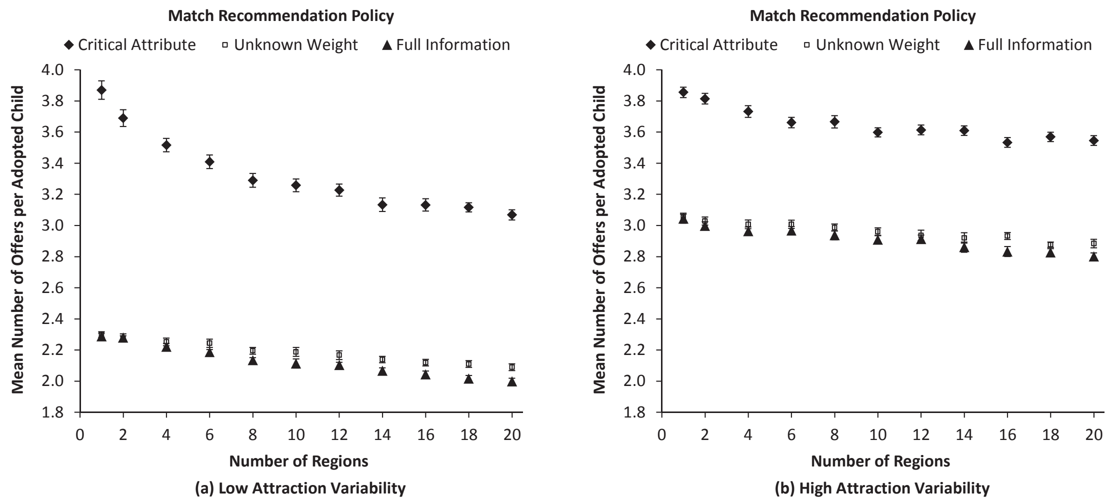


Figure 4.4: Improving the information available for matching reduces the average number of attempts before a successful adoption. Bands represent 95% confidence intervals.

and FI policies cause the mean number of match attempts until success to decrease between 32% and 41% when attraction variability is low and between 17% and 21% when attraction variability is high.

We further investigate the effect of the attraction variability, which represents the unpredictability of attraction between an individual family and child. When the attraction variability is high, match success is inherently more difficult to predict, which results in an increase in the mean number of attempts per successful adoption of up to 0.81 (Figure 4.4). Comparing the change in the mean adoption rate as the number of regions decreases, the difference between the low and high attraction variability cases (in relation to the calibration point) is almost always less than 1%. This indicates that the underlying match unpredictability has relatively little impact on the mean adoption rate compared to the matching rule and number of regions. The only exceptions are for the CA policy with 1 or 2 regions when high attraction variability results in an adoption rate that is 1.0-2.0 percentage points higher than the adoption rate when the attraction variability is low. In these cases, the lower match success predictive power of the CA policy for ranking families, the high attraction variability, and the small number of regions means that it is relatively more likely for families who are offered children later in the sequence of ten offers to be stronger candidates.

These results justify the value of a statewide adoption network and show that the quality of information about family preferences is critical to its success. Furthermore, we have shown that better information improves secondary metrics of system performance that can be interpreted as reducing case workers' workloads and time in system for children. If the adoption network can even just double the size of its regions so that the number of regions is 10 instead of 20 while improving how it elicits family preferences for matching (i.e., follow the FI policy), the number of successfully adopted children can increase by approximately 21 children per year. Additional results and managerial insights related to improvements in match quality are discussed in Appendix C.

4.8 Process Improvements and Results

To achieve the possible improvement in the adoption rate demonstrated in the simulation, we worked with PAE to improve the information it collects and its family ranking tool for matching. In this section, we discuss these changes, as well as the incentives that affect how participants reveal their preferences. A spreadsheet matching tool in use by PAE coordinators provides the most tangible evidence of improvements from our collaboration. While using a computerized tool for matching is not unique to our project, we are the first to connect the quality of preference information to the overall network adoption rate and show how underlying incentives for how families reveal their preferences can diminish the usefulness of recommendations. We also suggest improvements to the family ranking algorithm that are novel to the practice of matching in child adoptions.

4.8.1 Registration Information

Through our interviews with child case workers and discussion of the simulation results, PAE managers came to recognize new potential for gathering information during the registration process as a driver of the overall network adoption rate. Specifically, our research collaboration has led to the collection of additional information to use in making match recommendations. Although PAE managers believed revising the CY 130 and CY 131 forms to be an arduous process, especially since the forms had been recently revised, they are beginning to collect child and family information through an online survey format with a new set of questions. Their intuition about data that would be most valuable for predicting matches informed their selection of these new questions. In particular, these new questions focus on the child's positive attributes, such as interests or hobbies, that might predict attraction between children and families. Other questions focus on family attributes that could be compared to child or child case worker preferences for families without certain pets or other children in certain age ranges. The questions have received approval from the state for use by PAE and are in the process of being implemented.

Furthermore, as a result of our project, PAE managers have begun tracking the results of match recommendations for future analysis. While the set of information used for match recommendations is currently based on managerial intuition, we have encouraged PAE to maintain data about match attempts and their results in order to scientifically evaluate the power of different questions to predict matches. Econometric analysis of child attributes, family preferences, and results of match attempts would allow PAE managers to better estimate the probability of success of a child-family match and assess which questions are more or less important in predicting a match.

4.8.2 Spreadsheet Matching Tool

After several design and feedback iterations, PAE coordinators have begun to use our matching tool prototype to suggest families to county-level case workers. The matching tool prototype has also allowed them to gain insights into the matching rules and to begin to think about what matching rules help produce the best matches for children. We have supported the addition of features, such as geographical preferences for families, to improve the tool's value to PAE managers.

Compared to other matching tools discussed in Hanna and McRoy (2011), we use the same underlying framework of linearly weighted questions to score a family's suitability for a child, and add three simple innovations. First, the user can directly specify the weights for each attribute to help determine which attributes are most important for selecting a family for a child. PAE managers and coordinators observed that children are "labelled" by certain special needs (e.g., fire starter, animal abuser) when the underlying behavior that prompted the label was viewed as innocuous. They felt having the ability to adjust matching tools weights based on knowledge of the severity of a child's special needs could produce better matches. This feature also allowed the easy change of default weights for factors identified as important in the regression analysis. Second, the user can state geographical preferences for the family's county of residence, which can be important in assessing the feasibility of a match if continuing community or familial relationships is important for children. Finally, as noted in Hanna and McRoy (2011), social workers who assist families can use the tool to identify shortcomings in the family's capabilities for a child and prepare appropriate support mechanisms. To this end, we have added score summaries by category so that users can more easily identify a family's strengths and challenges, as well as output reports for use by the matching committee that show how the child and family compare for each attribute.

4.8.3 Information Incentives

We also examined how families and child case workers interact with PAE to understand intentional or unintentional behavior patterns that may reduce the matching process's effectiveness. Through conversations with PAE managers and preliminary runs of the matching system, we noticed that the system is vulnerable to strategic manipulation by families in how they complete their CY 131 forms. Because rejecting a child is very easy for families — a telephone hotline is available to review details of and accept/reject an available child for which the family is recommended — families have an incentive to overstate their willingness to accept children with special needs. This allows a family to gain additional information and be considered for more children. Furthermore, this behavior does not necessarily result from conscious manipulation of the PAE system; different families may be inherently more or less strict in how they interpret the difference between an "acceptable" and "will consider" response to a specific special need. The current system gives families the incentive to err on the side of choosing "accept," which makes differentiating between families more difficult for PAE.

We recommended a process change and an algorithmic feature to overcome the challenge of families' overstating their tolerance for children with special needs. First, we recommended that matching occur in small batches so that PAE coordinators — who use the matching tool — can observe if families are chosen too frequently and further investigate the appropriateness of those families as recommended matches. PAE managers initially welcomed this suggestion, and decided that monthly matching meetings would work well with the adoption framework. They also augmented it based on their experiences with a rule that PAE coordinators should wait 30 days between successive recommendations of the same family. Second, the family-child match score was adjusted for three criteria — race, age, and gender — to reward families whose preferences more closely fit the child's attributes. For example, a family who indicates a preference of male or female receives a higher score on the gender attribute than a family who indicates a preference of "either" if the child is of the preferred gender.

Using a matchmaking experiment, we show that the rewarding of narrow preferences more effectively spreads the recommendations over the pool of families. For each active child, we calculated the top five matches (plus ties) from the list of active families using a scoresheet with and without rewards for narrow preferences over age, race, and gender. Without rewards for narrow preferences, we noticed that only 7.7% of families received at least one match, although we should note that this number would increase in practice when geographical preferences are applied. With rewards for narrow preferences, the number of families who received at least one match increased by 41%. We proposed further rewarding narrow preferences based on the number of child special needs that the family states are acceptable, based on the acceptable number of all special needs or the ten special needs identified as significant in the regression analysis.

Another difficulty that PAE managers emphasized is placement decision making by child case workers. SWAN managers and even case workers themselves indicate that some case workers also struggle with emotional attachment to children to an extent that dispassionately making a placement decision according to the child's best interests can be very challenging. In other words, some case workers' emotional attachment can cause them to hold out for the "perfect" family, when another family likely to be suitable for the child is available. PAE managers have found the spreadsheet matching tool to be valuable as a mechanism to enforce the conceptualization of trade-offs. In conversations between case workers and PAE coordinators who use the matching tool, observing that no family is likely to be a 100% match can lead to discussions about the strengths and weaknesses of a family-child match.

4.9 Conclusions

In the collaboration described in this chapter, we helped PAE improve the processes for recommending prospective families for children in state custody. We believe that these changes increase PAE's value to case workers in their efforts to find find families, and will increase the percentage of children who find permanent placements. Furthermore, PAE has begun collecting additional data about the matching process to scientifically compare families' stated preferences to their actual decisions. This will enable future work that would analyze the matching weights and relative value of the registration questions. The challenge of making match recommendations in a two-sided matching market certainly extends beyond the adoption of children in state custody, and insights from this chapter may apply to matching for foster care placements.

Chapter 5

Conclusions

The work in this thesis demonstrates the potential of operations management to contribute to the design of important and innovative services, both in the commercial and non-profit sectors. Insights from this work pertain to both tactical and strategic management questions, and are motivated by and have influenced practice.

Approaches grounded in stochastic inventory theory can help increase the profitability and service levels of online rental businesses. Work in this thesis has focused on decision variables such as the pre-season inventory ordering decision, the recirculation rules for assigning rental units to customer demands, and whether or not to accept reservation requests. It has incorporated novel model elements of usage-based loss of rental units and service requests by customers in advance of the service start time.

Future work on rental systems could look at additional management decisions into account or broaden the scope of the model. In-season reordering decisions provide a natural extension of the work presented in Chapter 2, as managers might have a chance to replenish inventory in response to lost rental units or inferences about the demand rate. Stock-out based substitution in a multiproduct model provides a potentially valuable extension for rental businesses. Jointly considering the inventory levels in an assortment planning problem then provides additional challenges.

Based on a variety of techniques, the insights from this thesis can also help an adoption network increase the rate of successfully adoption children, as well as improve secondary quality metrics such as the time until adoption and the quality of the adoptive placement. Through a discrete-event simulation motivated by the Pennsylvania Statewide Adoption Network, we have shown that the successful outcome rate for children depends on both the predictive value of the information collected and the network size. This justifies the network's value and motivates efforts to expand and improve the family and child registration data. We also used an analysis of child outcomes and family preference incentives to create a revised match recommendation algorithm that was implemented as a spreadsheet tool.

The practice of child adoption could greatly benefit from data collection and analysis related to family and child preferences. By observing successful and unsuccessful match recommendations, the weights of the match recommendation could be improved to better predict successful matches through knowledge of which criteria should be emphasized and which should be ignored. In particular, estimates of participants' sensitivity to geographical distances could help to inform the match recommendation algorithm and improve the understanding of the network's value.

The challenge of making match recommendations in a two-sided matching market certainly extends beyond the Pennsylvania Adoption Exchange. Many other states, multi-state partnerships, and non-profit organizations operate similar exchanges in which recommendations can play an important role. Furthermore, the procedure for selecting foster families as children enter the foster care system also faces similar difficulties for managing and evaluating prospective families but in a more compact time frame. Foster care system managers additionally must consider the engagement of families over time and weigh trade-offs between family skill-building, family engagement, and the quality of care when placing the child.

An important application that combines both capacity planning and non-profit operations management is nurse aide staffing in long-term care facilities. In recent efforts to improve care provided in nursing homes, the goal of "consistent assignment" — i.e., having the same nurse aides care for the same residents over time — has emerged as an important pillar of person-centered care paradigms. Despite the prominence of consistent assignment as a quality goal for nursing homes, little attention has been given to the underlying operational challenges and their relationship to consistency expressed as a patient-centered metric. In future work, we plan to study how operational staffing decisions affect both staffing costs and consistency of care.

Appendix A

Proofs for Chapter 2

Proof of Lemma 2.1. For notational convenience, we denote the first forward difference $L_n(y+1,\boldsymbol{\xi}) - L_n(y,\boldsymbol{\xi})$ by $\Delta L_n(y)$, indexing this difference to the time period n and the initial inventory of y rental units, with the objective of comparing the inventory system with y rental units to the system with $y+1$ rental units. We also define $\mathbb{1}(\cdot)$ as the indicator function that takes the value of one if the argument is true and zero otherwise. In addition, $\Delta I_n(y)$, $\Delta R_n(y)$, and $\Delta W_n(y)$ indicate the effect of increasing the value of y by one on the available inventory to rent, the number of rentals, and the number of rental units returned, respectively; i.e., $\Delta I_n(y) = I_n(y+1,\boldsymbol{\xi}) - I_n(y,\boldsymbol{\xi})$, $\Delta R_n(y) = R_n(y+1,\boldsymbol{\xi}) - R_n(y,\boldsymbol{\xi})$ and $\Delta W_n(y) = W_n(y+1,\boldsymbol{\xi}) - W_n(y,\boldsymbol{\xi})$.

In this proof, we represent the effect of a unit increase in y on the total number of rentals in terms of the aggregate effect over all N periods; i.e., $\Delta\mathcal{R}(y,\boldsymbol{\xi}) = \sum_{n=1}^{N} \Delta R_n(y) = \sum_{n=1}^{N} \Delta I_n(y)\mathbb{1}(d_n > I_n(y,\boldsymbol{\xi}))$ because

$$\Delta R_n(y) = \begin{cases} 0 & \text{if } d_n \le I_n(y,\boldsymbol{\xi}) \\ \Delta I_n(y,\boldsymbol{\xi}) & \text{otherwise} \end{cases}$$

by the first-forward differences of the state equations in (2.1). Thus, an additional rental unit results in an additional rental in period n only if there is an additional rental unit available in period n (i.e., $\Delta I_n(y,\boldsymbol{\xi})$) and there is still unsatisfied demand (i.e., $d_n > I_n(y,\boldsymbol{\xi})$) in period n. We separate the summation term into two components — one for period n and another for periods $1, 2, \ldots, n-1$. We further substitute $\Delta I_n(y)$ with its counterpart from the state equations in (2.1) and $W_{n+1}(y,\boldsymbol{\xi}) = \sum_{a=A_{min}}^{A_{max}} W_{a,n+1}(y,\boldsymbol{\xi})$ to obtain the following characterization for $\sum_{n=1}^{N} \Delta R_n(y)$:

$$\begin{aligned}
\sum_{n=1}^{N} \Delta R_n(y) &= \Delta I_N(y)\mathbb{1}\left(d_N > I_N(y,\boldsymbol{\xi})\right) + \sum_{n=1}^{N-1} \Delta R_n(y) \\
&= \left(\Delta I_{N-1}(y) + \sum_{a=A_{min}}^{A_{max}} \Delta W_{a,N}(y) - \Delta R_{N-1}(y)\right)\mathbb{1}\left(d_N > I_N(y,\boldsymbol{\xi})\right) + \sum_{n=1}^{N-1} \Delta R_n(y).
\end{aligned}$$

Next, we recursively substitute $\Delta I_n(y,\boldsymbol{\xi}) = \Delta I_{n-1}(y,\boldsymbol{\xi}) - \Delta R_{n-1}(y,\boldsymbol{\xi}) + \Delta W_n(y,\boldsymbol{\xi})$ and rearrange terms using $\Delta R_n(y) - \Delta R_n(y)\mathbb{1}\left(d_N > I_N(y,\boldsymbol{\xi})\right) = \Delta R_n(y)\mathbb{1}\left(d_N \le I_N(y,\boldsymbol{\xi})\right)$ to get

$$\sum_{n=1}^{N} \Delta R_n(y) = \left(1 + \sum_{n=1}^{N} \sum_{a=A_{min}}^{A_{max}} \Delta W_{a,n}(y)\right) \mathbb{1}\left(d_N > I_N(y,\boldsymbol{\xi})\right) + \sum_{n=1}^{N-1} \Delta R_n(y) \mathbb{1}\left(d_N \leq I_N(y,\boldsymbol{\xi})\right).$$

Finally, we continue by separating $\sum_{n=1}^{N-1} \Delta R_n(y)$ into $\Delta R_{N-1}(y) + \sum_{n=1}^{N-2} \Delta R_n(y)$ and repeating this process until all $R_n(y)$ terms have been removed by substitution. The resulting equation is

$$\sum_{n=1}^{N} \Delta R_n(y) = \sum_{n=1}^{N} \left(1 + \sum_{t=1}^{n} \sum_{a=A_{min}}^{A_{max}} \Delta W_{a,n}(y)\right) \mathbb{1}(d_n > I_n(y,\boldsymbol{\xi})) \prod_{v=n+1}^{N} \mathbb{1}(d_v \leq I_v(y,\boldsymbol{\xi})).$$

In the case of experiencing at least one lost sale over the entire horizon, this expression reduces to an equivalence between $\sum_{n=1}^{N} \Delta R_n(y)$ and $1 + \sum_{t=1}^{u} \Delta W_t(y)$ with $u = \max\{n \in \{1, 2, \ldots, N\} : d_n > I_n(y,\boldsymbol{\xi})\}$. This expression further reduces to $1+\sum_{t=1}^{u} \sum_{a=A_{min}}^{A_{max}} \Delta W_{a,u}(y)$ with $u = \max\{n \in \{1, 2, \ldots, N\} : d_n > I_n(y,\boldsymbol{\xi})\}$ when the system experiences at least one lost sale. Thus, that $\mathbb{E}[\sum_{t=1}^{n} \sum_{a=A_{min}}^{A_{max}} W_{a,t}(y,\boldsymbol{\xi})]$ is concave and non-decreasing in y for $n = 1, 2, \ldots, N$ is a sufficient condition for the expected number of rentals $\mathbb{E}[\mathcal{R}(y,\boldsymbol{\xi})]$ to be concave and non-decreasing in y, and for the expected number of lost sales $\mathbb{E}[\mathcal{L}(y,\boldsymbol{\xi})]$ to be convex and non-increasing in y.

Proof of Proposition 2.1. We use induction to show the satisfaction of the sufficiency condition in Lemma 2.1 for the concavity of $\mathbb{E}[\mathcal{R}(y,\boldsymbol{\xi})]$ in the initial inventory of y rental units. First, we note that there are no rental units returned in period 1. Therefore, the sufficiency condition is trivially satisfied for $n = 1$. We next assume that $\mathbb{E}[\sum_{i=1}^{t} R_i(y,\boldsymbol{\xi})]$ is concave in y for $t = 1, 2, \ldots, n-1$ with $n \geq 2$. What is important to recognize here is that the expected number of returns by period n, $\mathbb{E}[\sum_{t=1}^{n} \sum_{a=A_{min}}^{A_{max}} W_{a,t}(y,\boldsymbol{\xi})]$ can be written as $(1-p)\sum_{a=A_{min}}^{A_{max}} h(a) \sum_{t=1}^{n-a} \mathbb{E}[R_t(y,\boldsymbol{\xi})]$. Therefore, $\mathbb{E}[\sum_{t=1}^{n} \sum_{a=A_{min}}^{A_{max}} W_{a,t}(y,\boldsymbol{\xi})]$ is a concave function of y, and it follows from Lemma 2.1 that $\mathbb{E}[\mathcal{R}(y,\boldsymbol{\xi})]$ is concave and non-decreasing in y.

Building on this structural property of the expected number of rentals, we show the concavity of the expected profit function in two steps: (1) The revenue acquired from all rentals is given by $\sum_{a=A_{min}}^{A_{max}} r_a \sum_{n=1}^{N} R_{a,n}(y,\boldsymbol{\xi})$. Because the duration of a rental that begins in period n is independent of the durations of any rentals that begin in periods $1, 2, \ldots, n-1$, it holds that $\mathbb{E}[R_{a,n}(y,\boldsymbol{\xi})] = h(a)\mathbb{E}[R_n(y,\boldsymbol{\xi})]$. Consequently, we obtain the expected rental revenue as $\sum_{a=A_{min}}^{A_{max}} r_a h(a)\mathbb{E}[\mathcal{R}(y,\boldsymbol{\xi})]$. (2) To account for different salvage values of lost rental units, we consider the expectation of the difference $\sum_{n=1}^{N} R_n(y,\boldsymbol{\xi}) - \sum_{n=1}^{N+A_{max}} \sum_{a=A_{min}}^{A_{max}} W_{a,t}(y,\boldsymbol{\xi})$. Because $\sum_{a=A_{min}}^{A_{max}} h(a) = 1$, the expected total number of returns $\mathbb{E}[\sum_{t=1}^{N+A_{max}} \sum_{a=A_{min}}^{A_{max}} W_{a,t}(y,\boldsymbol{\xi})] := (1-p)\sum_{a=A_{min}}^{A_{max}} h(a) \sum_{t=1}^{n-a} \mathbb{E}[R_t(y,\boldsymbol{\xi})]$ can be rewritten as $(1-p)\sum_{t=1}^{N} \mathbb{E}[R_t(y,\boldsymbol{\xi})]$. Consequently, we conclude that the expected profit function $\pi(y) := -k^S y - c\mathbb{E}[\sum_{n=1}^{N} D_n] + (\sum_{a=A_{min}}^{A_{max}} r_a h(a) + c - p(k^L - k^S))\mathbb{E}[\mathcal{R}(y,\boldsymbol{\xi})]$ to be concave in the initial inventory of y rental units for any rental unit recirculation rule.

Proof of Proposition 2.2. We first compare first forward differences of systems with y rental units and $y+1$ rental units to show that $\mathbb{E}\left[\Delta\mathcal{R}^{SP}(y,\boldsymbol{\xi})\right] \geq \mathbb{E}\left[\Delta\mathcal{R}^{SP}(y+1,\boldsymbol{\xi})\right] \geq 0$

for $y \geq 0$. Since all rental units are unconditionally stochastically equivalent in terms of their lifetimes, without loss of generality we focus on a marginal unit of inventory that has the lowest priority for when rental units are assigned to demands. By the definition of the static priority policy, the additional lowest-priority unit will not change the allocation of any other rental units, implying that $\Delta I_n^{SP}(y, \boldsymbol{\xi}(y)) \geq 0$. Thus, the forward difference of the state equations (2.1) for the number of rentals in a period $n+1$ is

$$\Delta R_{n+1}^{SP}(y, \boldsymbol{\xi}(y)) = \begin{cases} 0 & \text{if } d_{n+1} \leq I_{n+1}^{SP}(y, \boldsymbol{\xi}(y)) \\ \Delta I_{n+1}^{SP}(y, \boldsymbol{\xi}(y)) & \text{if } d_{n+1} > I_{n+1}^{SP}(y, \boldsymbol{\xi}(y)). \end{cases}$$

Next, we show that $\sum_{t=1}^{n} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) \geq \sum_{t=1}^{n} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))$ by induction. For period 1, we observe that $\Delta R_1^{SP}(y, \boldsymbol{\xi}(y)) \geq \Delta R_1^{SP}(y+1, \boldsymbol{\xi}(y+1))$ due to the forward difference of the state equations with $I_1^{\gamma}(y, \boldsymbol{\xi}(y)) = y$ and $\Delta I_1^{\gamma}(y, \boldsymbol{\xi}(y)) = 1$.

Now, we assume that $\sum_{t=1}^{n-1} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) \geq \sum_{t=1}^{n-1} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))$ for some period n. We account for the following cases:

1. $\sum_{t=1}^{n-1} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) = \mathbf{l}'$ **and** $\mathbf{l}' > \sum_{t=1}^{n-1} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))$**:** In this case, the additional rental unit is lost before period n for the system with y rental units but is not lost for the system with $y+1$ rental units. Because $\Delta R_n^{SP}(y+1, \boldsymbol{\xi}(y+1)) \leq 1$ by the state equations, $\sum_{t=1}^{n} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) \geq \sum_{t=1}^{n} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))$.

2. $\mathbf{l}' > \sum_{t=1}^{n-1} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y))$ **and** $\sum_{t=1}^{n-1} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) > \sum_{t=1}^{n-1} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))$**:** As in the previous case, $\sum_{t=1}^{n} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) \geq \sum_{t=1}^{n} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))$ because $\Delta R_n^{SP}(y+1, \boldsymbol{\xi}(y+1)) \leq 1$.

3. $\sum_{t=1}^{n-1} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) = \sum_{t=1}^{n-1} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))$**:** If $\sum_{t=1}^{n-1} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) = \mathbf{l}'$ and $\sum_{t=1}^{n-1} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1)) = \mathbf{l}'$, then the additional unit is unavailable for either system, and $\sum_{t=1}^{n} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) = \sum_{t=1}^{n} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1)) = \mathbf{l}'$. Otherwise, it suffices to show that $\Delta I_n^{SP}(y, \boldsymbol{\xi}(y)) \geq \Delta I_n^{SP}(y+1, \boldsymbol{\xi}(y+1))$. Let $i = \sum_{t=1}^{n-1} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y))$. By the inductive hypothesis, the ith rental of the additional unit occurred no later for the system with $y+1$ units than the system with $y+2$ units. Thus, after a rental duration of A_i' periods, the additional unit returns to become available in an earlier period for the system with $y+1$ units than with $y+2$ units, which implies that $\Delta I_n^{SP}(y, \boldsymbol{\xi}(y)) \geq \Delta I_n^{SP}(y+1, \boldsymbol{\xi}(y+1))$.

Having shown that the change in the total number of rentals up to period n from one additional rental unit is non-increasing in y on coupled sample paths (i.e., $\sum_{t=1}^{n} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y)) \geq \sum_{t=1}^{n} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))$), we note that the property that $\mathbb{E}\left[\sum_{t=1}^{n} \Delta R_t^{SP}(y, \boldsymbol{\xi}(y))\right] \geq \mathbb{E}\left[\sum_{t=1}^{n} \Delta R_t^{SP}(y+1, \boldsymbol{\xi}(y+1))\right]$ follows because the lifetime and rental durations of the additional unit for the two systems being compared are independent and identically distributed. Naturally, this property implies that $\mathbb{E}\left[\Delta \mathcal{R}^{SP}(y, \boldsymbol{\xi}(y))\right] \geq \mathbb{E}\left[\Delta \mathcal{R}^{SP}(y+1, \boldsymbol{\xi}(y+1))\right]$.

Finally, we show that $\pi^{SP}(y)$ is concave and non-decreasing in y; i.e., $\Delta \pi^{SP}(y) \geq \Delta \pi^{SP}(y+1)$ for $y \geq 0$. To determine when the concavity of the expected number of rentals in the initial inventory level implies the concavity of the expected profit, we must analyze whether the $(y+1)$st unit for a system with y units and the $(y+2)$nd unit for a system with $y+1$ units are lost. To do so, we compare $\Delta \Pi^{SP}(y, \boldsymbol{\xi}(y))$ to $\Delta \Pi^{SP}(y+1, \boldsymbol{\xi}(y+1))$

with the profit on a sample path written as

$$\Delta\Pi^{SP}(y,\boldsymbol{\xi}(y)) = \sum_{a=A_{min}}^{A_{max}} (r_a + c)\Delta\mathcal{R}_a^{SP}(y,\boldsymbol{\xi}(y)) - k^S - (k^L - k^S)\mathbf{1}\left\{\sum_{t=1}^{n} \Delta\mathcal{R}_t^{SP}(y,\boldsymbol{\xi}(y)) \geq \mathbf{l}'\right\}. \tag{A.1}$$

We immediately observe that $\Delta\Pi^{SP}(y,\boldsymbol{\xi}(y)) \geq \Delta\Pi^{SP}(y+1,\boldsymbol{\xi}(y+1))$ if $\Delta\mathcal{R}^{SP}(y,\boldsymbol{\xi}(y)) = \Delta\mathcal{R}^{SP}(y+1,\boldsymbol{\xi}(y+1))$. If $\Delta\mathcal{R}^{SP}(y,\boldsymbol{\xi}(y)) > \Delta\mathcal{R}^{SP}(y+1,\boldsymbol{\xi}(y+1))$, then the expected effect for the system with y rental units of the $\Delta\mathcal{R}^{SP}(y,\boldsymbol{\xi}(y)) - \Delta\mathcal{R}^{SP}(y+1,\boldsymbol{\xi}(y+1))$ extra rentals of an additional unit must be non-negative. Considering the loss probability for the extra $\Delta\mathcal{R}^{SP}(y,\boldsymbol{\xi}(y)) - \Delta\mathcal{R}^{SP}(y+1,\boldsymbol{\xi}(y+1))$ of the system with y rental units to which the $(y+1)$st unit is being added, it suffices that $\sum_{a=A_{min}}^{A_{max}} r_a h(a) + c \geq (k^L - k^S)\ell_i$, $i \geq N/A_{min}$ for $\Delta\pi^{SP}(y) \geq \Delta\pi^{SP}(y+1)$ to hold, completing the proof.

Proof of Proposition 2.3. We first show that the forward difference of the number of rentals is decreasing in y when the rental system follows the even spread recirculation rule. To do so, we define a restricted allocation of rental units to a reduced demand level $d'_n(y,\boldsymbol{\xi}(y)) = d_n - R^{ES}_{n,y+1}(y+1,\boldsymbol{\xi}(y+1))$, $n = 1,2,\ldots,N$; i.e., any unit demand served by the $(y+1)$st rental unit for the system with $y+1$ units is not allowed to be satisfied by any rental unit when the system has only y units. We denote the number of units rented in period n with this restricted allocation by $R_n^{ESr}(y,\boldsymbol{\xi}(y))$. By definition of the restricted allocation, we immediately recognize the equivalence between $R^{ESr}_{n,m}(y,\boldsymbol{\xi}(y))$ and $R^{ES}_{n,m}(y+1,\boldsymbol{\xi}(y+1))$ for $n = 1,2,\ldots,N$ and $m = 1,2,\ldots,y$. In other words, the satisfaction of the ith demand by the rental unit $m \in \{1,2,\ldots,y\}$ occurs in the same period for the restricted system with y units and the unrestricted system with $y+1$ rental units. Thus, the addition of one rental unit with lifetime $\mathbf{l}'$ and random durations $\{a'_1, a'_2, \ldots\}$ to the system with y rental units has the same effect for the system with $y+1$ rental units; i.e., $\Delta R_n^{ESr}(y,\boldsymbol{\xi}(y)) = \Delta R_n^{ES}(y+1,\boldsymbol{\xi}(y+1))$. Hence, we also have $\sum_{n=1}^{N} \Delta R_n^{ESr}(y,\boldsymbol{\xi}(y)) = \sum_{n=1}^{N} \Delta R_n^{ES}(y+1,\boldsymbol{\xi}(y+1))$.

Next, we show that $\sum_{n=1}^{N} \Delta R_n^{ES}(y,\boldsymbol{\xi}(y)) \geq \sum_{n=1}^{N} \Delta R_n^{ESr}(y,\boldsymbol{\xi}(y))$ by removing the allocation restrictions so that the recirculation rule obeys the even spread policy and that state equation $R_n^{\gamma}(y,\boldsymbol{\xi}) = \min\{d_n, I_n^{\gamma}(y,\boldsymbol{\xi})\}$ in each period n. Specifically, we show that $\sum_{t=1}^{n} \Delta R_t^{ES}(y,\boldsymbol{\xi}(y))$ is non-decreasing in $\{d_1, d_2, \ldots, d_n\}$, which implies that

$$\sum_{t=1}^{n} \Delta R_n^{ES}(y,\boldsymbol{\xi}(y)) \geq \sum_{t=1}^{n} \Delta R_t^{ESr}(y,\boldsymbol{\xi}(y))$$

holds because $d_t \geq d'_t(y,\boldsymbol{\xi}(y))$ for $t = 1,2,\ldots,n$. In period 1, the inductive hypothesis is true by the state equations because $\Delta R_1^{ES}(y,\boldsymbol{\xi}(y)) = \mathbf{1}\{d_1 > y\}$ is non-decreasing in d_1. For some period n with $\sum_{t=1}^{n-1} \Delta R_t^{ES}(y,\boldsymbol{\xi}(y)) \geq \sum_{t=1}^{n-1} \Delta R_t^{ESr}(y,\boldsymbol{\xi}(y))$, we consider the following cases:

1. $\Delta R_n^{ES}(y,\boldsymbol{\xi}(y)) \geq \Delta R_n^{ESr}(y,\boldsymbol{\xi}(y))$. The result follows immediately.
2. $\Delta R_n^{ES}(y,\boldsymbol{\xi}(y)) < \Delta R_n^{ESr}(y,\boldsymbol{\xi}(y))$. We must show that

$$\sum_{t=1}^{n-1} \Delta R_t^{ES}(y,\boldsymbol{\xi}(y)) + \Delta R_n^{ES}(y,\boldsymbol{\xi}(y)) \geq \Delta \sum_{t=1}^{n-1} R_t^{ESr}(y,\boldsymbol{\xi}(y)) + \Delta R_n^{ESr}(y,\boldsymbol{\xi}(y)).$$

To do so, we will demonstrate that $\Delta I_n^{ESr}(y,\boldsymbol{\xi}(y)) - \Delta I_n^{ES}(y,\boldsymbol{\xi}(y)) \leq \sum_{t=1}^{n-1} \Delta R_t^{ES}(y,\boldsymbol{\xi}(y)) - \sum_{t=1}^{n-1} \Delta R_t^{ESr}(y,\boldsymbol{\xi}(y))$. Each unit difference comprising $\Delta I_n^{ESr}(y,\boldsymbol{\xi}(y)) - \Delta I_n^{ES}(y,\boldsymbol{\xi}(y))$ can only occur when $\sum_{t=1}^{n-1} \Delta R_{t,m}^{ES}(y,\boldsymbol{\xi}(y)) > \sum_{t=1}^{n-1} \Delta R_{t,m}^{ESr}(y,\boldsymbol{\xi}(y))$ for some rental unit m. Therefore, $\sum_{t=1}^{n-1} \Delta R_t^{ES}(y,\boldsymbol{\xi}(y)) + \Delta R_n^{ES}(y,\boldsymbol{\xi}(y)) \geq \sum_{t=1}^{n-1} \Delta R_t^{ESr}(y,\boldsymbol{\xi}(y)) + \Delta R_n^{ESr}(y,\boldsymbol{\xi}(y))$.

Thus, the inductive hypothesis holds to show that the value of an additional rental unit cannot decrease with the conversion of the restricted allocation to an unrestricted allocation while maintaining the even spread policy; i.e., $\sum_{n=1}^{N} \Delta R_n^{ES}(y,\boldsymbol{\xi}(y)) \geq \Delta R_n^{ESr}(y,\boldsymbol{\xi}(y))$. Since it was also shown above that $\mathbb{E}[\sum_{n=1}^{N} \Delta R_n^{ESr}(y)] \geq \mathbb{E}[\sum_{n=1}^{N} \Delta R_n^{ES}(y+1)]$, we see that $\mathbb{E}[\mathcal{R}^{ES}(y,\boldsymbol{\xi})]$ is concave and non-decreasing in y for the even spread recirculation rule. Using this same argument, we observe that the expected total number of returns $\mathbb{E}[\sum_{n=A_{min}+1}^{N+A_{max}} W_n^{ES}(y,\boldsymbol{\xi})]$ is also concave and non-decreasing in y.

To prove that $\pi^{ES}(y)$ is concave in y, we rewrite the profit function as

$$\Pi^{\gamma}(y,\boldsymbol{\xi}) = c\sum_{n=1}^{N} d_n + \sum_{a=A_{min}}^{A_{max}} (r_a + c - k^L + k^S)\mathcal{R}_a^{\gamma}(y,\boldsymbol{\xi}) - k^S y + (k^L - k^S) \sum_{n=A_{min}+1}^{N+A_{max}} W^{\gamma}(y,\boldsymbol{\xi}).$$

As the sum of concave functions, the expected profit function is concave in y.

Proof of Proposition 2.4. We first focus on the proof of Step 5 listed in the text leading up to the proposition, and begin by showing that $\mathcal{R}^S(\xi_{(1)},\xi_{(2)}) + \mathcal{R}^S(\xi_{(2)},\xi_{(1)}) \geq \mathcal{R}^V(\xi_{(1)},\xi_{(2)}) + \mathcal{R}^V(\xi_{(2)},\xi_{(1)})$. First, by the increasing failure rate property and the coupling of sample paths, we note that for a given value of $\eta_{(1)}$ or $\eta_{(2)}$, rental unit i will have a remaining lifetime that is at least as long as that of rental unit j.

Next, we show that $\mathcal{R}^S(\xi_{(1)},\xi_{(2)}) \geq \mathcal{R}^V(\xi_{(2)},\xi_{(1)})$ and $\mathcal{R}^S(\xi_{(2)},\xi_{(1)}) \geq \mathcal{R}^V(\xi_{(1)},\xi_{(2)})$. We consider each inequality separately:

- $\mathcal{R}^S(\xi_{(1)},\xi_{(2)}) \geq \mathcal{R}^V(\xi_{(2)},\xi_{(1)})$: In this case, we compare the loss period of rental unit i defined by $\xi_{(1)}$ in the switched allocation to rental unit j defined by $\xi_{(1)}$ in the violating allocation. Due to sample path coupling, unit j has a remaining lifetime in the violating allocation that is less than or equal to that of unit i in the switched allocation. By induction, we can show that $\sum_{t=1}^{n} R_t^V(\xi_{(1)},\xi_{(2)}) \leq \sum_{t=1}^{n} R_t^S(\xi_{(1)},\xi_{(2)})$ for $n = 1, 2, \ldots, N$. We omit this induction argument for its similarity to that of Proposition 2.3.

- $\mathcal{R}^S(\xi_{(2)},\xi_{(1)}) \geq \mathcal{R}^V(\xi_{(1)},\xi_{(2)})$: In this case, we compare the loss period of rental unit i defined by $\xi_{(2)}$ in the switched allocation to rental unit j defined by $\xi_{(2)}$ in the violating allocation. Again by sample path coupling, unit j has a remaining lifetime in the violating allocation that is less than or equal to that of unit i in the switched allocation. As in the previous case, we can show that $\sum_{t=1}^{n} R_t^V(\xi_{(2)},\xi_{(1)}) \leq \sum_{t=1}^{n} R_t^S(\xi_{(2)},\xi_{(1)})$ for $n = 1, 2, \ldots, N$.

Because rental unit lifetimes and durations are independent and identically distributed, the inequality $\mathcal{R}^S(\xi_{(1)},\xi_{(2)}) + \mathcal{R}^S(\xi_{(2)},\xi_{(1)}) \geq \mathcal{R}^V(\xi_{(1)},\xi_{(2)}) + \mathcal{R}^V(\xi_{(2)},\xi_{(1)})$ implies that $\mathbb{E}\left[\mathcal{R}^S(y)\right] \geq \mathbb{E}\left[\mathcal{R}^V(y)\right]$. Thus, under the assumption that $\sum_{a=A_{min}}^{A_{max}} r_a h(a) + c \geq (k^L - k^S)\ell_i$, our coupling argument implies that the profit from the even spread policy is stochastically larger than that of all other count-based recirculation rules.

Proof of Proposition 2.5. The proof proceeds analogously to Proposition 2.3 by analyzing the marginal effect of an additional rental unit in a sample path coupling framework for which $\tilde{s}_i$ is the state of the additional unit after it has been rented i times. We analyze the best-first policy BF but note that the same logic applies for the worst-first policy. We rely on the reasoning of Proposition 2.3 but must confirm the inductive argument for the condition-based model. Specifically, we define a restricted allocation BFr analogous to that of Proposition 2.3 and show that the change in the number of rentals is non-decreasing as the allocation restrictions are removed. Based on the definition of s_{mi}, the period in which the ith allocation of a rental unit m occurs is non-decreasing as the restrictions are relaxed. As each demand restriction is relaxed, $\sum_{n=1}^{N} R_n(y)$ either remains the same or increases by one. Thus, the expected number of rentals can only increase with the conversion of the restricted allocation to an unrestricted allocation while maintaining the best-first policy, and $\sum_{n=1}^{N} \Delta R_n^{BF}(y) \geq \sum_{n=1}^{N} \Delta R_n^{BFr}(y)$. Since it also holds that $\mathbb{E}[\sum_{n=1}^{N} \Delta R_n^{BFr}(y)] \geq \mathbb{E}[\sum_{n=1}^{N} \Delta R_n^{BF}(y+1)]$, $\mathbb{E}[\mathcal{R}^{BF}(y, \boldsymbol{\xi})]$ is concave in y for the best-first (and worst-first) policies. The remainder of the proof follows as in Proposition 2.3 to show that the concavity of the expected number of rentals implies the concavity of the expected profit function.

Proof of Proposition 2.6. The proof proceeds as in Proposition 2.4 with the need to only modify the inverse probability mass function values $\eta_{(1)}$ and $\eta_{(2)}$ for the conditional lifetime distributions of the two rental units. For P that is totally positive of order 2 and given either $\eta_{(1)}$ or $\eta_{(2)}$, rental unit i will always have a longer remaining lifetime (i.e., the number of possible allocations after period $n-1$) than rental unit j. The remainder of the proof follows as in Proposition 2.4. Thus, with the assumption that $\sum_{a=A_{min}}^{A_{max}} r_a h(a) + c \geq (k^L - k^S)P(i, S)$ for $i = 1, \ldots, S-1$, the profit under the best-first recirculation rule is stochastically larger than that of all other condition-based recirculation rules.

Appendix B

Matching Tool Spreadsheet

Match Scoring Tool for Ranking Families

Child ID	**(Fictitious)**			**Points**	
Family ID	**(Fictitious)**			Earned	Possible
				384	425
Score	**90.35%**				
Weight		**Child Info**	**Family Pref**	**Points**	**Pts Possible**
	DEMOGRAPHIC INFORMATION				
100	**Age**	13		100	100
	Low Age		10		
	High Age		14		
100	**Race/Ethnicity**				
	African American	NA	Not Preferred	0	0
	Hispanic	NA	Not Preferred	0	0
	White	Applicable	Preferred	100	100
	American Indian/Alaskan Native	NA	Not Preferred	0	0
	Asian	NA	Not Preferred	0	0
	Native Hawaiian/Other Pacific Islander	NA	Not Preferred	0	0
100	**Child Gender**	Female	Either	100	100
	SPECIAL NEEDS INFORMATION				
10	Drug Exposed Infant	Applicable	Not Approved	0	10
10	Emotional Disability	Applicable	Approved	10	10
100	HIV	NA	Not Approved	0	0
10	MH Diagnosis	NA	Approved	0	0
100	MR Diagnosis	NA	Not Approved	0	0
10	Multiple Placement History	Applicable	Approved	10	10
100	Physical Disability	NA	Approved	0	0
10	Runaway History	NA	Approved	0	0
100	Sexual Abuse History	NA	Approved	0	0
100	Siblings	NA	Approved	0	0
10	Special Education Student	Applicable	Approved	10	10
100	Special Medical Care	NA	Approved	0	0
10	Abuse History	Applicable	Not Approved	0	10
10	Neglect History	Applicable	Approved	10	10

Figure B.1: PAE regional coordinators use a spreadsheet with customizable attribute weights that automatically computes scores for all families for a given child. ("NA" refers to an attribute that is not applicable for a child.)

Weight	CHILD CHARACTERISTICS	Child Info	Family Pref	Points	Pts Possible
	1. Does child have significant health issues?	No	Acceptable		
1	2. Does child have allergies or asthma? (may require treatment)	Yes	Acceptable	1	1
10	3. Is child hyperactive? (may require treatment)	Yes	Acceptable	10	10
1	4. Does child have speech problems? (may require treatment)	No	Acceptable		
1	5. Does child have hearing problems? (may require treatment)	No	Acceptable		
1	6. Is child legally deaf?	No	Will Consider		
1	7. Does child have vision problems? (may require treatment)	No	Will Consider		
10	8. Is child legally blind?	No	Unacceptable		
1	9. Does child have dental problems? (may require treatment)	No	Acceptable		
1	10. Does child have orthopedic problems (special shoes, braces, etc)	No	Acceptable		
10	11. Does child have seizures?	No	Will Consider		
1	13. Is child a high achiever in school?	No	Acceptable		
1	14. Does child achieve at grade level in regular classes?	Yes	Acceptable	1	1
1	15. Does child achieve below grade level in regular classes?	No	Acceptable		
1	16. Is child in special education classes?	No	Acceptable		
1	17. Does child have a learning disability?	No	Acceptable		
1	18. Does child need classes for the emotionally or behaviorally handicapped?	Yes	Acceptable	1	1
1	19. Does child need tutoring in one or more subjects?	No	Acceptable		
10	20. Does child have serious behavior problems in school?	Yes	Will Consider	5	10
1	21. Is child generally quiet and shy?	No	Acceptable		
1	22. Is child generally outgoing and noisy?	Yes	Acceptable	1	1
1	23. Does child have emotional issues that requires therapy?	Yes	Acceptable	1	1
1	24. Does child tend to reject father figures?	No	Will Consider		
1	25. Does child tend to reject mother figures?	No	Will Consider		
1	26. Does child have difficulty relating to others and relating to other children?	Yes	Acceptable	1	1
1	27. Does child frequently wet the bed?	No	Acceptable		
1	28. Does child frequently soil him/herself?	No	Will Consider		
100	29. Does child masturbate frequently or openly?	No	Unacceptable		
1	30. Does child have poor social skills?	Yes	Acceptable	1	1
10	31. Does child have problem with lying?	Yes	Will Consider	5	10
10	32. Does child have problem with stealing?	No	Will Consider		
10	33. Does child frequently start physical fights with other children?	Yes	Will Consider	5	10
100	34. Does child abuse animals?	No	Unacceptable		
10	35. Is child destructive with clothing, toys, etc.?	Yes	Will Consider	5	10
10	36. Does child use foul or bad language?	No	Acceptable		
1	37. Does child have frequent temper tantrums?	Yes	Acceptable	1	1
1	38. Does child have difficulty accepting and obeying rules?	Yes	Acceptable	1	1
100	39. Does child exhibit inappropriate sexual behavior?	No	Will Consider		
100	40. Does child have a history of running away?	No	Will Consider		
100	41. Does child have history of playing with matches, setting fires?	No	Unacceptable		
1	42. Does child have strong ties to birth family?	Yes	Acceptable	1	1
1	43. Does child have strong ties to foster family?	No	Acceptable		
1	44. Is continued contact with siblings desirable?	No	Acceptable		
1	45. Does child have a previous adoption disruption?	No	Acceptable		
1	46. Was child sexually abused?	No	Will Consider		
1	48. Was child exposed to promiscuous sexual behavior?	No	Will Consider		
1	49. Was child conceived by rape?	No	Will Consider		
1	50. Was child conceived as a result of prostitution?	No	Unacceptable		
1	51. Are one or both parents addicted to alcohol?	Yes	Will Consider	0.5	1
1	52. Are one or both parents dependent on substances other than alcohol?	Yes	Acceptable	1	1
1	53. Do one or both parents have a criminal record?	Yes	Acceptable	1	1
1	54. Are one or both parents mentally retarded?	No	Unacceptable		
1	55. Do one or both parents have a mental illness?	Yes	Will Consider	0.5	1
1	56. Does agency lack information about one or both parents?	No	Acceptable		
1	57. Is child in contact with birth parents?	Yes	Acceptable	1	1
	58. Is child in contact with siblings?	No			
	59. Is child in contact with extended birth family?	No			
	60. Is child in contact with former foster family?	No			

Figure B.2: The spreadsheet tool also includes a section for "Child Characteristics" information.

Appendix C

Matching Model and Simulation Details

Children are defined by a type $c = \{a, s, r\}$, which reflects that child's desirability on two attributes — age $a \in [0, 19]$ years and number of significantly negative special needs $s \in \{0, \ldots, 10\}$ — and a residence region $r \in \{1, \ldots, R\}$, where R is the total number of regions.

Families are defined by their type $f = \{a_{MIN}, a_{MAX}, s', r, \alpha\}$, which represents their range of acceptable ages $[a_{MIN}, a_{MAX}]$ with $a_{MIN}, a_{MAX} \in \{0, \ldots, 19\}$ and $a_{MIN} < a_{MAX}$, their tolerance for child special needs $s' \in \{0, \ldots, 10\}$, a weight $\alpha \in [0, 1]$ to express the relative importance of age and special needs, and a region attribute $r \in \{1, \ldots, R\}$. We also define a utility function for the purpose of indicating whether a family will accept an offered child. A child age component and a child special needs component comprise the utility function, and their relative weight is dictated by the weighting term α. With only limited information about family preferences, we use a uniform distribution for the weighting term, as justified for preference modeling with limited information by Kennan (2006).

We define a family's utility for a match with a child of type c as

$$u(c; f) := \alpha \left(u^{AGE}(a, a_{MAX})\right) + (1 - \alpha) \left(u^{SN}(s, s')\right) + \epsilon, \tag{C.1}$$

where ϵ is an error term that represents the randomness of a child's attractiveness to a family. We let ϵ be an independent random variable that follows a normal distribution with mean μ and standard deviation σ. The term σ represents families' variability in their attractiveness for individual children. Without data to connect families' stated preferences to their acceptance decisions, we consider cases of $\sigma = 0.1$ and $\sigma = 0.2$, to which we refer as low attraction variability and high attraction variability, respectively. Given σ, μ then becomes a tuning parameter for the simulation. The value of ϵ is only revealed when a match is attempted between a family and a child.

Lacking data to directly estimate families' preferences, we instead rely on the analysis of factors related to child outcomes from the previous section to create a model for family preferences. Specifically, we use the resulting coefficients from a linear regression model based on the child's managerially weighted outcome as a response variable and factors of age (linear), age (quadratic), and the number of significant negative special needs. The resulting model is

$$Outcome(c) = 0.8356 + 0.0426a - 0.0045a^2 - 0.0476s, \tag{C.2}$$

for which the intercept and all three coefficients are significant at a 99.9% confidence level. We use these coefficients from this model to estimate the age and special needs components of the utility function. Given a child's age a and a family's preferred minimum age a_{MIN} and maximum age a_{MAX}, we define the age component of a family's utility for a child as

$$u^{AGE}(a, a_{MIN}, a_{MAX}) := \begin{cases} 0.0426(a - a_{MAX}) - 0.0045(a^2 - a_{MAX}^2) & \text{if } a \geq a_{MIN} \\ 0 & \text{if } a < a_{MIN} \end{cases},$$

which represents the difference in the effect of age upon outcome between the child's age upon registration and the family's maximum preferred age. For a family that prefers a child between 0 and 12 years old, the age component of the utility function is 0.235 for a child of age 4, 0.190 for age 8, and 0 for age 12. For older children, the value is -0.149 for a child of age 14, -0.334 for age 16, and -0.554 for age 18. We note that, due to the quadratic term, the utility component is not strictly decreasing in age for very young children, but we ignore this effect just for simulation purposes as children younger than 3 only represent about 6% of the population. In general, these children are not difficult to place and are not the focus of PAE. Similarly, the special needs component of the family's utility for a child is defined as

$$u^{SN}(s, s') := -0.0476(s - s'),$$

which represents the difference in the effect of the number of significant negative special needs and the family's number of corresponding acceptable special needs.

With this model, we can more precisely define the three family ranking methods:

1. **Critical Attribute (CA):** If $0.0426a - 0.0045a^2 < -0.0476s$, then families are sorted according to $u^{AGE}(a, a_{MIN}, a_{MAX})$. Otherwise, families are sorted according to $u^{SN}(s, s')$.

2. **Unknown Weight (UW):** Families are sorted according to their nominal utilities, which disregards the error term, for the child with the two attributes equally weighted (i.e., $\alpha = 0.5$) to represent α unknown.

3. **Full Information (FI):** Families' types are known to the matchmaker, which given a child of type c can rank the families according to their nominal utility $\alpha \left(u^{AGE}(a, a_{MAX})\right) + (1 - \alpha)\left(u^{SN}(s, s')\right)$.

The simulation is initialized by starting with a pool of 1,000 randomly generated families. The replication length is five years and is preceded by a one-year warm-up period. The family attraction variability tuning parameter is set to $\mu = -0.1875$ for low attraction variability $\sigma = 0.1$ and $\mu = -0.315$ for high attraction variability $\sigma = 0.2$, which corresponds to a 64% success rate for the Critical Attribute rule and matches the expected quality-adjusted outcome value for children in the PAE system between 2005 and 2013. We calibrated the simulation using the CA decision rule to represent the process by which county case workers

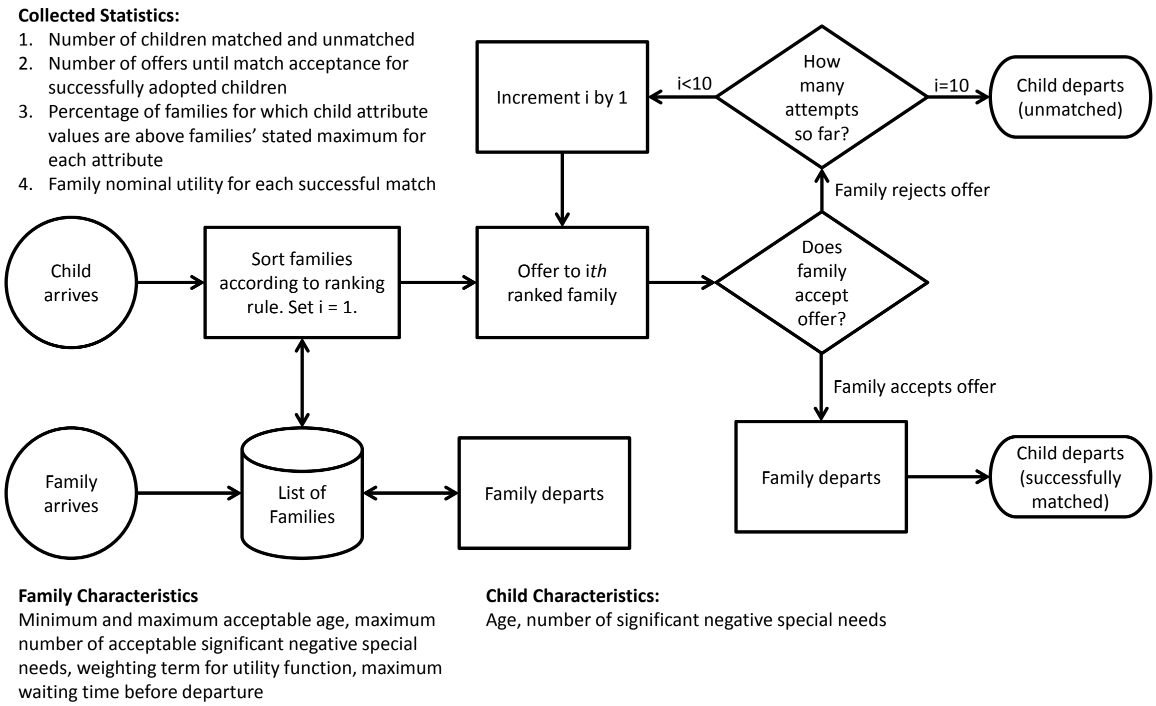


Figure C.1: When a child becomes available, we rank prospective families and sequentially make up to ten match attempts. A child is successfully adopted if at least one family accepts the child. Otherwise, the child is not adopted.

manually searched through families' records focusing on their suitability for a small subset of child attributes, which most accurately describes PAE's functioning before changes were implemented as part of our collaboration. For each scenario — defined by a matching policy and number of regions — we used 25 replications so that we are 95% confident that the resulting mean match rate is within 1% of the true mean match rate. The simulation was implemented in Java and relied upon the Java Simulation Library described in Rossetti (2008) for simulation functions.

Bibliography

Abdulkadiroğlu, A., T. Sönmez. 2003. School choice: A mechanism design approach. *American Economic Review* **93**(3) 729–747.

Adelman, D. 2008. A simple algebraic approximation to the Erlang loss system. *Operations Research Letters* **36**(4) 484 – 491.

Alexandrov, A., M. Lariviere. 2012. Are reservations recommended? *Manufacturing & Service Operations Management* **14**(2) 218–230.

Arikan, M., B. Ata, J. Friedewald, R. Parker. 2012. What drives the geographical differences in deceased donor organ procurement in the United States? Working paper, University of Kansas.

Ata, B., A. Skaro, S. Tayur. 2012. OrganJet: Overcoming geographical disparities in access to deceased donor kidneys in the United States. Working paper, Northwestern University.

Baccara, M., A. Collard-Wexler, L. Felli, L. Yariv. 2014. Child-adoption matching: Preferences for gender and race. *American Economic Journal: Applied Economics* **6**(3) 133–158.

Baron, O., I. Hajizadeh, J. Milner. 2011. Now playing: DVD purchasing for a multilocation rental firm. *Manufacturing and Service Operations Management* **13**(2) 209 – 226.

Bassamboo, Achal, Sunil Kumar, Ramandeep S. Randhawa. 2009. Dynamics of new product introduction in closed rental systems. *Operations Research* **57**(6) 1347–1359. doi: 10.1287/opre.1080.0629. URL `http://or.journal.informs.org/content/57/6/1347`.

Bertsimas, D., R. Shioda. 2003. Restaurant revenue management. *Operations Research* **51**(3) 472–486.

Biller, B., S. Ghosh. 2006. Multivariate input processes. B. L. Nelson, S. G. Henderson, eds., *Handbooks in Operations Research and Management Science*, vol. 13. Elsevier Science, Amsterdam, 123–153.

Binkley, C. 2011. Fashion 101: Rent the Runway targets students. *Wall Street Journal* URL `http://online.wsj.com/article/SB10001424052748703806304576244952860660370.html`.

Brown, M., N. R. Chaganty. 1983. On the first passage time distribution for a class of Markov chains. *The Annals of Probability* **11**(4) 1000 – 1008.

Children's Bureau, U.S. Department of Health and Human Services. 2014. Child welfare outcomes 2009-2012. `http://www.acf.hhs.gov/programs/cb/resource/cwo-09-12`. Online; accessed 16 April 2015.

Coffman Jr, E., P. Jelenkovic, B. Poonen. 1999. Reservation probabilities .

Cohen, M. A., W. P. Pierskalla, S. Nahmias. 1980. A dynamic inventory system with recycling. *Naval Research Logistics Quarterly* **27**(2) 289 – 296.

Coles, P., J. Cawley, P. Levine, M. Niederle, A. Roth, J. Siegfried. 2010. The job market for new economists: A market design perspective. *The Journal of Economic Perspectives* **24**(4) 187–206.

Duenyas, I., W. Hopp. 1995. Quoting customer lead times. *Management Science* **41**(1) 43–57.

Dworsky, A., M. Courtney. 2010. The risk of teenage pregnancy among transitioning foster youth: Implications for extending state care beyond age 18. *Children and Youth Services Review* **32**(10) 1351–1356.

Eisenmann, T. R., L. Winig. 2012. *Rent the Runway*. Harvard Business Publishing, Boston.

Emstad, P., B. Feng. 1990. Traffic models for reservation systems. *Proc. 13th Int. Teletraffic Congress*. 647–652.

Erlang, A. 1917. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Elektrotkeknikeren* **13** 5–13.

Erlanger, S., M. De La Baume. 2009. French ideal of bicycle-sharing meets reality. *The New York Times* URL `http://www.nytimes.com/2009/10/31/world/europe/31bikes.html`.

Gale, D., L. Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* **69**(1) 9–15.

Gans, N., S. Savin. 2007. Pricing and capacity rationing for rentals with uncertain durations. *Management Science* **53**(3) 390 – 407.

Gerchak, Y., D. Gupta, M. Henig. 1996. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science* **42**(3) 321–334.

Green, L. V., P. J. Kolesar, J. Soares. 2001. Improving the SIPP approach for staffing service systems that have cyclic demands. *Operations Research* **49**(4) 549 – 564.

Greenberg, A., R. Srikant, W. Whitt. 1999. Resource sharing for book-ahead and instantaneous-request calls. *IEEE/ACM Transactions on Networking* **7**(1) 10–22.

Hanna, M. D., R. G. McRoy. 2011. Innovative practice approaches to matching in adoption. *Journal of Public Child Welfare* **5**(1) 45–66.

Harel, A. 1988. Sharp bounds and simple approximations for the erlang delay and loss formulas. *Management Science* **34**(8) 959 – 972.

Hariharan, R., P. Zipkin. 1995. Customer-order information, leadtimes, and inventories. *Management Science* **41**(10) 1599–1607.

Hitsch, G. J., A. Hortaçsu, D. Ariely. 2010. What makes you click?—mate preferences in online dating. *Quantitative Marketing and Economics* **8**(4) 393–427.

Howard, J., S. Brazin. 2011. Never too old: Achieving permanency and sustaining connections for older youth in foster care. Tech. rep., Evan B. Donaldson Adoption Institute, New York, NY.

Huang, C. C., S. L. Brumelle, K. Sawaki, I. Vertinsky. 1977. Optimal control for multi-servers queueing systems under periodic review. *Naval Research Logistics Quarterly* **24**(1) 127 – 135.

Jain, A., K. Moinzadeh, A. Dumrongsiri. 2015. Priority allocation in a rental model with decreasing demand. *Manufacturing & Service Operations Management* .

Jung, M., E.S. Lee. 1989. Numerical optimization of a queueing system by dynamic programming. *Journal of Mathematical Analysis and Applications* **141**(1) 84 – 93.

Kaheel, A., H. Alnuweiri, F. Gebali. 2006. A new analytical model for computing blocking probability in optical burst switching networks. *Selected Areas in Communications, IEEE Journal on* **24**(12) 120–128.

Kapuscinski, R., S. Tayur. 2007. Reliable due-date setting in a capacitated mto system with two customer classes. *Operations research* **55**(1) 56–74.

Kennan, J. 2006. A note on discrete approximations of continuous distributions. Unpublished manuscript.

Kleywegt, A. J., A. Shapiro, T. Homem-de Mello. 2002. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* **12**(2) 479 – 502.

Landes, E, R Posner. 1978. The economics of the baby shortage. *Journal of Legal Studies* **7** 323–348.

Lee, S. 2007. Preferences and choice constraints in marital sorting: Evidence from korea. Unpublished manuscript.

Lee, S., M. Niederle, H. Kim, W. Kim. 2011. Propose with a rose? signaling in internet dating markets. Tech. rep., National Bureau of Economic Research.

Levi, R., G. Janakiraman, M. Nagarajan. 2008. A 2-approximation algorithm for stochastic inventory control models with lost sales. *Mathematics of Operations Research* **33**(2) 351–374.

Levi, R., C. Shi. 2011. Revenue management of reusable resources with advanced reservations. *Working Paper* .

Lindvall, T. 1992. *Lectures on the Coupling Method*. Wiley, New York.

Lu, Y., A. Radovanović. 2007. Asymptotic blocking probabilities in loss networks with subexponential demands. *Journal of Applied Probability* **44**(4) 1088–1102.

Luss, H. 1977. A model for advanced reservations for intercity visual conferencing services. *Operational Research Quarterly* 275–284.

Mahajan, S., G. van Ryzin. 2001. Stocking retail assortments under dynamic consumer substitution. *Operations Research* **49**(3) 334 – 351.

Miller, B. 1969. A queueing reward system with several customer classes. *Management Science* **16**(3) 234–245.

Miller, C. 2014. Is owning overrated? the rental economy rises. *New York Times* URL `http://www.nytimes.com/2014/08/30/upshot/is-owning-overrated-the-rental-economy-rises.html?abt=0002&abg=1`.

Muth, E. J. 1979. The reversibility property of production lines. *Management Science* **25**(2) 152 – 158.

Niederle, M., A. E. Roth. 2003. Unraveling reduces mobility in a labor market: Gastroenterology with and without a centralized match. *Journal of Political Economy* **111**(6) 1342–1352.

PA Department of Human Services. 2015. Governor's 2015-2016 executive budget. `http://www.dhs.state.pa.us/publications/budgetinformation/dhsbudget/index.htm`. Online; accessed 16 April 2015.

Papier, F., U. Thonemann. 2010. Capacity rationing in stochastic rental systems with advance demand information. *Operations research* **58**(2) 274–288.

Papier, F., U. W. Thonemann. 2008. Queuing models for sizing and structuring rental fleets. *Transportation Science* **42**(3) 302 – 317.

Patrick, J., M. Puterman, M. Queyranne. 2008. Dynamic multipriority patient scheduling for a diagnostic resource. *Operations Research* **56**(6) 1507–1525.

Rees, M. A., J. E. Kopke, R. P. Pelletier, D. L. Segev, M. E. Rutter, A. J. Fabrega, J. Rogers, O. G. Pankewycz, Janet Hiller, A. E. Roth, T. Sandholm, M. U. Ünver, R. A. Montgomery. 2009. A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine* **360**(11) 1096–1101.

Reilly, T. 2003. Transition from care: status and outcomes of youth who age out of foster care. *Child Welfare: Journal of Policy, Practice, and Program* **82**(6) 727–746.

Riordan, J. 1962. *Stochastic Service Systems*. Wiley, New York.

Rossetti, M. D. 2008. Java simulation library (JSL): an open-source object-oriented library for discrete-event simulation in Java. *International Journal of Simulation and Process Modelling* **4**(1) 69–87.

Roth, A., E. Peranson. 1999. The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American Economic Review* **89** 748–780.

Roth, A., T. Sönmez, M. U. Ünver. 2005. Pairwise kidney exchange. *Journal of Economic Theory* **125** 151–188.

Roth, Alvin E., Tayfun Sönmez, M. Utku Ünver, Francis Delmonico, Susan Saidman. 2006. Utilizing list exchange and nondirected donation through 'Chain' paired kidney donations. *American Journal of Transplantation* **6**(11) 2694–2705.

Savasaneril, S., P. Griffin, P. Keskinocak. 2010. Dynamic lead-time quotation for an m/m/1 base-stock inventory queue. *Operations research* **58**(2) 383–395.

Savin, S. V., M. A. Cohen, N. Gans, Z. Katalan. 2005. Capacity management in rental businesses with two customer bases. *Operations Research* **53**(4) 617 – 631.

Shanthikumar, J. G., D. D. Yao. 1987. Optimal server allocation in a system of multi-server stations. *Management Science* **33**(9) 1173 – 1180.

Tainiter, M. 1964. Some stochastic inventory models for rental situations. *Management Science* **11**(2) 316 – 326.

Takács, L. 1962. *Introduction to the Theory of Queues*. Oxford University Press, New York.

Tang, Christopher S., Sarang Deo. 2008. Rental price and rental duration under retail competition. *European Journal of Operational Research* **187**(3) 806–828. doi:10.1016/j.ejor.2006.03.061. URL `http://www.sciencedirect.com/science/article/pii/S0377221706007843`.

Tayur, S. 1993. Structural properties and a heuristic for kanban-controlled serial lines. *Management Science* **39**(11) 1347 – 1368.

van de Vrugt, M., N. Litvak, R. Boucherie. 2014. Blocking probabilities in erlang loss queues with advance reservation. *Stochastic Models* **30**(2) 187–196.

Virtamo, J., S. Aalto. 1991. Stochastic optimization of reservation systems. *European Journal of Operational Research* **51**(3) 327–337.

Whisler, W. 1967. A stochastic inventory model for rented equipment. *Management Science* **13**(9) 640 – 647.

Wortham, J. 2009. Rent the Runway offers designer dresses in the Netflix model. *New York Times* URL `http://www.nytimes.com/2009/11/09/technology/09runway.html`.

Yannetta, T. 2013. No kidding: Rent the Runway's sample sale returns April 1st. URL `http://ny.racked.com/archives/2013/03/26/rent_the_runway_4.php`.

Zhang, Y., M. L. Puterman, M. Nelson, D. Atkins. 2012. A simulation optimization approach to long-term care capacity planning. *Operations Research* **60**(2) 249 – 261.

Zhu, X., M. Veeraraghavan. 2008. Analysis and design of book-ahead bandwidth-sharing mechanisms. *Communications, IEEE Transactions on* **56**(12) 2156–2165.