ABSTRACT

Title of dissertation: ESSAYS ON BUSINESS ANALYTICS

Liyi Gu
Doctor of Philosophy, 2019

Dissertation directed by: Professor Ilya O. Ryzhov
Department of Decision, Operations
& Information Technologies

With the rapidly increasing availability of business-related big data in recent years as well as the advancements in statistical and machine learning techniques, business analytics (BA) is becoming an essential practice to explain past events, predict future trends and optimize decision making. Using BA, the two essays in this dissertation aim to address some important questions in two emerging topics: humanitarian fleet management and social behaviors in online gaming industry.

In the first essay, we analyse how vehicle management is carried out in a humanitarian setting. In humanitarian fleet management, the performance of purchase, assignment, and sales decisions is determined by dynamic interactions between the fleet composition, the time-varying and uncertain demands on the fleet, and the depreciation of the vehicles as they are exploited. We propose to evaluate purchase, assignment, and sales policies in a realistic simulation environment that directly models heterogeneous vehicle attributes and tracks their evolution over time. Using data from a large international humanitarian organization (LIHO), the simu-

lator can identify the rationale behind seemingly ad-hoc decisions by field managers at LIHO. For instance, by selling vehicles later than LIHO recommends, managers are actually reducing their costs; similarly, managers decline to switch vehicles between mission types because the benefits to the operational cost turn out to be marginal at best.

In the second essay, we conduct an empirical study of the relationship between social interaction and user engagement, retention, and purchase behavior, based on a high-resolution player-level dataset from a major international video game company for one of its premier titles. We engineer a set of features that characterize social behavior within the game, and link these behaviors to several measures of user engagement using statistical and econometric models. Our results show that user engagement is highly correlated with certain social dynamics; meanwhile, social interaction does not always translate to better retention rates or more purchases. In some cases, high dependence on a small set of friends is positively correlated with churn, indicating a tradeoff between engagement in one title and adoption of others. Early adopters are generally more responsive to the social experience than late adopters.

ESSAYS ON BUSINESS ANALYTICS

by

Liyi Gu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Ilya O. Ryzhov, Chair/Advisor
Professor Zhi-Long Chen
Professor Sean Barnes
Professor Kunpeng Zhang
Professor Paul M. Schonfeld

© Copyright by
Liyi Gu
2019

Acknowledgments

First and foremost, I would like to express my deep gratitude to my advisor, Professor Ilya Ryzhov for giving me the opportunity to work on various interesting and challenging problems over the last five years. His continued support, insightful comments, kind encouragement and useful advice helped me throughout my PhD study. From him I have learned the way of reading, thinking, developing ideas, writing and overall, researching.

I would like to thank my co-authors, Professor Shawn Mankad of Cornell University and Professor Mahyar Eftekhar of Arizona State University, for their cooperation, help and advice in my research. Special thanks go to my dissertation committee members, Professor Zhi-Long Chen, Professor Sean Barnes, Professor Kunpeng Zhang and Professor Paul Schonfeld, for their insightful comments on this dissertation.

I would also like to thank my friends and fellow PhD students, Weiming Zhu, Liu Ming, Wenchang Zhang, Tianshu Sun, Ai Ren and many more for their help and comments on my research, as well as making these years a colorful period in my life.

Last but not the least, I cannot say enough thanks to my parents, without whom I could never come to this day.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Introduction to Business Analytics

Data is changing the way business is run drastically. According to a report by Marr [2018], 2.5 quintillion bytes of data are created each day in 2018; Facebook manages 1.56 billion daily active users who generate 300 million photo uploads and 4.75 billion pieces of content shared every 24 hours [Noyes, 2019]; Walmart is building a state-of-the-art data center to process 2.5 petabytes of data every hour [Marr, 2017]. All the data collected will enable us to spot business trends, better manage risks and enhance competitiveness [Acito and Khatri, 2014]. Business analytics is the practice emerging in recent years that leverages this enormous amount of data and translates it into actionable insights. It has evolved to become part of every major business decision making process, and is one of the fastest growing fields in the business world [Bayrak, 2015].

Business analytics (BA), or business intelligence as is used more often in the information technology community [Sircar, 2009], is defined by Davenport et al. [2006] as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions". According to Holsapple et al. [2014], this definition indicates that BA capabilities

include:

1. Using techniques that are quantitative, qualitative, and combinations;
2. Using statistical techniques;
3. Using systematic reasoning;
4. Working effectively with models that are descriptive/explanatory, predictive or prescriptive;
5. Working effectively with evidence.

This gives a three-fold taxonomy of BA: Descriptive analytics, Predictive analytics, and Prescriptive analytics [Robinson et al., 2010, Liberatore and Luo, 2011, Delen and Demirkan, 2013, Evans, 2012]. The first phase of BA, descriptive analytics, mines historical data and summarizes it into meaningful charts and reports that can understand past performance and rationalize behind past successes or failures [Evans and Lindner, 2012]. Some typical questions that may be answered by descriptive analytics are: What was the revenue and profit in the last year? Which regions are the highest and lowest selling regions? What types of customers did we have and how much did they spend? Descriptive analytics allows for a multifaceted approach to data analysis due to the acceptance of either quantitative or qualitative data, and may serve as a precursor to future research. Limitations include the bias caused by the role researchers play, incapability of correlating variables or determining cause and effect, non-repeatable results and open interpretation of findings [McNabb, 2012].

The second phase of BA, predictive analytics, utilizes a variety of statistical and analytical techniques from data mining, predictive modelling and machine learning on the data to detect patterns or relationships and make predictions about future or unknown events [Nyce and Cpcu, 2007]. Typical questions answered by predictive analytics include: What will be the revenue and profit in the next year? If we invest one million dollars, how much might we get back and what is the risk of losing money? What will the stock price be in one month's time? One of the well-known applications of predictive analytics is credit scoring, where the customers' likelihood of paying future credit on time is extrapolated by their credit history, outstanding loans, personal data, etc. Limitations of predictive analytics include cost of implementing predictive analytics techniques, inherent inaccuracy of predictive models, and the need for clean, accurate data [Nyce and Cpcu, 2007].

Prescriptive analytics, the third and final phase of BA, goes beyond predicting future outcomes and uses optimization/simulations to suggest actions to benefit from predictions and show implication of each option, often identifying the best alternative to optimize certain objective [Stewart and McMillan, 1987, Evans and Lindner, 2012]. Many questions in OM literature are answered by prescriptive analytics, for example, how many newspapers should a newsvendor order; what inventory level should the reseller stack up to when making orders; what is the best pricing strategy to maximize profit, and so on. The techniques used in predictive analytics are often used along with optimization techniques to make decisions when there is uncertainty in the data. Limitations of prescriptive analytics include: the definition of a good objective function, often in the case of multi-objective problems; difficulty

in obtaining optimal solutions, given the scale and dimension of the problem, the characteristics of the objective function, and the complexity of the constraints; and the accuracy, sparseness and possible anomalies in the data which may render the prediction and optimization unreliable.

Business analytics has been applied in numerous domains, including financial analytics [Smelyanskiy, 2008], supply chain analytics [Trkman et al., 2010, O'Dwyer and Renner, 2011], transportation analytics [Simao et al., 2009], service analytics [Tian et al., 2008], marketing analytics [Spais and Veloutsou, 2005, Hauser, 2007], customer analytics [Davenport and Harris, 2007, Lichtenstein et al., 2008], enterprise optimization [Bryan, 2003], web analytics [Burby and Atchison, 2007, Kaushik, 2009], software analytics [Buse and Zimmermann, 2012], crisis analytics [Tomaszewski et al., 2007, Jern et al., 2010], risk analytics [Ray et al., 2008], and so forth.

1.2 Humanitarian Fleet Management

A sub-field of management science and one domain in which business analytics is applied, enterprise optimization, aims to answer the question "How to earn best-possible profit under ever-changing external and internal conditions?" [Bryan and Bielat, 2013]. Enterprise optimization can be viewed as the optimization of procurement options, sales opportunities and production capabilities [Bryan, 2003]. The first essay of this dissertation, with the needs to optimally procure, sell and assign resources (vehicles), falls under the category of prescriptive analytics and

enterprise optimization.

A number of international humanitarian organizations such as World Food Programme (WFP), World Health Organization (WHO) and International Committee of the Red Cross (ICRC) operate humanitarian programs in various countries. These programs can generally be divided into two categories: relief and development. Contrary to the speed requirement of relief operations which aim at providing relief to disaster-stricken civilians, development programs hold the objective of improving people's quality of life in the relevant developing countries, and needs to maintain large fleets of vehicles for elongated period of time, sometimes as long as tens of years. Therefore, development programs require cost effective fleet operations through centralized vehicle procurement [Kunz et al., 2015]. We interviewed field fleet managers from one large international humanitarian organization (LIHO), and found out that the three main decisions they made constantly were vehicle procurement, vehicle disposal (selling) and vehicle-to-mission assignment. However, during the interviews, we found that there are no official guideline from the headquarters on certain decision making, such as the vehicle-to-mission assignment, and field managers make ad hoc assignments. Sometimes even if there exist guidelines from HQ, managers still deviate from them and act differently. For instance, LIHO suggests that managers sell a vehicle as soon as it reaches 60 months in age or 150,000 kilometers in odometer, whichever comes first. However, Fig. 1.1 shows that around 70% of the vehicles are sold after at least one of the recommended thresholds. Therefore, we are interested in understanding the rationale behind the field managers' ad hoc decision making, and furthermore explore the possibility of better policies on the

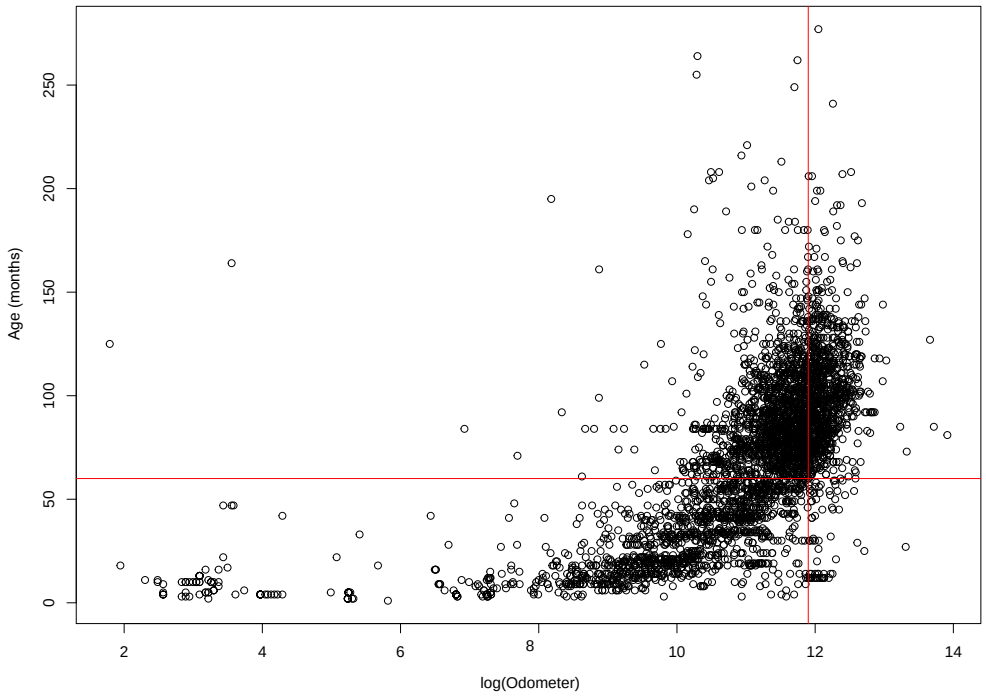


Figure 1.1: Selling age and odometer of all vehicles in LIHO. Red lines indicate the recommended selling thresholds.

three major decision making processes.

Counterfactual analysis is required in order to evaluate different policies without testing them in field operations, and in light of the complex system dynamics regarding the fleet and missions, we decide to model detailed attributes for each vehicle and mission, such as age, odometer, model type and utilization for vehicles and location, mission type and distance to cover for missions. This creates a huge state space in which searching for optimal policies is infeasible, thus we build a simulation framework to evaluate those "intuitive" policies that may be adopted by field managers; this also allows us to come up with better policies that are easily and readily implemented by field managers. The output of the simulation framework is

twofold: cost and mission completion rate. Mission completion rate is computed by the throughput in the simulator, but cost needs to be linked with the attributes, so we build empirical models that estimate how the attributes affect the operational cost.

One major difficulty stems from the demand model. From the data we observe that demands (missions) occur with a degree of randomness on top of a steady trend. Since multiple simulations are run to evaluate each policy, we need to generate demands that are similar to the historical data, but at the same time differ from each run. We solve this problem by calibrating the underlying demand level using the historical fleet size, and adding randomness to the demand generation process with the Compound Poisson model, whose parameters are estimated from the monthly mileage dataset of vehicles.

Simulations of policies across 3 mission types in two countries show that under different demand levels, it is always more cost effective to sell the vehicles later than the recommended thresholds, suggesting that field managers' decision to operate vehicles longer is financially justified. For the vehicle-to-mission assignment decision, it is better to concentrate vehicle utilization to only part of the fleet if the demand level is low, and balance loads among all vehicles when the demand level is high. Also, even though the idea of allowing vehicles to switch between mission types is enticing, simulation shows that only minimal benefit at most can be extracted from such switching, justifying managers' decision not to adopt this practice. From a research point of view, this essay shows that multi-attribute simulation-based approach may yield interesting results for humanitarian fleet management, and pos-

sibly humanitarian operations as a whole.

1.3 Social Behaviors in Competitive Online Gaming

The second essay of this dissertation concerns an already massive yet still rapidly growing industry: online gaming. One aspect of gaming experience, social interaction, has been widely recognized as an important motivating factor for gamers [Yee, 2006, Cole and Griffiths, 2007]. Jansz and Tanis [2007], Lenhart et al. [2008], [Weibel et al., 2008] andWohn et al. [2011] all found that social interaction is a significant part in the overall user gaming engagement and enjoyment. However, existing literature has not clearly shown how social interaction translates to revenue. While Bapna and Umyarov [2015] and Cheung et al. [2015] have shown that on a broader scale, engagement due to peer influence has positive effect on online sales, Guo et al. [2018] and Hamari [2015] argued that social influence and engagement is a weak driver of paid contents.

In this essay we look for new insights on whether and how social gaming behavior is related to player engagement, retention and purchasing behavior, with an empirical analysis on data of a leading international video game company for one of its major online first-person shooter titles. The data focuses on a sample of 1,309 players, randomly chosen from two cohorts of players classified by the company as early and late adopters, respectively. From data covering every round played by each focal player over a period of 123 weeks, we construct a set of covariates that characterize different aspects of social behaviors by these focal players. For each

focal player, we construct a "friends list" containing players (possibly non-focal) with whom he played repeatedly, each with a "degree" indicating the frequency. Then, the size of such friends lists can be treated as a measure of social behavior. We also decompose players in each friends list into "casual" and "close" friends depending on his "degree". This enables us to identify nuanced social effects stemming from each player's dependence on casual or close friends for enjoyment of the game.

Using regression analysis and survival analysis on the time-dependent panel data and the aggregated cross-sectional data, we make the following findings. In terms of engagement and retention, we find that while for both cohorts, social behavior is positively correlated with user engagement, this type of engagement is short-term and does not always translate into higher retention rates. In terms of product purchase, early adopters are more likely to convert social experience into purchases, but it comes at the expense of lower engagement in the focal title. Also, significant heterogeneities can be seen between early and late adopters, not only in social behavior and its effect, but also in their recent performance in the game. This may help to explain some of the mixed findings in the literature. One last surprising finding is that social interaction with casual friends appears to be better for both short- and long-term engagement. This could be caused by the wider network created by playing with casual friends, which comes with less intensity, and thus the network is more robust to shocks. Therefore for gaming companies, a better strategy may be to create wider communities rather than promoting intense social interactions.

Chapter 2: Evaluating Humanitarian Fleet Management Policies Using Simulation

2.1 Introduction

The complex settings of humanitarian operations have given rise to a considerable body of analytical work on decision-making in problems such as supply chain integration [Vanajakumari et al., 2016, Ni et al., 2018], facility location [Balcik and Beamon, 2008, Charles et al., 2016], prepositioning strategies [Acimovic and Goentzel, 2016, Salmeron and Apte, 2010, Rawls and Turnquist, 2012], inventory pooling mechanisms and field coordination [Toyasaki et al., 2017, Ergun et al., 2014], and aid distribution and last mile delivery [Balcik et al., 2008, Jahre et al., 2016]. However, implementing the results of these analytical approaches in the field has proven challenging, perhaps because the assumptions used in analytical studies often do not reflect the full complexity of the humanitarian context (see, e.g., Gralla et al., 2014 for discussion along these lines). Eftekhar and Van Wassenhove [2016] presented empirical evidence that field managers do not follow standard policies recommended either by researchers or by international agencies, suggesting that "what seems logical from the headquarters' perspective may be illogical or inconvenient for

the field."

We encountered similar statements, expressed by logistics officers and fleet managers at several major international humanitarian organization (HOs) in a series of interviews.[1] To give a specific example, one of these HOs recommends that field managers sell or dispose of vehicles once they are used for either 5 years or 150,000 km, whichever comes first. The consensus among interviewees, however, was that field managers did not follow this policy in practice, and in fact continued to exploit vehicles for much longer than either of these thresholds, although there appeared to be no single agreed-upon reason for this. In this essay, we propose an approach to counterfactual analysis in humanitarian fleet management that can explain this and similar discrepancies.

The main contribution of this essay is a holistic simulation environment that models and evaluates the acquisition, assignment, and disposition of multi-attribute vehicles in field operations. Specifically, we consider last mile delivery of humanitarian services from local offices to affected communities [Balcik et al., 2008]. The economic performance of a given set of policies (e.g., disposition policies) is subject to high uncertainty due to unpredictable demand, budget constraints, infrastructure problems and other issues [Eftekhar et al., 2014, McCoy and Lee, 2014], and depends on complex interactions between the fleet composition, stochastic and non-stationary demand, and the depreciation of trucks as they are exploited. Existing analytical and empirical research examines various elements of this problem in isola-

[1] We interviewed two freelance consultants, a logistics officer and a fleet manager at GOAL International, an executive fleet manager and two logistics officers at the International Committee of the Red Cross, a senior advisor of supply chain management unit at Catholic Relief Services, and two logistics officers at the World Food Program.

tion, such as fleet sizing and procurement [Keshvari Fard et al., 2017, Eftekhar et al., 2014], vehicle maintenance and replacement [McCoy, 2012, McCoy and Lee, 2014, Pedraza-Martinez and Van Wassenhove, 2013], and field vehicle fleet management [Pedraza Martinez et al., 2011, Eftekhar and Van Wassenhove, 2016]; however, the dynamic interactions between these factors are too complex to allow any analytical (or even numerical) treatment of optimal policies. An alternative approach is to use simulation to evaluate and compare various policies in a realistic setting. Simulation has been used in this way to study policies for many applications of interest to the public sector, such as HIV/AIDS prevention [Rauner, 2002], kidney allocation [Zenios et al., 1999], and emergency response [Kaplan et al., 2002].

The difficulty of policy evaluation in humanitarian fleet management is largely driven by the fact that both vehicles and demands possess heterogeneous *attributes*. For example, missions can be categorized by type and location; different types of missions arrive at different rates and impose different loads (kilometers to be traveled) on the fleet. Vehicles are acquired at different times, and their individual ages and odometers not only affect their performance, but also change dynamically over time. Managers' decisions are influenced by the attributes of the fleet (for example, they may prefer to place higher loads on newer or older vehicles). Our ability to model these attributes for each individual vehicle, track their evolution over time, use them to carry out decisions, and evaluate the impact of these decisions on costs incurred, constitutes the main distinguishing feature of our work as compared to the vast majority of the humanitarian OM literature. It also allows us to investigate many dimensions of fleet management within the same simulation environment.

Undoubtedly, any one dimension can be approached with an analytical model, but the deep insights that might be obtained come at the expense of simplifying other dimensions. In a certain sense, we do not know if existing models "agree" or "disagree," because they are not mutually compatible in their choices of what factors to include or exclude. Our approach, however, enables us to investigate a wide variety of practical questions within a single environment, which serves as a standardized test bed for ideas or observations that originated from previous literature.

We calibrated and validated the simulator using data provided by a large international humanitarian organization (LIHO). Using detailed data on purchase and maintenance costs, odometers over time, and salvage values, for multi-attribute vehicles in several field offices, we design several modules of the simulator: 1) We treat odometer data as a (censored) stand-in for demand, and develop a stochastic model of non-stationary, attribute-dependent demands over time. The simulated demand trajectories follow the same overall trend as what was observed historically, but incorporate random variation, allowing us to test different "what-if" scenarios. 2) Salvage data are used to calibrate a statistical model for the *depreciation* of vehicles as they are exploited over time; a vehicle is automatically removed from the fleet once it has lost all of its value, but can be sold earlier to redeem a portion of that value. 3) Refueling and maintenance data are used to calibrate statistical models that calculate short-term operating costs (fuel and maintenance). All of these costs are modeled as functions of vehicle attributes, which in turn are impacted by fleet managers' decisions. We evaluate various candidate policies for the following problems:

1. *Disposition.* We compare LIHO's recommended 5 year/150,000 km disposition policy with other combinations of age/odometer thresholds in realistic demand scenarios.

2. *Mission switching.* In practice, LIHO assigns each new vehicle to carry out missions of a particular type, and does not change this assignment for the remainder of the vehicle's lifetime. We counterfactually evaluated the possible benefits of switching types dynamically.

3. *Centralized procurement.* Managers are often able to choose between ordering a new vehicle through headquarters and waiting six months for it to join the fleet, or buying the vehicle locally with a 50% markup. We compared both schemes under realistic demand.

In some cases, our results agree with previous work: for instance, we find that procuring vehicles through headquarters ultimately reduces costs, as does Keshvari Fard et al. [2017]. In other cases, our results are quite surprising. For example, Pedraza-Martinez and Van Wassenhove [2013] studied vehicle disposition (using an analytical model) in a similar context, and recommended replacing vehicles much earlier than the HO's recommendation; this conclusion relied on various simplifying assumptions, one of them being that the monthly mileage[2] of every vehicle is equal to the same constant. Our simulation results consistently indicated the opposite, namely that LIHO's policy was too quick to dispose of vehicles. The best-performing sales threshold does become lower as the load on the fleet increases, but even for

[2]In fact, distance is measured in kilometers in the LIHO data; we use the word "mileage" informally in this essay to mean "distance travelled," but all the numbers are actually in km.

very high loads it is still higher than LIHO's recommendation. We observe this very consistently in many scenarios, even when our model is completely recalibrated with data from a different developing country on another continent. It is noteworthy that this observation is much closer to what actually happens in the field.

The mission switching problem presents another surprise. Many researchers have observed that coordination is challenging and often lacking in the humanitarian setting [McClintock, 2009, Balcik et al., 2010]. The general consensus is that more coordination would be desirable; thus, Pedraza Martinez et al. [2011] argues that improved coordination would reduce the unpredictability of demand on fleets, while Bhattacharya et al. [2014] finds that asset transfer between programs in an HO leads to more efficient operations. Pedraza-Martinez et al. [2013] investigates the design of mechanisms to encourage coordination between headquarters and local offices for more efficient procurement. However, we found that the economic benefits of switching are marginal at best, even in artificial "off-sync" scenarios where demand for one mission type increases just as demand for another type ramps down. We are not suggesting that coordination can never be useful, but we believe that our results explain why field managers do not seem to view it as an issue of primary importance: under current operating conditions (which may be subject to time constraints and suboptimal cost allocations; see Dolinskaya et al., 2011), the benefits may not be worth the effort.

In both cases where our results disagree with existing literature, they nonetheless agree with current practice. In other words, a simulation environment may not produce an "optimal" policy, but through a greater degree of modeling complexity

it helps to show how certain choices made by field managers are rational given the circumstances. In practice, their decisions do not seem to be based on any specific policy, including those considered in our study. Nonetheless, the managers' intuitive perception of the situation [Kahneman and Klein, 2009], based on their experience or other factors, may lead them to reject policies that are clearly suboptimal, as in the case of LIHO's recommended disposition policy. Similarly, the lack of improvement from mission switching in our simulation results helps to explain why, in the field, switching does not seem to be widely practiced or even encouraged. Of course, these decisions by fleet managers may not be based on any formal analysis; however, even if they do not articulate the rationale for their decisions [Tazelaar and Snijdes, 2013], the expertise behind these decisions often leads to successful completion of complex tasks, particularly under time pressure [Hayashi, 2001] and in unstable environments [Khatri and Ng, 2000]. Our results likewise suggest that these intuitive decisions should perhaps be regarded more carefully than has heretofore been the case in the literature.

In this way, this essay shows the power of simulation as a tool for counterfactual analysis in humanitarian fleet management. Previous applications of simulation to humanitarian operations include agent-based models [Crooks and Wise, 2013, Altay and Pal, 2014] that look for emergent patterns from interactions between AI agents, or discrete-event models that study facility location and configuration in rapid-onset disasters [Sahebjamnia et al., 2017]. Most of these papers either do not study resource allocation at all, or evaluate static decisions that are made once before the simulation starts. When it comes to simulating *dynamic* decisions that depend

on evolving system state variables, the literature is mostly limited to inventory management or budget allocation [Beamon and Kotleba, 2006, Iakovou et al., 2014, Chacko et al., 2016]. In fleet management, however, costs are determined by the management of operating assets (vehicles), in a way that changes over time based on the changing attributes of the fleet. This essay offers a way to handle these complex dynamics with a great degree of modeling granularity, enabling a detailed counterfactual evaluation of a very broad variety of policies.

2.2 Simulation for policy evaluation

Section 2.2.1 provides a high-level description of the simulation environment. Section 2.2.2 gives a rigorous discussion of the dynamics of the simulator that enable us to model and track the evolution of fleet attributes.

2.2.1 Overview of the simulator

Figure 2.1 provides a conceptual model of the simulation environment. The main *inputs* of the simulator, to be provided by the user, are as follows:

1. *Fleet composition.* First, we have to describe the vehicles in the fleet, by which we mean, not only the fleet size, but the specific attributes of each individual vehicle. In the LIHO dataset, these are the *age* of the vehicle (in months) at the start of the time period to be simulated; the *odometer* of the vehicle; the vehicle *type* (can be thought of as the make and model, but can also be aggregated by size, engine power or other factors); the *mission type* to which

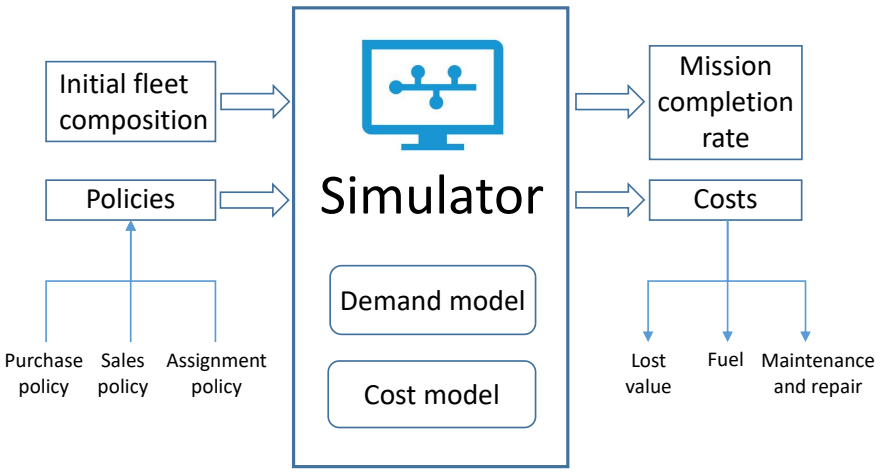


Figure 2.1: Conceptual model of the simulation environment.

the vehicle is assigned; the vehicle's *location* at the start of the planning period; the vehicle's *accident history* (e.g., number of accidents) and the *residual value* of the vehicle, which is bounded above by its original purchase price. It is entirely possible for every vehicle to have a unique combination of attributes (that is, no two vehicles in the fleet are exactly alike).

2. *Policies.* The decision rules to be evaluated may pertain to the *purchase* of new vehicles, the *sales* or disposition of aging vehicles, or the *assignment* of vehicles to missions. LIHO's recommended 5 year/150,000 km rule is one example of a sales policy; additional policies can be created and compared by considering different values for these thresholds. An example of a simple purchase policy is to immediately replace any vehicle that is sold, thus keeping the fleet size constant. An example of a simple assignment policy is to prioritize new vehicles for incoming missions (as is often done in practice; see Pedraza-Martinez

and Van Wassenhove, 2013). Policies should be sufficiently detailed to enable the simulator to automate all purchase, sales and assignment decisions based on the composition of the fleet at any given moment.

With this starting configuration, the simulator then begins to generate *demands* on the fleet from a stochastic simulation model; this *demand model* is a key module of the simulation environment, and should be customized to reflect the structure of the available data. In the LIHO dataset, demands have two key attributes, namely *type* (a simple categorical value in our dataset, but potentially could reflect the type of work that is required, the priority of the work, or the degree of danger involved in carrying it out) and *distance* that the vehicle has to travel (in km). Using the historical data, we can estimate the probability distributions of demands with different attributes, to give ourselves the flexibility of generating demands that resemble the historical data in a general sense, but may deviate from the precise historical values that happened to have been observed.

After generating the demands, the simulator automatically assigns demands to vehicles using the specified assignment policy, updates the attributes of the fleet (for example, assigning a mission to a vehicle will increase its odometer), and repeats the process for the duration of the planning period. In each decision epoch, a *cost* is incurred based on vehicle attributes across the whole fleet. Cost models comprise the other key module of the simulator, and also require extensive customization based on the data. Our approach was to fit a number of statistical models to predict fuel and maintenance costs, as well as on depreciation, as functions of the evolving

fleet composition. These models are necessary because our historical records do not provide detailed costs for every possible combination of vehicle attributes, and so costs must be inferred when running simulated scenarios that do not precisely match what was observed.

Once the required number of decision epochs has been simulated, the environment returns the *total costs* that were incurred by following the pre-specified set of policies over the course of the simulation run. Costs are determined by adding the total fuel and maintenance costs that were computed by the cost module, plus the costs of any new vehicles that were acquired by following the purchase policy, minus the residual values of any vehicles that were sold before the end of their lifespan (as specified by the sales policy). Since the demands are generated using stochastic simulation, running the simulator multiple times with the same starting conditions will produce different cost values; one can then perform many runs and apply standard output analysis techniques to estimate the mean performance of the given set of policies. In addition to cost, the simulator also returns the average *completion rates* for all mission types, which may be useful as a second objective for policymakers since minimizing cost is not the only goal in humanitarian contexts.

2.2.2 Dynamics of the fleet composition

In this section, we formalize the dynamics used inside the simulator to model the evolution of fleet attributes over time. We use the framework and notational system of stochastic dynamic resource allocation (see ch. 14 of Powell, 2011), in

which a "resource" (vehicle) is used to serve "demands" (missions). The state of a single vehicle is defined by an attribute vector a, composed of multiple attributes that may be numerical or categorical:

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \end{pmatrix} = \begin{pmatrix} \text{Age} \\ \text{Odometer} \\ \text{Location} \\ \text{Model type} \\ \text{Mission type} \\ \text{Accident history} \\ \text{Residual value} \end{pmatrix} \tag{2.1}$$

Let $\mathcal{A}$ denote the set of all possible attribute vectors. Let $t = 0, 1, ..., T$ be a time index representing the tth "decision epoch," or instant in time when it is necessary to make a purchase, sales, or assignment decision. Our odometer data is aggregated by month (unfortunately, we do not have access to more granular data), so we assume that one month elapses between time $t-1$ and t for each t; however, one could potentially use the same framework to model more frequent decisions.

Let R_{ta} denote the number of vehicles with attribute vector $a \in \mathcal{A}$ at time t, and let $R_t = (R_{ta})_{a \in \mathcal{A}}$ represent the overall vehicle inventory.[3] One may think of R_t as a vector that has very high dimensionality, but is very sparse (as $R_{ta} > 0$ for a very small number of attribute vectors at any given time). Of course, when

[3]To accommodate purchase decisions within the same modeling framework, one could also add a "dummy" attribute to (2.1) to represent vehicles that are available to be purchased, or that have been purchased and are scheduled to join the fleet, but that are not yet available for assignment.

implementing the simulator, one does not explicitly code R_t as a vector; rather, we use this notation to make our presentation more rigorous. Next, we denote by $\hat{R}_{ta}$ the exogenous (randomly occurring) change in the number of vehicles with attribute vector a between time $t-1$ and time t (for example, such changes could occur due to accidents). We let $\hat{R}_t = (\hat{R}_{ta})_{a \in \mathcal{A}}$ denote all such changes to the fleet.

The attributes of each demand (mission) are given by

$$b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \begin{pmatrix} \text{Location} \\ \text{Mission type} \\ \text{Travel distance} \\ \text{Order} \end{pmatrix},$$

and we similarly denote by $\mathcal{B}$ the set of all attribute vectors b. The first three attributes are self-explanatory. We assume that all missions with the same location and mission type are sorted in some order to be used for assignment, and the fourth attribute describes the (integer-valued) position of a given mission in this ordering. In the simplest case, missions could be sorted in the order in which they become known to the manager; if, however, the manager has advance knowledge of multiple missions, it is possible to order them in other ways, e.g., by travel distance or priority.

The quantity B_{tb} is defined to be the total number of missions with attribute vector b that are known to exist at time t, while $\hat{B}_{tb}$ denotes the number of *new* missions with attribute vector b that first appear at time t. In our study, we assume that demands are not backlogged (any demand that cannot be satisfied immediately

is lost), so $B_{tb} = \hat{B}_{tb}$; however, it is straightforward to incorporate backlogs into the model. As before, we let $B_t = (B_{tb})_{b \in \mathcal{B}}$ denote the full vector of all currently-existing demands. Because of the fourth attribute, two missions that exist at the same time cannot have two identical attribute vectors, thus making B_t a binary vector.

The *system state vector* $S_t = (R_t, \hat{B}_t)$ represents all the information that is known to the decision-maker at time t, before the next decision is made. We now proceed to modeling the decisions themselves. In the time interval of one month, a vehicle can be assigned to multiple missions; for this reason, an assignment decision d that can be applied to a vehicle is represented by an N-vector, where N is the total number of distinct demands (that is, the number of distinct attribute vectors b for which $B_{tb} = 1$) available at time t. These missions are sorted by their b_4 attribute values, and the kth element d_k of the vector d corresponds to a unique existing mission with attribute vector $b^{(k)}$ satisfying $b_4^{(k)} = k$. The statement $d_k = 1$ means that decision d assigns this mission to the given vehicle. If all entries of d are equal to zero, the vehicle remains idle for the entirety of the next time period. Denote $\mathcal{D}^D$ as the set that includes all possible d, and let $\mathcal{D}^M$ be a set of additional decisions not related to specific demands (for example, purchase decisions). Thus, every decision that can possibly be applied to a vehicle is an element of the set $\mathcal{D} = \mathcal{D}^D \cup \mathcal{D}^M$.

Now, let x_{tad} be the number of vehicles with attribute vector a to which we apply the decision d at time t, with $x_t = (x_{tad})_{a \in \mathcal{A}, d \in \mathcal{D}}$. This decision variable must

satisfy the conditions

$$\sum_{d\in\mathcal{D}} x_{tad} = R_{ta}, \qquad \forall a \in \mathcal{A}, \tag{2.2}$$

$$\sum_{a\in\mathcal{A}}\sum_{d\in\mathcal{D}^D} x_{tad} \leq B_{tb^{(k)}}, \qquad \forall k = 1, ..., N, \tag{2.3}$$

$$x_{tad} \geq 0, \qquad a \in \mathcal{A}, d \in \mathcal{D}, \tag{2.4}$$

Condition (2.2) means that, for any a, exactly R_{ta} vehicles are available to act on. Condition (2.3) means that N distinct missions are available to be assigned. Condition (2.4) is straightforward. We may also impose problem-specific conditions: for example, we may wish to prohibit assignments where the vehicle and mission are not in the same location, or if the mission is not of the type to which the vehicle has been assigned by LIHO; we may also limit the total number of missions that a single truck can fulfill per month (for example, by setting a maximum daily travel distance).

Let $\mathcal{X}_t$ be the set of all x_t that satisfy (2.2)-(2.4). The decisions are determined based on user-specified policies; thus, $x_t = X_t^\pi(S_t)$, where X_t^π is a mapping on the state space into $\mathcal{X}_t$, with the superscript π denoting the "name" of a particular policy. In words, a policy sees the system state S_t at time t and converts this state into a set of feasible assignment decisions x_{tad} for all a and d.

When we act on a vehicle with attribute a using decision d, the attribute vector of the vehicle changes. The new attribute vector $a' = a^M(a, d)$ is calculated using the *transition function* a^M, which has to be explicitly coded. For example, if $d = \{0, 0, ..., 0\}$ (that is, no missions are assigned to the vehicle), then $a'_1 = a_1 + 1$,

$a'_2 = a_2$, $a'_3 = a_3$, $a'_4 = a_4$, $a'_5 = a_5$, $a'_6 = a_6$, and $a'_7 = a_7 - K$, where K is the decrease in residual value (which must be estimated via a statistical model to be introduced later) of the vehicle resulting from leaving it idle for one month. For notational purposes, we define the indicator function

$$\delta_{a'}(a,d) = \begin{cases} 1, & \text{if } a^M(a,d) = a' \\ 0, & \text{otherwise.} \end{cases}$$

Then, the new fleet composition arising *as a result of our decision* is given by

$$R^x_{ta'} = \sum_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \delta_{a'}(a,d) x_{tad},$$

and the resource transition from t to $t+1$ is given by $R_{t+1} = R^x_t + \hat{R}_{t+1}$ if there are any random changes to the fleet. Again, since we assume that unfulfilled missions are not backlogged, we have $B_{t+1} = \hat{B}_{t+1}$ as mentioned earlier.

Finally, we describe the evaluation of the policy π. Let c_{tad} be the cost of applying decision d to a resource with attribute a at time t. The cost includes maintenance, repair, and purchase costs, which are obtained from statistical models to be discussed later. The cost may also be negative if the vehicle is sold (in that case, the revenue is equal to the vehicle's residual value attribute at the time of sale). The total single-period cost is given by

$$C_t(S_t, x_t) = \sum_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} c_{tad} x_{tad},$$

and the performance of the policy is calculated as

$$V^{\pi} = \mathbb{E} \sum_{t=0}^{T} C_t \left(S_t, X_t^{\pi} \left(S_t \right) \right), \tag{2.5}$$

the expected total cost incurred over the given planning period when policy π is used to calculate decisions. The expectation in (2.5) is taken over the joint distribution of $\left(\hat{R}_{t+1}, \hat{B}_{t+1} \right)_{t=0}^{T-1}$, and is computationally intractable due to the complex dependence of $(S_t)_{t=1}^{T}$ on these random quantities. However, it is quite straightforward to estimate (2.5) through simulation: given an initial state S_0 and a policy π, we can generate M independent trajectories $\left(\hat{R}_{t+1}^m, \hat{B}_{t+1}^m \right)_{t=0}^{T-1}$ for $m = 1, ..., M$ and report the sample average

$$\bar{V}^{\pi} = \frac{1}{M} \sum_{m=1}^{M} \sum_{t=0}^{T} C_t \left(S_t^m, X_t^{\pi} \left(S_t^m \right) \right), \tag{2.6}$$

where $(S_t^m)_{t=0}^{T}$ is the sequence of states visited in the mth simulation run. The average mission completion rate can be estimated in a similar way (the costs c_{tad} should be redefined accordingly).

2.3 Development and calibration of cost and demand modules

Section 2.3.1 describes the LIHO data used to calibrate the models that follow. Section 2.3.2 presents our proposed stochastic model for simulating demands on the fleet. Sections 2.3.3-2.3.5 discuss statistical models used to estimate costs due to depreciation, refueling, and maintenance, while Section 2.3.6 briefly discusses the simulation of accidents. Our presentation here focuses on data from a single

developing country, but the analysis is repeated on a second country in Appendix A.

2.3.1 Description of LIHO data

The LIHO data describes the fleet composition in a developing country (which we refer to as the "focal country" or FC) from 2004 to 2015, in which there are 16 local offices. Information about individual vehicles in FC was extracted from four separate datasets. First, we have an aggregate cross-sectional dataset of 454 vehicles and their attributes collected at the end of the observation period. The attributes of the vehicles used in our study include:

- Unique vehicle ID and local office where the vehicle was stationed;
- Vehicle model type, mission type given to the vehicle;
- Date registered, date on which odometer is recorded, final odometer;
- Purchase cost, sales value (if sold) or booking value (if not sold);
- Total number of accidents during the observation period.

We mainly use this dataset to estimate the residual value of vehicles of different model types, mission types, ages and odometers. Age is obtained by subtracting the starting date of registration from the final date of record.

The second dataset includes monthly traveling distances for individual vehicles from 2004 to 2015. Unfortunately, distances for individual missions are not available and we were required to work with aggregate distance traveled by each vehicle in

one month. We use this dataset to calibrate the mission arrival process; this task required the development of a probabilistic model to generate individual mission distances from a distribution that was calibrated using the aggregate data.

The third dataset contains the refueling history of vehicles, and provides the refueling cost as well as the age and odometer of the vehicle associated with each refuel. We use this dataset to estimate the fuel cost resulting from completing various missions by vehicles with different attributes. The fourth dataset pertains to the maintenance of vehicles, and is used to estimate the maintenance cost of vehicles with different attributes. For individual vehicles, the cost of each maintenance as well as the age and odometer of the vehicle is recorded.

In order to estimate the costs coherently for vehicles of different attributes, we eliminated vehicle model types and mission types that do not appear with an adequate number of records in any of the three datasets related to costs. Because many models and mission types appeared very infrequently in the data, most of our analysis was carried out on one vehicle model type with two possible mission types, involving 160 vehicles in the cross-sectional dataset, 122 vehicles in the refueling dataset, and 39 vehicles in the maintenance dataset.

2.3.2 Demand model

In the LIHO data, the monthly mileage of a vehicle ranges from $20 - 30$ to upwards of 3000, and there are also many zeroes that may represent idle vehicles or missing data. This level of variation is consistent for vehicles operating in dif-

ferent sub-delegations or handling different mission types. However, it is difficult to evaluate assignment policies based purely on the historical data. First, because monthly mileage is a consequence of the historical assignment decisions, we cannot directly calculate how a different policy would have performed in the same time frame. Second, the LIHO data only provides aggregate monthly mileage for each vehicle, and does not show how many individual tasks were performed or the size of each task. Third, monthly mileage only provides information about missions that were *completed*, and there is no way to know how many additional tasks there might have been that were visible to the fleet managers, but that could not be completed due to lack of resources or other factors.[4]

The goal of the demand module (see Figure 2.1) is to generate, in each time period, a stochastic number of missions whose individual mileage attributes also vary stochastically. The total monthly loads on the fleet should be "realistic," that is, they should exhibit the same general trends and magnitudes as the historical loads (for example, gradually rising and falling similarly to the historical mileage), but they should not be identical to the historical data because 1) we would like to have the flexibility to consider different scenarios, and 2) as discussed above, the historical data are not a perfect representation of demand to begin with. For example, if we view the historical data as being censored (since we only see mileage for completed missions), we might wish to generate demand trajectories that are consistently *higher* than historical, while following similar trends over time.

[4]Of course, we expect that there is a strong correlation between the observed load on the fleet and the total size of all visible missions, so we are using the former as a stand-in for the latter. This is consistent with other work in this area, and in any case, the LIHO dataset does not provide any other information that might serve as a more accurate observation of demand.

We found that the monthly mileages for vehicles did not exhibit any significant correlation between locations and mission types. For this reason, we assume that demands are independent across all location/mission type combinations, and so we estimate an independent demand model for each such combination separately from the others. In the following discussion, we take one mission type in the capital city of FC to illustrate how the model works.

To model the time-varying behaviour of the demand, we first construct a stochastic process $(L_t)_{t=0}^{T}$ that takes on positive integer values. The value of L_t for given t can be viewed in terms of the number of different humanitarian "projects" (distinct relief or development efforts) that are currently active and can generate tasks for the fleet to perform. This is typical in humanitarian logistics; however, the LIHO dataset provides no information about any such projects, so the process (L_t) is a modeling construct rather than an empirically observed quantity. One could also think about (L_t) as a kind of "latent fleet size," that is, the number of vehicles that we *should* have on hand in order to complete all the tasks. The actual fleet size is modeled with the resource variable R_t and may be completely different from L_t; in particular, the actual fleet size may be lagging behind the "latent" one, if the fleet managers make delayed reactions to sudden growth in demand.

The process (L_t) is modeled as a $G/G/\infty$ queue with batch arrivals. Each new batch represents a new active "project," for which we should have a certain number of vehicles (represented by the batch size) in order to complete all the tasks. The "service time" in this system represents the lifetime of the project; once this time runs out, the project disappears and stops generating tasks, leading to a reduction

in demand. The number of "servers" is infinite because we do not impose any limit on how many projects may be active at the same time. Thus, as the number of busy servers fluctuates over time, we will see some periods of very high demand.

Because our goal is to generate realistic demands that resemble the overall trajectory of the historical loads, the interarrival times and batch sizes are bootstrapped from the data. Since, in the monthly data, we can see the exact time periods when new vehicles enter the fleet, we simply use the times elapsed between two such "arrivals." Usually, multiple new vehicles are added to the fleet at the same time, providing us with the batch size. Of course, here we run into the issue that the data only provides information about actual vehicles that were acquired, which can at best be viewed as censored observations of L_t.[5] We do not see any way to eliminate this issue with the LIHO data that we have; however, the purpose of bootstrapping from the data is to provide us with a rough trajectory for L_t over time, and we rely on the stochasticity of the simulation to provide us with sufficiently many scenarios that deviate from the data in other ways (for example, by generating higher demands).

The service times for the queueing system could also be obtained by bootstrapping from the observed vehicle lifetimes; however, in the data, these lifetimes are right-censored since many vehicles are still in the fleet at the end of the observation period. For this reason, we used a Weibull distribution for the service time, and computed the parameters using maximum likelihood estimation with censoring.

[5]Note, however, that we do not even know *which* observations are censored since there may well have been months when the fleet was able to meet all of the visible demand. This is one of the reasons why it is difficult to apply standard statistical techniques for handling censored data in this particular setting.

The total mileage of all new missions that become visible at time t (for the given, fixed mission type/location pair) is then calculated as

$$D_t = \sum_{\ell=1}^{L_t} Y_\ell, \qquad Y_\ell = \sum_{n=1}^{N_\ell} X_n, \tag{2.7}$$

where N_ℓ is a positive integer representing the number of distinct tasks generated by project ℓ in time period t, the values X_n for $n = 1, ..., N_\ell$ represent the mileages required to complete each individual task, and Y_ℓ then represents the total mileage generated by project ℓ. We would like the flexibility to model demand at the level of individual tasks because this may affect the performance of various assignment policies (for example, consider a simple policy that always assigns tasks in order of increasing mileage, with the quickest tasks being completed first).

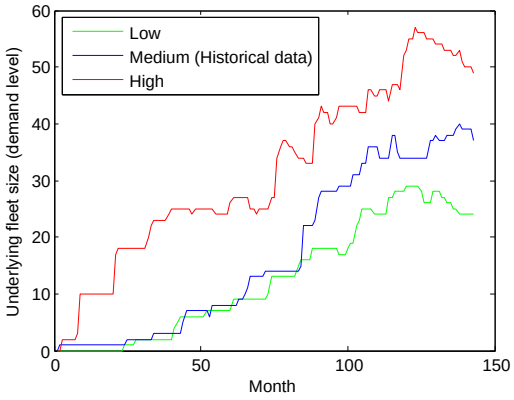
Since we are using the historical fleet size as a stand-in for L_t, we can treat the monthly mileage of each vehicle as an observation of Y_ℓ. Although the data do not tell us anything explicit about N_ℓ or X_n, nonetheless we can estimate distributions for these random variables by interpreting Y_ℓ as a compound Poisson random variable, that is, $N_\ell \sim Poisson\,(\lambda)$ and $X_n \sim Gamma\,(\alpha, \gamma)$. Under a particular transformation of $(\lambda, \alpha, \gamma)$, this becomes the well-known Tweedie distribution [Zhang, 2013], which is widely used in applications (e.g., in insurance claims modeling, see Smyth and Jørgensen, 2002) where it is necessary to "reconstruct" the individual components of an observed sum. The Tweedie distribution also allows

$(\lambda, \alpha, \gamma)$ to depend on covariates such as the historical fleet size.[6]

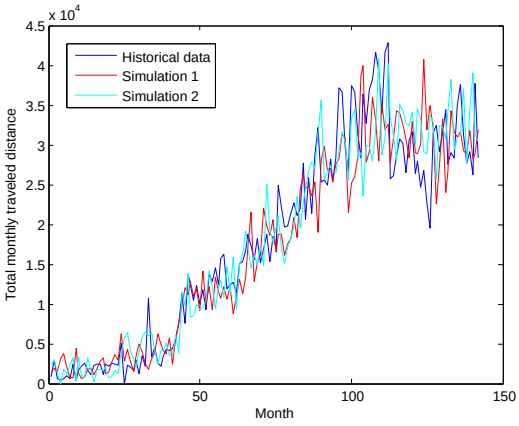
Figure 2.2 provides an illustration of how the output of the demand model is "realistic" without being identical to the real data. First, Figure 2.2(a) shows two realizations of the demand-generating process (L_t) compared to the actual historical fleet size (again, for this particular location/mission type combination). All three trajectories grow over time, and so the simulation output has the same general trend as historical. However, each individual simulated scenario deviates quite a bit from the historical trajectory; in particular, the number of "active projects" generated in the "high" scenario is consistently much greater than the historical fleet size. This trajectory is useful if we believe that the historical data are significantly under-reporting the visible demands. Second, Figure 2.2(b) fixes the process L_t to have the same values as the historical data, and uses (2.7) to randomly generate individual tasks. In other words, we are comparing the simulated values of Y_ℓ against the historical values. Again, we see that, while the simulation output is not identical to the historical data, all of the trajectories follow the same general trend. Of course, when L_t is also simulated, we will expect to see greater deviation from historical.

We close this discussion by reiterating that demand simulation must take two diametrically opposed concerns into consideration. On one hand, we want the simulation output to resemble the historical data, as otherwise the results of policy evaluation may not be relevant to LIHO. On the other hand, we also want to en-

[6] Eftekhar et al. [2014] points out that such a dependence may constitute evidence that the demand is censored; unfortunately, as stated earlier, we are not able to see which particular observations are censored. It may also be the case that, if there is a large number of projects, the managers may be able to combine some individual tasks together into a single "trip," leading to a slightly smaller number of tasks.



(a) Sample paths of (L_t).



(b) Sample paths of (D_t).

Figure 2.2: Simulated demands compared with historical data.

courage a certain amount of deviation from historical, both because we do not want to "overfit" our results to the one set of numbers that happened to have been observed, and also because the dataset itself is not completely reliable with respect to the demand.

2.3.3 Cost models: depreciation

From (2.1), recall that our resource model tracks the residual value of every vehicle over time. This value serves as a constraint on the vehicle's lifespan: once it reaches zero, the vehicle is automatically removed from the fleet. Residual value also

plays two roles in cost modeling. First, it determines our revenue in the event that we sell the vehicle. Second, at the final time T, the residual value of every vehicle remaining in the fleet is subtracted from the total cost, to avoid a bias in favour of policies that deliberately make fewer purchases late in the planning period.

We formulate and estimate a statistical model that can be used to calculate changes in residual value as a function of vehicle attributes and assignment decisions. Essentially, this model serves as part of the transition function a^M. Specifically, we use the zero-intercept Tobit model

$$\text{Dep\%} = \beta_1 \text{Age} + \sum_j \beta_{2j} \log(\text{Odometer}) \times \text{MissionType}_j + \beta_3 \text{NumAcc}, \tag{2.8}$$

where Dep% is the depreciation expressed as a percentage of the original purchase price, MissionType_j is a dummy variable that is equal to 1 if the vehicle is assigned to the jth distinct mission type, and NumAcc is the total number of accidents in the vehicle's history at the time of disposition. The Tobit model is used because the LIHO dataset contains an inflated number of zeroes (and no negative numbers) reported as residual values; we assume that any negative value for the right-hand side of (2.8) indicates that the vehicle has been rendered unusable and can only be salvaged.

We use the percentage loss rather than the actual residual value because the initial purchase price of a vehicle depends on factors such as accessories, payload, special design etc., that are not explicitly captured in (2.8). By modeling the loss relative to the original purchase price, we implicitly keep these factors in the esti-

mated residual value. The age and odometer variables are self-explanatory: both should affect the residual value negatively, or the percentage loss positively. However, the effect of odometer should vary with the mission type on which the vehicle operates, as some mission types inflict more wear on the vehicle, resulting in lower residual value. We apply a log transformation to the odometer because we expect residual values to drop more sharply in the early stages of a vehicle's life than in the later stages (an additional 100 km traveled should have less impact if the vehicle has already accrued $100,000$ km). Due to the small number of accidents in the data, we assume that the number of accidents has the same linear impact on depreciation across all mission types.

In (2.8), the intercept is set to zero, because a vehicle's residual value is initialized to its purchase price. The use of a zero-intercept model is further justified by the way in which the model is used inside the simulator. The LIHO data, which are used to estimate (2.8), record only the initial and final values of any given vehicle, but do not provide values at intermediate times. However, inside the simulator, the model is used to calculate *monthly* value reductions; essentially, we evaluate (2.8) with the ages and odometers in R_t, then apply the decision $X_t^\pi(S_t)$, evaluate (2.8) again with the new ages and odometers, and take the difference to obtain the depreciation that resulted from that particular decision. Any intercept term would be cancelled out in this calculation, so we do not include it.

Next, observe that (2.8) includes interaction terms between $\log(\text{Odometer})$ and the mission types, but does not include either main effect. The main effects for mission types are not included because (2.8) is a zero-intercept model (essentially

	Independent variable:
	Depreciation%
Age	4.190×10^{-4} ***
log(Odometer)×Mission1	0.06417***
log(Odometer)×Mission2	0.07720***
NumAcc	0.08300***
Observations	164
Log Likelihood	−4,196.056
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2.1: Estimation results for the depreciation model (2.8).

the reasons for omitting the intercept term also apply to these effects). The main effect of log (Odometer) is not included because this would create linear dependence among independent variables; as an alternative, one could omit one value of j from (2.8) and include the main effect of log (Odometer) instead. Since the primary purpose of this model is to generate costs inside the simulator, this is not a major issue.

Table 1 shows the estimation results for FC with one location and two mission types. The coefficients for all the regressors are positive and significant, which means that age, odometer, and number of accidents are all positively correlated with depreciation, as one might expect.

2.3.4 Cost models: fuel

Each data point in the LIHO refueling dataset represents one refueling for one vehicle. We make the assumption (in the absence of any information whatsoever in

this regard) that the tank is filled on each refueling, and thus the amount of fuel purchased corresponds to the amount that has been consumed during the distance traveled since the previous refueling of that vehicle (which we can obtain from the monthly odometer data). Then, for each refueling record, we can calculate the fuel cost per kilometer between two refuels of the same vehicle and relate this quantity to the vehicle's mission type, age and odometer at that time. We propose the linear model

$$\text{FuelCostPerKm} = \beta_0 + \beta_1 \text{Odometer} + \beta_2 \text{Age} + \sum_j \beta_{3j} \text{MissionType}_j, \qquad (2.9)$$

with additional fixed effects for each vehicle ID. To improve estimation quality, we removed the outliers and only considered data with traveled distance above 300 km.

	Dependent variable:
	Cost per Km
Odometer	1.632×10^{-7} ***
Age	2.506×10^{-4} ***
Mission2	0.009759**
Constant	0.1292***
Observations	11,040
R^2	0.301
Adjusted R^2	0.293
Residual Std. Error	0.029 (df = 10917)
F Statistic	38.486*** (df = 122; 10917)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2.2: Estimation results for the fuel cost/km model (2.9).

Table 2.2 shows the OLS regression result of the model. We can see that, since

all coefficients are significant, there exists a baseline cost of 0.1292 per km traveled, each extra kilometer on the odometer adds 1.632×10^{-7} to this cost, each extra year on the age adds 2.506×10^{-4}, and if the vehicle operates on mission type 2, an additional 0.009759 is added on the per km cost. Note that, even if the cost/km follows a linear model, the resulting fuel costs are *not* linear in the distance traveled: if the vehicle is new or has a low odometer, its fuel efficiency is at a higher level, reflected by the lower fuel cost per kilometer. The cost incurred for a fixed distance increases with the age and utilization of the vehicle.

2.3.5 Cost models: repair/maintenance

In the LIHO data, vehicles do not appear to follow a strict schedule of maintenance and repairs. Rather, they appear to receive several "levels" of maintenance approximately every 5000, 15000 and 50000 km, but also other services at seemingly unsystematic times and odometer readings. From this, we conjecture that field managers do schedule maintenance and repairs based on age and odometer, but that any such schedule is only followed roughly. We formulate the linear model

$$\text{CC} = \sum_j \beta_{1j}\text{Age}\times\text{MissionType}_j + \beta_{2j}\text{Odometer}\times\text{MissionType}_j + \beta_{3j}\text{Odometer}^2\times\text{MissionType}_j, \tag{2.10}$$

where CC is the *cumulative* maintenance/repair costs of one vehicle that were incurred since it entered the fleet. We use a zero-intercept model since the cumulative cost for a brand-new vehicle should equal zero; for the same reason, we do not include main effects for the various mission types. We also do not include main

effects for age and odometer, but (2.10) is equivalent to the model where one mission type is removed and the main effects are included. We include the nonlinear term Odometer^2 since we expect that the vehicle will receive more repairs as it is utilized more. Interaction terms between age/odometer and mission types are included because vehicles operating on different mission types may have different maintenance schedules, and more strenuous missions may require more repairs. We do not explicitly include accidents in (2.10) because the repair data do not provide sufficient detail to allow us to match specific repairs to specific accidents; however, since repair costs are still included in the data, (2.10) can be viewed as indirectly incorporating costs due to accidents in an average sense.

The LIHO dataset only provides maintenance data for 39 vehicles, and testing with a mixed linear model suggests that the level of between-group variability is sufficient to warrant the inclusion of a random effect representing vehicle ID. The estimation results are shown in Table 2.3. We can see that the coefficients of the quadratic terms are positive and significant, indicating that maintenance/repair costs do indeed grow more quickly as the odometer increases.

2.3.6 Generation of accidents

In Section 2.3.3, we estimated the impact of accidents on the residual value of a vehicle. In order to include accidents in our simulations, we also require a probabilistic model of how likely they are to occur in a single decision epoch. Unfortunately, our data are not sufficient to estimate such a model; for the one location

	Dependent variable:
	Cumulative costs
Age×Mission1	25.553^{***}
Age×Mission2	15.011^{***}
Odometer×Mission1	0.014^{***}
Odometer×Mission2	0.0083^{*}
Odometer^2×Mission1	3.21×10^{-7} ***
Odometer^2×Mission2	7.06×10^{-7} ***
Observations	735
Log Likelihood	−5,386.827
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2.3: Estimation results for the cumulative maintenance/repair cost model (2.10).

and two mission types considered in Section 2.3.3, we have records of only 12 accidents. For this reason, we constructed an artificial model in which the accident occurrence probability is linearly related to a vehicle's age and odometer. The coefficients of this linear model can be viewed as tunable parameters, and we chose them so that the total numbers of accidents in our simulations resembled those in the data (we refer to this as the "base case"). However, we also considered other parameter values with 2x and 4x the base accident probability, for the purpose of sensitivity analysis and also because there may have been accidents that were not recorded in the data.

2.4 Results and insights

We present a case study in which our simulator compared a number of purchase, assignment, and sales policies using the models in Section 2.3 calibrated to the LIHO data. Recall from Section 2.3 that, due to the specifics of the LIHO data, we focus on a single location with two mission types.

The various policies that were compared are described in Section 2.4.1. The first case considered in Section 2.4.2 focuses on sales and assignment policies under stable (stationary) demand. Section 2.4.3 considers realistic demand (in the sense discussed in Section 2.3.2) and introduces purchase policies. Section 2.4.4 compares centralized and decentralized procurement. Section 2.4.5 investigates the impact of allowing vehicles to change their assigned mission type (a practice not currently implemented by LIHO).

2.4.1 Description of policies

Recall from Section 2.2.2 that a policy provides a way to calculate a decision when given any system state. Thus, to run the simulator, the user must choose policies for purchasing, assigning, and selling vehicles. We consider a number of simple and intuitive choices for each of these categories.

Purchase policies. The simplest purchase policy is *pure replacement*, where we purchase a new vehicle only when an existing vehicle is removed from the fleet (either sold, or disposed after reaching zero residual value). Under this policy, the fleet size is constant. We mainly consider this policy in Section 2.4.2, where demand

is assumed to be stationary.

We also consider simple "reactive" policies that purchase new vehicles when the recent mission completion rate appears to be "low" (i.e., falls below some tunable threshold) or when the utilization of the existing fleet appears to be "high." In most cases, we assume that new vehicles join the fleet instantaneously upon request; however, in Section 2.4.4, we investigate the issue of lead time.

Assignment policies. We assume that managers cannot anticipate the arrival of new missions and must assign them to vehicles in the order in which they appear (are generated by the simulator). The following simple assignment rules are considered:

- *Balance*. This rule assigns an incoming mission to the vehicle that has currently (based on previously assigned missions) has the least traveling distance assigned to it for the month. Essentially, this rule attempts to balance the monthly load on the fleet.

- *Least/Most Odometer*. These rules assign an incoming mission to the vehicle with the least (most) mileage on the odometer.

- *Oldest/Newest*. These rules assign an incoming mission to the vehicle with the largest/smallest age attribute.

- *Myopic*. Assigns an incoming mission to a vehicle to minimize the immediate cost of the assignment (i.e., assigning the mission to this vehicle incurs less cost than assigning it to any other vehicle), calculated by adding fuel, maintenance, and depreciation costs.

Among these policies, Balance and Least Odometer can be viewed as workload-balancing rules, which have been widely studied in commercial transportation [Matl et al., 2017]. The Balance policy attempts to evenly divide the monthly load on the fleet, whereas Least Odometer attempts to balance total odometers across all vehicles. The Newest policy is inspired by tendencies observed in practice [Pedraza-Martinez and Van Wassenhove, 2013].

We impose a limit on the number of missions that can be assigned to a single vehicle. First, we calculate the maximum monthly traveling distance of any vehicle in the LIHO data, and divide this quantity by 30 to obtain a cap on daily traveling distance (approximately 120 km). The "travel distance" attribute of each generated mission (the value X_n in (2.7)) can be divided by this daily cap to obtain the number of days required by the mission. A vehicle cannot take more than 30 days' worth of missions; if it is unable to receive any more assignments, the next best matching vehicle (depending on the assignment policy) is used. If no vehicles are eligible to receive the next incoming mission, it is simply dropped (not completed).

Sales policies. LIHO recommends selling a vehicle once it has reached *60 months* in age or *150,000 km* of odometer, whichever comes first. We consider other combinations of these thresholds, such as 40 mos./100,000 km (40/100), 100 mos./225,000 km (100/225), 140 mos./300,000 km (140/300), as well as the option to run the vehicle into the ground (RIG), which means that we continue to exploit it until it reaches zero value and is automatically removed from the fleet.

2.4.2 Case 1: stable demand, sales/assignment policies

In this case, we assume that demand is stationary: the process (L_t) in (2.7) is set to a constant, but the Tweedie distribution is still used to randomly generate individual tasks. A pure replacement policy is used to procure new vehicles, thus keeping the fleet size constant.

We consider a time horizon of 200 months. The initial fleet composition consists of four vehicles, of which two are new while the other two record 20 mos./30,000 km and 40 mos./60,000 km on age and odometer, respectively. All vehicles are assumed to have been initially purchased for $30,000, which is also used as the fixed purchase price of all new vehicles.

Five levels of stationary demand were considered by setting $L_t \equiv i$ for $i \in \{1, 2, 3, 4, 5\}$. For levels 1 and 2, the demand is lower than the fleet capacity, meaning that the vehicles will tend to be under-utilized. Levels 3 and 4 are roughly equal to the fleet capacity (level 3 is slightly below on average, but can occasionally generate high loads), and level 5 is above the fleet capacity, meaning that the fleet will not be able to complete all the missions.

We fix each successive sales policy and run it together with every possible assignment policy. The smallest total cost across all assignment policies, as estimated by (2.6), is then reported as the performance of that particular sales policy. We found that, for any fixed demand level, there was very little variation in completion rates between policies; thus, demand levels 1-4 resulted in completion rates close to 1, and demand level 5 resulted in a completion rate of around 77%. For this reason,

we use total cost as the primary performance metric in this case.

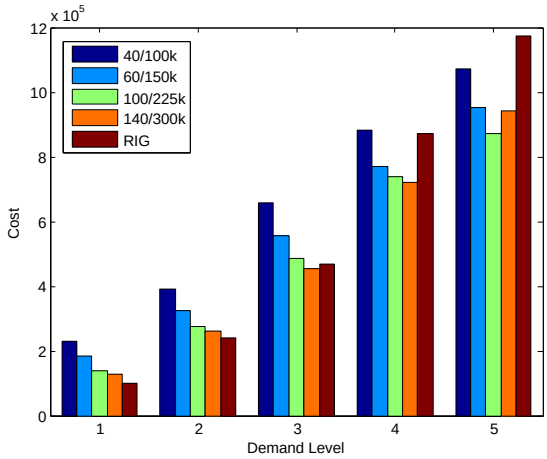


Figure 2.3: Cost comparison of sales policies under demand levels 1-5.

Figure 2.3 shows the results of the comparison with accidents not included (i.e., the accident generation probability was set to zero). The performance differentials between any two sales policies are all statistically significant as a result of running sufficiently many simulation runs. When demand is very low, the best performance is achieved by the RIG policy, which never sells vehicles; however, as the demand increases, the gaps between RIG and other policies close. At levels 3-4, the 140/300 policy starts to perform better, and at level 5, 100/225 becomes the best. Two observations can be made from these results:

1. The optimal sales threshold is always later/higher than the one recommended by LIHO;

2. However, the optimal threshold moves earlier/lower as the demand increases.

The first observation appears to dovetail with the sales decisions made by LIHO field managers, who generally continue to exploit vehicles after the 60/150 threshold has

been reached. We saw this in our data, where over 70% of dispositions occurred after at least one of these thresholds (with ages reaching 250 months and odometers exceeding 180,000 km), and of the remaining 30% many were listed as having zero age, suggesting that they were not sold but transferred elsewhere in the HO. Our interviewees agreed that the 60/150 rule was rarely followed, and similar observations were made by Pedraza-Martinez and Van Wassenhove [2013] in an empirical study.

This behaviour can be explained in terms of the tradeoff between utilization and residual value, first highlighted by Eftekhar and Van Wassenhove [2016] in the humanitarian context. Costs related to vehicle value (purchase and depreciation) are mostly incurred early on in the vehicle's lifetime; on the other hand, operational costs (fuel, maintenance and repair) become steeper late in the vehicle's lifetime. When the demand is low, value-related costs account for a greater share of the total cost and thus we prefer to continue exploiting the fleet rather than making new purchases. When the demand is high, the fleet is exploited more heavily and it becomes better to avoid keeping vehicles with high age/odometer in the fleet. Even then, however, the optimal tradeoff is made after LIHO's recommended threshold.

Next, we examine both sales and assignment policies and bring accidents into the picture. Table 2.4 reports optimal sales/assignment policy combinations for all demand levels and accident frequencies. With regard to sales, the pattern is consistent with Figure 2.3: RIG is optimal for low demand levels, followed by 140/300 and finally 100/225 as the demand increases. This holds for all accident frequencies, although the precise point at which we switch, e.g., from RIG to 140/300 may vary

since vehicles become less likely to reach the later stages of their lifetime when there are more accidents.

An interesting pattern can be observed with regard to assignment policies. In general, for low demand levels Myopic is preferable, followed by Least Odometer and Balance for very high demand.[7] At demand level 1, the Myopic policy essentially concentrates the demand on a portion of the fleet, exploiting those vehicles very heavily while leaving the others idle. This is optimal in low-demand level settings because residual value is then the most important cost driver. As the demand increases, it becomes necessary to use all the vehicles to complete the missions, and so workload-balancing rules start to perform better as they essentially control the growth of maintenance costs across the fleet, while also making vehicles reach the sales threshold later. For very high demand, all assignment policies become very similar as all the vehicles are fully utilized and have very similar monthly mileages; however, the Balance policy has a slight edge since, when the fleet is running at close to full capacity, only small missions can be "squeezed into" the loads, and the Balance policy will tend to assign more of these missions.

If cost is the main performance criterion, it appears that the presence of accidents in the simulator does not substantially change our conclusions regarding sales and assignment policies. For this reason, accidents are mostly omitted from the remainder of our study. We note, however, that if a field manager is concerned about accidents for reasons other than cost (e.g., personnel safety), reducing the

[7] In Table 2.4, there are two occasions when the Newest policy achieves the lowest cost. However, in both cases it is within 1% of either Myopic or Least Odometer, so there is no significant change in our conclusions.

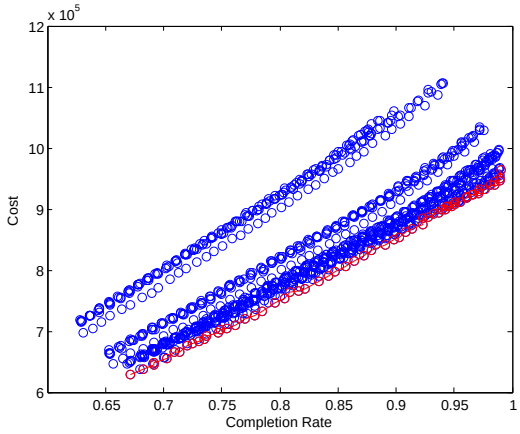
sales threshold will be an effective way to control these risks. Under the same accident level, the number of accidents tends to steadily increase with demand, but this increase can be reduced or even reversed by reducing the sales threshold, as illustrated by demand levels 4 and 5 in the baseline case.

2.4.3 Case 2: realistic demand, assignment/purchase policies

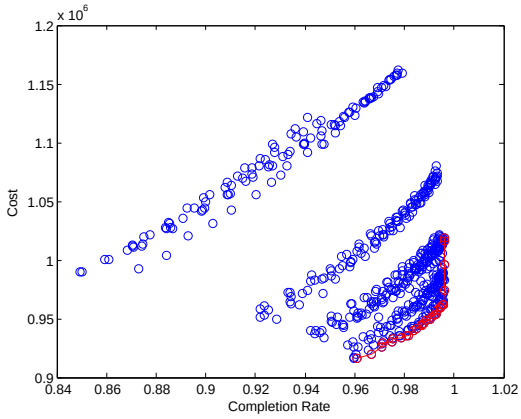
We now consider nonstationary demand, based on the model in Section 2.3.2 calibrated to 143 months of historical data for a single local office. We focus on the "high" demand trajectory shown in Figure 2.2, which demonstrates the same rising trend as in the historical data, but represents a hypothetical situation where the data under-reported the actual demands.

Our dataset tells us when new vehicles were purchased historically; however, because the historical fleet size was also used to calibrate the demand process in Section 2.3.2, the historical purchase schedule will tend to appear as if it is ordering too many vehicles relative to the simulated demand. For a more informative comparison, we construct two reactive purchase policies. The first policy purchases a new vehicle when the average mission completion rate over the past three months dips below a tunable threshold α (we label this policy "COMP" for "completion"), while the second policy purchases a new vehicle when the average vehicle utilization over the past three months, calculated as the number of days needed to complete all missions divided by the total vehicle-day units in the fleet, is above a tunable threshold β (we label this policy "UTIL" for "utilization"). Thus, high α and low

β lead to more purchases, while low α and high β lead to fewer.



(a) COMP policy: $\alpha = \{0.6, 0.61, ..., 0.99\}$.



(b) UTIL policy: $\beta = \{0.8, 0.81, ..., 0.99\}$.

Figure 2.4: Cost and completion rate comparison of reactive purchase policies under high demand.

Figure 2.4 shows the costs and completion rates, in the high-demand scenario, for different assignment/sales combinations together with each of the two reactive policies. The Pareto fronts in each subplot are highlighted in red. For the COMP policy, out of 64 points on the Pareto front, 44 use the Balance assignment policy, and 40 use the 100/225 sales policy. The Pareto-optimal points are nearly equidistant and monotonic in α, suggesting that the manager could easily use α as the "knob" for achieving a desired tradeoff between cost and completion rate. Including accidents

in the model did not substantially change the shape or composition of the Pareto front in Figure 2.4(a), but only moved it upward; for this reason, accidents are omitted from this discussion. For the UTIL policy, more points are concentrated on the right as most of the low β values produce very high completion rates. There are 28 Pareto points, of which 23 use the Least Odometer assignment policy, while 11 and 16, respectively, use the 140/300 and RIG sales policies.

Although the COMP policy tends to work well with Balance and 100/225, while UTIL prefers RIG and 140/300, these results are actually consistent with our observations from Section 2.4.2, where we saw that the optimal assignment/sales combinations depend on the demand level. Under COMP, purchase decisions lag behind the demand to some extent, and so this scenario resembles the high-demand scenario in Section 2.4.2, where Balance and 100/225 were indeed optimal. On the other hand, UTIL acts "preventively" when the load on the fleet appears to be growing, instead of reacting to a perceived shortage of vehicles; as a result, the fleet size is more closely matched to the demand, producing results that resemble low- to medium-demand levels in Section 2.4.2, where RIG and 140/300 were preferable. However, despite this distinction, the Pareto points for COMP and UTIL achieve similar costs under the same completion rate, though COMP provides more flexibility for trading off the two objectives.

2.4.4 Case 3: centralized purchases with lead times

Previously, we assumed that new vehicles join the fleet as soon as they are ordered. In practice, this corresponds to a situation where field managers purchase new vehicles from the local market. However, it often happens that managers have the additional option to purchase vehicles through headquarters, in which case there may be a lead time, perhaps as long as six months, before the vehicle will become available. Purchasing locally can be 50% more expensive than purchasing through the HQ [Besiou et al., 2014], creating a dilemma for fleet managers.

Let us return to the setting of stable demand from Section 2.4.2. Consider a hypothetical situation where the manager has some partial knowledge of demand in the near future: specifically, the manager knows the underlying value of L_t, but not the random number of tasks generated from L_t or their random magnitudes. (The manager does know the distributional parameters of these random variables, as these can be estimated from data.) Using this information, the manager can make a crude forecast of the next time that a vehicle in the current fleet will reach zero residual value; if this is predicted to occur within the next six months, the manager orders a new vehicle for $\$30,000$. We then simulate all possible sales/assignment policy combinations together with this method of purchasing, and report the lowest cost achieved for five demand levels. Likewise, we simulate all possible sales/assignment policies for a setting where the manager can instantly obtain new vehicles with a 50% markup in cost.

Table 2.5 reports the results of the comparison; since completion rates were

similar under both purchasing schemes, we focus on cost. Since the manager does not have perfect knowledge of the future, the savings obtained from centralized purchases are partially offset by increased utilization costs (since there may now be times when the remaining vehicles in the fleet are forced to take on more missions while waiting for a new vehicle to arrive). Nonetheless, there is a clear net gain from centralized purchases, with typical savings of around 10% relative to the decentralized scheme.

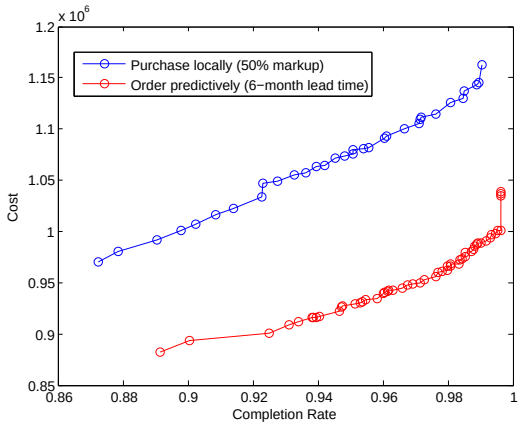


Figure 2.5: Comparison of centralized and decentralized purchases.

We also conducted a similar comparison for the realistic demands in Section 2.4.3. Here, in order for the two schemes to produce comparable completion rates, the decentralized setting uses the COMP policy ($\alpha = 0.8, 0.81, ..., 0.99$). In the centralized setting, the purchase policy anticipates future increases in demand (again, we assume that the fleet manager knows the trajectory of L_t) by keeping the fleet size at a higher level than what may be needed at the moment. Since UTIL also aims to preventively increase the fleet size, we use it as the purchase policy ($\beta = 0.6, 0.61, ..., 0.99$). The resulting Pareto fronts for both schemes are shown in Figure 2.5, and a fair comparison can be made by looking at the cost figures under

a fixed completion rate. Again, centralized purchases result in substantial savings (about 15%).

2.4.5 Case 4: switching of mission types

Current practice at LIHO locks in a single mission type for every new vehicle, and no vehicle is ever observed to complete missions from two different types. However, it is possible for a particular *type* of vehicle to be feasible for multiple mission types; in the data, we frequently see different vehicles with the same model type serving different mission types. It is natural to ask whether any improvement in efficiency can be achieved by allowing "switching" of mission types, or reassignment of vehicles from one type to another. Intuitively, allowing switching can help to use existing vehicles more efficiently without having to purchase new ones, which would reduce the cost of purchasing new vehicles and then incurring massive depreciation in the early stages of their lifespan. Since the data from FC include one vehicle type that is allowed to operate on two mission types, we can use our simulator to evaluate the effectiveness of switching.

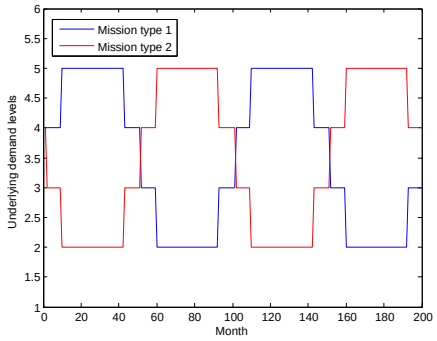
For illustrative purposes, we first consider an artificial scenario that one might expect to be biased in favour of switching. The initial fleet composition is the same as in Section 2.4.2, but the underlying demand levels (trajectories of L_t) are now two sine functions rounded to the nearest integer (see Figure 2.6(a)). Since switching should be most useful when the demand is high for one mission type and low for the other, one of the sine functions is lagged by half of its period, causing the two

demand levels to be completely off-sync. There is still stochasticity in the demands; the distributions of (X_n) are calibrated using data for the two mission types.

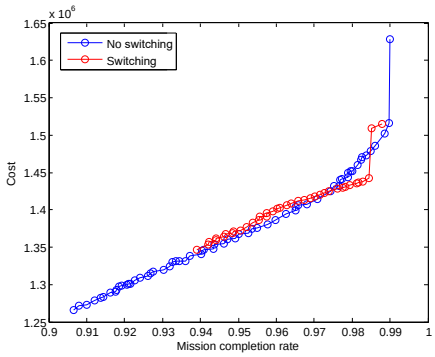
The switching policy works as follows. Every month, a switching decision is made based on the 3-month average mission completion rates and vehicle utilizations for both mission types: if the average completion rate for type 1 is 100%, the average completion rate of type 2 is *below* 100%, and the utilization for vehicles that currently operate on type 1 is under 70% (ensuring that we can afford to switch some of them), the oldest vehicle that operates on type 1 is switched to type 2.[8] Similar criteria are used to switch from type 2 to type 1. If no switching occurs either way, a purchase decision is then made using the COMP logic. To obtain depreciation due to switching, we compute (2.8) for a hypothetical vehicle that has been running on the new mission type, and has also been given the same distance to travel; we then take the difference in the estimated residual values to get the monthly reduction after switching, thus bypassing all previous mission type history.

Figures 2.6(b)-2.6(d) show the performance of different assignment/sales combinations under the switching policy described above, compared to applying the COMP policy independently to both mission types without switching. To avoid clutter, only the Pareto-optimal combinations are reported in Figure 2.6. We can observe that, for the most part, switching seems to offer very little benefit. The only exception is in Figure 2.6(c), where switching can achieve about a 3% reduction in cost with the same completion rate.

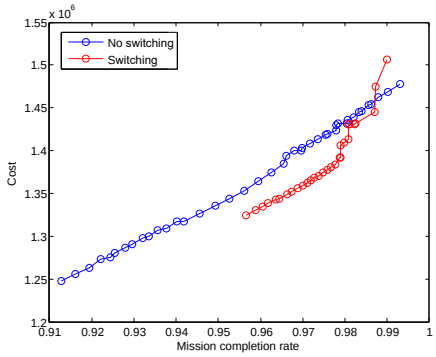
[8]We also considered other ways of choosing the vehicle to be switched (for example, based on least or most odometer), but this did not substantially change the conclusions.



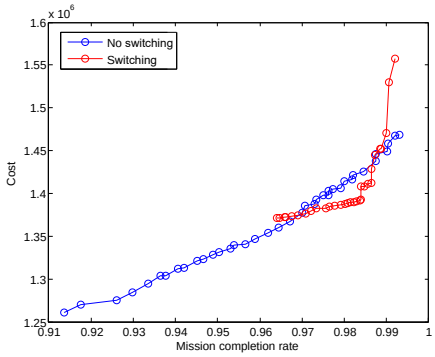
(a) Demand, period = 100.



(b) Performance, period = 20.



(c) Performance, period = 60.



(d) Performance, period = 100.

Figure 2.6: Demand levels and performance in switching vs. nonswitching with off-sync demands.

To obtain further insight into why switching achieves an improvement in this one setting but not in the others, we closely analyzed two scenarios with period 60 that have similar completion rates: one scenario uses switching, the Least Odometer assignment policy, the 100/225 sales policy, and $\alpha = 0.8$; while the other scenario uses Least Odometer, 100/225, $\alpha = 0.91$ and no switching. Figure 2.7 shows how costs grow over the course of the 200-month planning period in each of these scenarios. We see that *value-related costs* are consistently lower with switching than without switching, reflecting the fact that some purchases can be completely avoided by using switching. However, the operational cost is consistently *higher* with switching,

reflecting the fact that existing vehicles are being exploited more heavily. Switching is not able to completely avoid the tradeoff between these two costs, but in some cases it may be possible to achieve (small) total savings by using switching. For periods other than 60, the savings in value-related cost are outweighed by the increase in operational cost, and so switching does not produce any significant improvement.

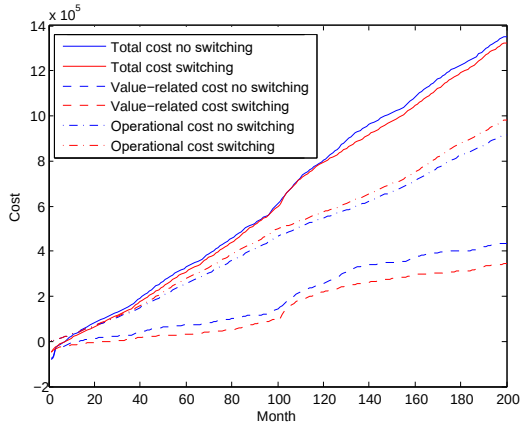
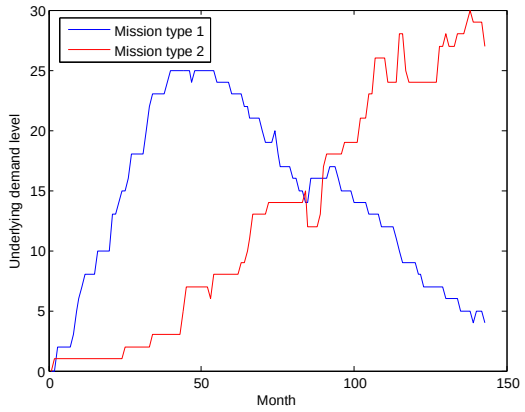
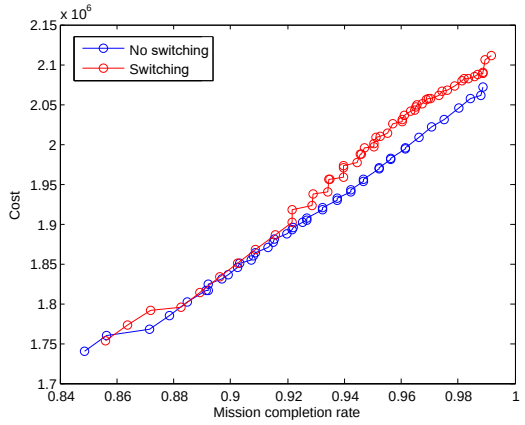


Figure 2.7: Change in costs over time for switching vs. not switching.

We repeated the same experiment with the purchase price increased to 40000 (from 30000), the idea being that higher purchase prices should lead to greater savings in value-related cost. Surprisingly, however, we found that this scenario did not yield any advantage for the switching policy; on the contrary, the advantage observed in Figure 2.6(c) disappeared. We omit the graphs due to space considerations, but essentially, this happens because, when the purchase cost is increased, the RIG sales policy becomes optimal in the nonswitching case. In other words, instead of switching vehicles, it is more efficient to just run them for as long as possible under their assigned mission type. As noted earlier, this result is not sensitive to the criterion used for switching; for example, if the newest vehicle, rather than the



(a) Realistic demand levels for two mission types.



(b) Cost and completion rate comparison for switching vs. not switching.

Figure 2.8: Comparison of switching vs. non-switching under historical demand.

oldest, is switched from type 1 to type 2, the oldest vehicle still has to shoulder more of the remaining type-1 demand, negating much of the benefit obtained by switching.

We also considered a scenario where the demand levels for both mission types were calibrated using LIHO data (with the model of Section 2.3.2). Figure 2.8(a) shows the trajectory of (L_t) for *each* of the two mission types; we see that type 1 generates heavy demand in the first half of the planning period, then gradually ramps down, while type 2 steadily ramps up over the course of 143 months. Potentially,

one might expect switching to offer some benefit when the two demand levels cross, i.e., we can switch some vehicles from type 1 to type 2. However, the Pareto fronts in Figure 2.8(b) show that this is not at all the case: in fact, switching leads to *higher* costs than not switching. We conclude that, while switching may seem attractive as a way of lowering value-related cost, any such advantage is almost always negated by a corresponding increase in operational cost. We believe that this offers a strong supporting argument for what is typically practiced at LIHO.

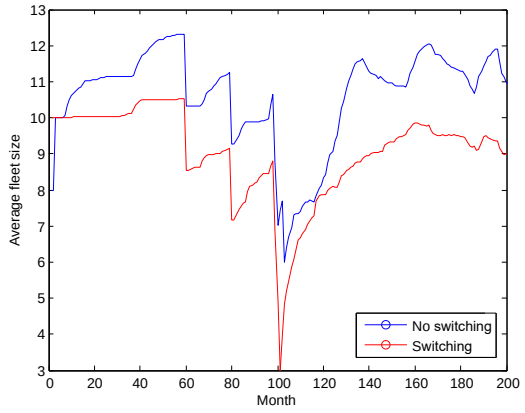


Figure 2.9: Change in fleet size over time under switching vs. not switching.

It is worth noting that switching may have an extra benefit in terms of fleet size. Figure 2.9 shows the fleet size over time for the two cases analyzed in the initial setting (the dip around month 100 occurs due to the 100/225 sales policy). It can be observed that the switching policy consistently keeps the total fleet size (both types combined) well below that of the non-switching policy. While this does not translate to big savings in total cost, it may have other benefits such as requiring less storage space, smaller maintenance crews, and so on.

2.5 Conclusion

We have presented a holistic simulation environment for counterfactual analysis in humanitarian fleet management. The simulator uses several modules, which are calibrated using real-world data, to evaluate policies for the purchase, assignment, and disposition of vehicles, in terms of both cost and mission completion rate. This environment models vehicles with heterogeneous attributes, and simulates how these attributes change over time in response to decisions that are made. In the context of the LIHO data, we used the simulator to obtain the following practical insights:

1. If cost is the primary objective, it is better to exploit vehicles longer than recommended by LIHO headquarters. The sales threshold should be reduced when the demand is higher, but there is no scenario in which the 60/150 threshold is cost-optimal.

2. When demand is low, it is better to assign it to a portion of the vehicles while leaving the others idle. For high demand, the best assignment policies are those that attempt to balance the load in some way.

3. Switching vehicles between mission types has marginal cost benefits at best: it is indeed possible to use them more efficiently and thereby reduce the fleet size, but the cost savings will be largely nullified by the accompanying increase in operational costs.

4. If reasonably accurate forecasts of future demands are available, one can order

new vehicles through headquarters and entirely compensate for the resulting long lead times. However, this imposes a greater burden of data collection and analysis on the field manager.

While these findings are valuable for LIHO and similar organizations, it is our view that the simulation-based approach presented in this essay is more broadly useful beyond the LIHO case. Using a single environment, we have investigated a broad variety of research questions in humanitarian fleet management, which previously had been studied independently using various mutually incompatible models. Our approach provides a level playing field for evaluating results obtained from such models together with practices observed in the field; in certain cases, we can identify the rationale behind "the facts on the ground," which we believe can lead to a more productive dialogue between researchers and practitioners.

Accident Level	Demand Level	Sales	Assignment	Total cost	Avg. no. of accidents
0	1	RIG	Myopic	100779	0
	2	RIG	Least Odometer	243211	0
	3	140/300	Least Odometer	452985	0
	4	140/300	Balance	721170	0
	5	100/225	Balance	870719	0
Baseline	1	RIG	Myopic	108951	1.749
	2	RIG	Least Odometer	258790	3.31
	3	140/300	Least Odometer	462249	3.876
	4	140/300	Least Odometer	744300	6.732
	5	100/225	Balance	887180	6.136
2x Baseline	1	RIG	Myopic	121923	3.356
	2	RIG	Least Odometer	272437	5.672
	3	140/300	Least Odometer	472207	7.664
	4	100/225	Newest	758999	9.722
	5	100/225	Balance	907844	12.319
4x Baseline	1	RIG	Myopic	136803	5.561
	2	RIG	Newest	290989	7.639
	3	RIG	Least Odometer	494179	14.846
	4	140/300	Least Odometer	790230	21.458
	5	100/225	Balance	948698	23.46

Table 2.4: Cost comparison of sales/assignment policies with varying levels of accidents.

Demand level	Purchase locally (no lead time, 50% markup)	Order predictively (6-month lead time)	% Cost saved from predictive ordering
1	108359	100670	7.10%
2	273837	242708	11.37%
3	501965	452871	9.78%
4	802372	719864	10.28%
5	997510	865034	13.28%

Table 2.5: Comparison of costs between purchasing locally and ordering predictively for stable demand.

Chapter 3: Social Behavior and User Engagement in Competitive Online Gaming: An Empirical Analysis

3.1 Introduction

Video games represent one of the most popular forms of entertainment among a broad range of individuals. More than 150 million Americans play video games each year, 42% of Americans play at least three hours per week [The Entertainment Software Association, 2015], and 97% of teens in the United States are gamers [Lenhart et al., 2008]. In 2014, over 135 million video games were sold to generate over $22 billion in revenue [The Entertainment Software Association, 2015]. Two of the largest publicly traded game development companies, Electronic Arts and Activision, have market capitalizations of over $40 billion and annual revenues of over $5 billion each, as of June 2018.

Social interaction has been widely recognized as an important motivating factor for gamers [Yee, 2006, Cole and Griffiths, 2007]. In recent years, the video game industry has undergone a massive transformation that has made the social interaction motive even more important. One major factor is the popularity of gaming on mobile devices [Soh and Tan, 2008]; the percentage of gamers that play primarily on

their smartphones has grown to 44% in 2014 from 0% of gamers just a decade earlier [The Entertainment Software Association, 2014]. Another factor is the emergence of social media platforms, which have themselves become a popular gaming venue: for instance, an estimated 20% of Facebook users play social games [Paavilainen et al., 2013]. The "traditional" console (e.g. Microsoft XBox and Sony Playstation) and computer game industry has also been heavily influenced by these developments, and new games on these platforms likewise heavily emphasize social features for communication, institutional formation, competition, and support networks [King et al., 2010]. These features are so prominent that cooperative online gameplay often overshadows the single-player features of a game.

The popular press has speculated [Huffington Post, 2014, Bloomberg News, 2013, CNN, 2012] whether console and computer game companies can survive in a market increasingly dominated by smartphones and tablets. In these market conditions of increased competition, a natural question is whether (and how) the social gaming phenomenon can be connected to business objectives such as customer retention and brand management. On one hand, it seems clear that social interaction is significantly correlated with user engagement and enjoyment. For example, Jansz and Tanis [2007] surveyed gamers and found that the social interaction motive was the "strongest predictor of the time actually spent on gaming." Lenhart et al. [2008] also found through surveys that 65% of game-playing teens play with other people who are in the room with them, and 27% play games with people who they connect with through the Internet. Playing against human opponents has been shown to positively influence user engagement [Weibel et al., 2008]. Additionally, Wohn

et al. [2011] found that perceived social interaction was a significant driver of participation in social network games, even if the games themselves lacked features to facilitate such interaction [Consalvo, 2011]. Game designers have developed gameplay mechanics for the express purpose of user acquisition and retention; many such mechanics are discussed in Hamari and Järvinen [2011].

On the other hand, it is not clear how well the social interaction motive translates to revenue. Terlutter and Capella [2013] identifies "social factors" as a largely unexplored direction for research on in-game advertising. In the broader context of social networks, Bapna and Umyarov [2015] has shown that peer influence causes a significant increase in the chances of buying a service. Within social *gaming*, however, Guo et al. [2018] finds that, while social influence may increase willingness to consume free services, the effect is much weaker when applied to paid content. The empirical literature is mixed on the issue: Cheung et al. [2015] found that "engagement exerted a positive influence on online sales," whereas Hamari [2015] concluded that enjoyment of a game does make the user more engaged, but actually "reduces the willingness to buy virtual goods."

In this article, we shed new light on these questions with a new empirical analysis of the potential links between social gaming and player engagement, retention, and purchasing behavior. Our analysis is based on player-level data provided by a major international video game company for one of its flagship first-person shooter titles. This game operates through online servers that players connect to using their console or computer to play with other players in real time within a shared game environment. The game has a variety of gameplay modes, though in general the ob-

jective is to defeat the other team by scoring the most points, "killing" members of the opposing team, or completing particular objectives, like capturing the opposing team's base. Players can play multiple rounds on a single gaming server or change servers at any time to play with different players and game settings.

Complete longitudinal data are collected for two cohorts of players, covering every round played by each cohort member (over $882,000$ unique game rounds) from product launch in late 2011 through early 2014. Players in the first cohort began playing the game within one month after its release date, whereas players in the second cohort began playing several months after the game's initial release. For all cohort members, data are available on game usage, game play, and historic behavior that we describe in detail in Section 3.3. For any round involving at least one cohort member, data are also available for other players that took part; however, their complete play history is unobservable.

From the data, we engineer a set of covariates that characterize different aspects of social behavior by cohort members, and link these social features to user engagement, which we measure using several metrics, mostly related to weekly play time. For each cohort member, we construct a "friends list" containing every player (not necessarily a cohort member) with whom the cohort member played repeatedly. This construction allows us to measure social behavior along multiple dimensions. First, the size of a player's friends list can be viewed as an absolute measure of social behavior, and thus allows us to test whether more social players exhibit different dynamics with respect to their playing and purchasing behavior. Second, we decompose each player's individual friends list into "casual friends" and "close

friends" (where the precise definition of these terms is player-specific), and subsequently measure the amount of engagement with each type of friend. This allows us to identify nuanced social effects stemming from each player's dependence on casual or close friends for enjoyment of the game.

We find that social behavior is indeed strongly correlated with both the decision to play each week and the resulting amount of play time. More social players, in both cohorts, exhibit a higher baseline level of engagement in the game. However, important nuances emerge when translating engagement to retention and revenue generation. We find that users appear to be harder to retain when their engagement depends more on playing with close friends, perhaps because they are more susceptible to churn when those close friends decide to stop playing. However, such players are also more likely to make more purchases. Conversely, more casual users, particularly those in the second cohort, make fewer purchases but demonstrate better long-term engagement in the focal title. Thus, there appears to be a tradeoff between user engagement in the focal title and user willingness to purchase new titles. We also observe an intuitive "cannibalization" effect, whereby the release of the next title in the series negatively impacts engagement and retention for the focal title.

We also find that the two cohorts react differently to the competitive side of the game. The early adopters in the first cohort demonstrate greater short-term engagement in the focal title after they perform particularly well, but this does not lead to more purchases, or even to better long-term retention for the focal title. By contrast, good performance does not seem to spur the more casual users in the

second cohort to become more engaged in the focal title, but does appear to have a more wide-ranging impact on their interest in gaming and willingness to buy other titles. This highlights the heterogeneity among gamers' preferences and may explain some of the mixed findings in the existing literature.

Thus, while our results verify the link between social interaction and user engagement in a game, they also imply that game designers should not simply expect highly social users to buy more. Nor will the encouragement of social behavior necessarily lead to increased revenues. If the game company's goal is to encourage early adopters of one title to shift to another, it may be useful to identify cliques of close friends, who are highly dependent on one another for engagement, for targeted marketing or special promotions. Alternately, if the goal is to increase retention of such users in the focal title, it may be useful to encourage them to make new contacts and social relationships online [Domahidi et al., 2014], to reduce their dependence on close friends. However, if the goal is to strengthen the company's support among casual players, social behavior appears to be unlikely to offer meaningful guidance. In those cases, it may be more beneficial to focus on the competitive side of the game, perhaps offering promotions or rewards to high-performing players. To support these recommendations, we demonstrate that our models have surprisingly high out-of-sample predictive power, and thus offer a solid basis for managerial action.

We briefly summarize the contributions of this essay. To our knowledge, this essay is the first large-scale empirical study to comprehensively examine different dimensions of social behavior in gaming and link them to engagement, retention, and revenue. We find evidence of significant correlations between social behavior and

subsequent increases in short-term engagement, which is largely in line with previous work in this area. However, our work is the first to identify significant differences between early adopters and casual users in terms of how this engagement relates to retention and purchases. In particular, we find that increased adoption of other titles comes at the cost of reduced engagement and retention in the focal title. The predictive power of our models also suggests that our proposed measures of social behavior may have significant practical utility, particularly when applied to early adopters.

3.2 Hypothesis Development and Literature Review

The increasing popularity of video games has generated research interest in a variety of fields, including education [Gee, 2003], economics [Meagher and Teo, 2005], psychology [Sherry et al., 2006], operations [Turner et al., 2011], and information systems [Zhu and Zhang, 2006]. Managerial problems studied by this literature include modeling of users' purchase behavior [Hanner and Zarnekow, 2015], sales forecasting [Yang et al., 2014], utility modeling [Park and Lee, 2011], marketing [Turner et al., 2011], and product innovation [Arakji and Lang, 2007]. Much of the existing empirical work is exploratory in nature and based on surveys or interviews; for example, Lehdonvirta [2009] identifies attributes of virtual goods that drive purchase decisions, while Westwood and Griffiths [2010] presents a classification of gamers into six types (one of which is "social gamers") influenced by different motivational factors.

Our main objective in this study is to use observational, player-specific usage data to understand and improve player engagement, retention and purchasing behavior. The business case for our analysis is supported by the literature on customer engagement, which emphasizes the influence of customer satisfaction on loyalty, customer retention, and hence profitability [Reichheld, 1992, Voss et al., 2008]. Greater customer engagement provides the game company with opportunities to study their more profitable customers, develop better services, and further improve customer loyalty [Mithas et al., 2005]. Specific to the game industry, Hsu and Lu [2007] has found that loyalty is positively impacted by enjoyment, while Cheung et al. [2015] presents evidence that increased engagement contributes positively to game sales.

Next, we present four research hypotheses and their theoretical motivations.

Hypothesis 1: Social gaming increases play time and retention rates.

The literature has proposed multiple detailed definitions of customer engagement, which can be measured through surveys or interviews [Cheung et al., 2011, 2015]. The game company in our study, however, relies on observational data on players' in-game behavior. Thus, in this essay, we use the player decisions to play and the amount of time spent playing as proxies for engagement. The idea is that play time is positively associated with satisfaction; conversely, a player is likely to have become dissatisfied or lost interest in the game if he or she stops playing it entirely for the remaining time covered by the data.

Under this framework, Hypothesis 1 can be motivated by the large literature on customer satisfaction and loyalty for experience-driven products, where customer engagement can be more important than simply providing quality service [Pullman

and Gross, 2004]. Deeply engaged customers are more likely to become enthusiastic fans who generate positive word of mouth and increase their spending [Pine and Gilmore, 1999, Vivek et al., 2012]. Thus, moving beyond operational factors, like timeliness and dependability [Heim and Sinha, 2001, Mollenkopf et al., 2007, Kumar et al., 2011], user interface navigation [Heim and Sinha, 2001], among others [Srinivasan et al., 2002, Craighead et al., 2004], the goal of the gaming company is to also provide experiences that create "emotionally engaged customers" [Voss et al., 2008].

Studies focusing on the video game industry in particular indicate that social gaming can be the basis for unique and memorable experiences for players [Jansz and Martens, 2005]. Peña and Hancock [2006] analyzes text messages from a similar online game and finds that players produce much more socioemotional than task content that is emotional and positive in tone. Sherry et al. [2006] identifies social interaction and peer pressure as the main reasons why many individuals play video games. The potential importance of social gaming to play time and retention relates to key CRM findings that customer satisfaction and retention can be driven from relationship commitments [Gustafsson et al., 2005]. In our context, players want to keep playing as long as their friends are also playing. Moreover, social gaming likely helps further differentiate the game from competitors, which can help win loyalty [Seenivasan et al., 2016]. Hsu and Lu [2007] finds that perceived social cohesion, through participation in an online community, can indirectly influence loyalty to a game company.

Social gaming can also be related to the broad body of work on social influ-

ence. For example, social interactions have been shown to impact participation in the stock-market [Hong et al., 2004] and in the political process [McClurg, 2003]. In the retail setting, Luo [2005] finds that the mere presence of peers increases the urge for impulsive purchases, and that the effect increases in magnitude with relationship strength. Siegel [2013] provides several examples from telecommunications, financial services, and public health illustrating the importance of social effects in understanding individual behaviors. Kempe et al. [2003] studies how such effects can be exploited for disseminating ideas or encouraging product adoption; see Bond et al. [2012] for a case study of political influence through online social networks. We note, however, that our work also does not belong entirely to this area: competitive online play is task-oriented and involves a great deal of randomness (e.g., playing against randomly chosen opponents), and does not lend itself to the kind of social influence analysis performed on, e.g., social networks [Tang et al., 2009, Ye and Wu, 2010]. Rather, we examine the role of social *interaction* in gaming, and use the social influence literature to motivate Hypothesis 1.

We also acknowledge some arguments against the hypothesis that social gaming improves retention rates. For instance, Hu et al. [2016] shows that, when consumers are sensitive to peer decisions, a product can suddenly become popular due to herding behavior of other social individuals amplifying adoption decisions. Bapna and Umyarov [2015] provides experimental evidence that an adopting friend causes a significant increase in the chances of buying a service, which however may also imply reduced engagement in previous services. Based on these and other similar findings [Hong et al., 2004, Young, 2009], in our context, if a large portion of the

utility of the gaming experience comes from social gaming, then players may quit the game simply because enough of their friends have also quit or moved on to a different game.

Hypothesis 2: Gamers who experience more social gaming will make more purchases.

It is widely believed that increased customer engagement results in revenue growth, since, when coupled with data collection, firms have a chance to study their more profitable customers, design and develop better services, and hence, further improve customer satisfaction and loyalty [Mithas et al., 2005]. Hanner and Zarnekow [2015] finds evidence that users are willing to spend more money over time as they become more engaged. Similarly, Srinivasan et al. [2002] finds that loyalty in e-commerce has an impact on word-of-mouth promotion, as well as on willingness to pay more. Both of these factors are prominent in social gaming. Cheung et al. [2015] finds a positive link between engagement and sales.

However, counter-arguments are also available. Hamari [2015] finds that enjoyment of the game (a form of engagement) does lead to increased willingness to play the game, but may in fact reduce users' willingness to buy additional content. Similarly, Guo et al. [2018] finds that, while social influence exerts a significant impact on play time, the corresponding effect on money spent is much weaker. Finally, the weekly time in which a user can play video games may be viewed as a kind of "budget constraint," reducing any positive effect that engagement may have on spending [Fornell et al., 2010].

Hypothesis 3: Social behavior has different effects for early adopters vs.

casual gamers.

Rogers [2010] defines five categories of adopters of a new service or technology. The first two groups to buy a product, so-called "Innovators" and "Early adopters," are motivated by social and hedonistic outcomes, that is, these sub-populations make their purchase decision by determining how it will define their social status and also by how much pleasure they can derive from the product [Brown and Venkatesh, 2003]. The remaining categories of adopters are analogous to casual gamers in our study, and are driven by other factors, including social influences from their peers and utilitarian judgements about how much the product will improve their overall effectiveness.

An important point is that there are systematic differences in the primary motivations of these different categories; thus, characteristics that appeal to innovators and early adopters may have a different effect on casual users [Moore, 2002]. In the context of the gaming industry, Westwood and Griffiths [2010] has also discovered important heterogeneities between "types" (e.g., "social" or "casual") of gamers, while Drachen et al. [2012] segmented users according to their in-game performance. The existence of these various heterogeneities has important practical implications, e.g., for the design of competitions or game mechanics [Huang et al., 2013], as well as content creation on social media, online forums, official websites, and so on.

Hypothesis 4: Playing and purchasing behaviors are sufficiently predictable for managerial action.

In this application, predictive power has important practical consequences.

For instance, with sufficient lead time, identifying players who are at risk of stopping their play indefinitely gives the gaming company opportunities to try targeted marketing campaigns to stop the player from leaving [Siegel, 2013]. Our work on this hypothesis relates to the literature on churn that aims to identify customers that are likely to stop using a product or service. Methodologically, we build on past work that applies supervised learning techniques to predict customer churn [Hadden et al., 2007]. This literature tends to focus on financial services [Glady et al., 2009], and telecommunications and cable subscriptions [Burez and Van den Poel, 2009]. In each of these industries, the firm uses the output of the model to target individuals with promotions that entice them to keep using the product or service. Our work is the first, to our knowledge, to utilize such modeling techniques in the video game context.

3.3 Data Description

Section 3.3.1 describes the basic gameplay and user information available in the data. We significantly transformed these data for our study: Section 3.3.2 describes additional data processing and generation we performed to create attributes describing in-game behavior, and Section 3.3.3 discusses how we inferred and extracted the social behavior of users from the available information.

3.3.1 User and gameplay information

The dataset tracks the in-game behavior of 1309 "focal users," divided into two *cohorts* of 674 and 635, respectively. Cohort 1 users all began playing the focus game within the first nine weeks of its release, whereas cohort 2 users started after that. Nine weeks is sufficiently far into the life cycle of the product to distinguish between early and late adopters of the game: the first major downloadable expansion pack was released around the time of this cutoff. Both cohorts were monitored for 123 weeks of play, which we refer to as the observation window. The window ends several months after the release of the next game in the series, at which time most users have abandoned the focus game. During the observed time period, five major expansion packs (downloadable content packs or DLCs) were released, each with new game modes and several new levels or "maps." Throughout the life cycle of the game, users also had the option to purchase a premium account, which allowed greater customization of their in-game characters and provided free access to all future DLCs.

Users play the game by logging on to a server and joining *rounds*, which are large ongoing battles between two teams. Although it is possible for a team to conclusively "win" or "lose" a round, this is not the primary incentive of the game, as users may join rounds that are already in progress, leave them before they end, or even switch teams mid-round. Rather, users accumulate points, or *score*, for completing individual in-game tasks, mostly situated around "killing" members of the opposing team. A user is not permanently removed from the round when killed,

but rather can "respawn" and continue. In this way, although gameplay is team-based, users are rewarded for their individual performance; a skilled player who happened to be on a losing team is still able to earn a high score and progress in the game.

The dataset contains records of all rounds involving at least one focal user from either cohort. Focal users were selected randomly by the service provider and thus rarely play with each other. Most participants in the rounds are thus non-focal users who were either randomly matched with the focal users, or interacted with the focal users somehow outside the game. Although the dataset includes information about the in-game behavior of non-focal users, this information is limited since we are only able to observe non-focal users when they play with focal users. However, non-focal users are assigned unique IDs, so we are able to identify non-focal users who repeatedly play against the same focal users.

For each participant in a round, we can observe the *time* spent by the user in the round; the user's in-game *rank* at the start and end of the round, the total *score* earned by the user from the round; and gameplay-specific information such as the game mode and map used for the round, the user's chosen *role* (offensive, defensive, and in between), any *vehicles* utilized by the user during the round, etc. Multiple metrics of the user's performance are recorded (such as number of kills and deaths), but some were prone to data collection errors; we focused on combat score as the most reliable metric of performance. The data also included information about teams, but we could not reliably map every participant to a team; furthermore, since players can easily switch teams, we felt that this information was less important for

quantifying player skill and performance.

For focal users, we can also see their home country, their signup date on which they first opened their account with the service provider, and the platform they used (PC, PlayStation3 or Xbox 360). Age and gender information is occasionally available, but self reported and unreliable, so we chose not to include it in our models (however, we included random effects to model this and other unobservable information). We also have limited information about *purchases* made by the focal users, most notably whether or not they owned the previous title in the series, and also whether or not they bought a premium account. We also know the number of products owned by any focal user at the start of the study, which includes all games sold by the service provider, not just the same series as the focus game. However, subsequent purchases made during the observation window (with the exception of the premium account for the focus game) were only tracked in detail for cohort 1. In particular, for cohort 1 users, we can see whether or not they purchased the *next* title in the series.

3.3.2 Data processing and feature generation

We designed and generated a number of attributes to describe user behavior, substantially transforming the data in the process. Features related to social behavior will be described in Section 3.3.3; here, we describe other attributes. First, we aggregated the dataset by week: for each week and each focal user, we calculated the total play time of the user, i.e., the total hours spent in rounds during that

week. The aggregated dataset is smaller and thus more computationally tractable; furthermore, aggregation obviates the need to model day-of-week patterns (e.g., the fact that users generally have more time to play on the weekends).

We also generated additional control covariates describing the user's performance and in-game behavior during the week. We used the average score earned per hour of play as a stand-in for the user's performance or "skill" level during the week. We also calculated the proportions of rounds (among all rounds involving the user during the week in question) where the user employed particular roles or game modes. Similarly we controlled for the proportion of time spent by the user on maps from various DLCs (some maps are simply more popular than others), and the proportion of rounds where the user utilized various in-game vehicles. To further model the effect of new content on player engagement, we defined an indicator variable for each DLC, which was set to 1 during the four weeks following the DLC's release date, and 0 in other weeks. Similarly, another indicator variable is set to 1 once the user upgrades to a *premium* account. This variable remains at 1 for all weeks after the upgrade was made; if the player never upgrades, the variable is always zero. For cohort 1, we include similar variables to describe whether the user owns the next (or the previous) title in the series.

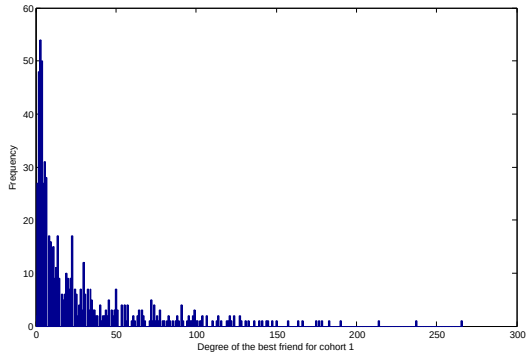
Our statistical models in Section 3.4 use various measures of user engagement as dependent variables. These are derived from users' weekly play time in the following ways: 1) a simple indicator set to 1 if a particular user plays at all in the given week; 2) an indicator representing whether or not the user's weekly play time was unusually high (see Section 3.4.2 for the exact definition); 3) the actual

play time, in hours, for the week; 4) the last week during the observation window in which the user played the focus game. We also consider models in which the dependent variable is the total count of products purchased during the observation window. These variables are described in more detail in Section 3.4.

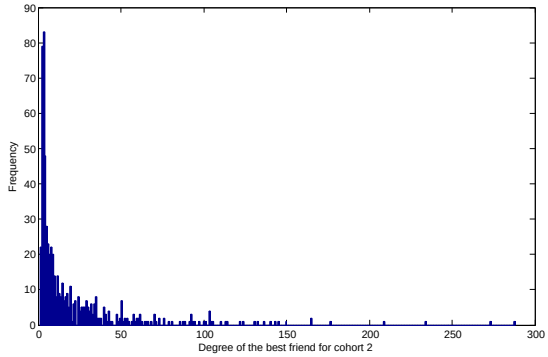
3.3.3 Features related to social behavior

Because focal users almost never play with each other, and non-focal users are only observable when they play with focal users, the social behavior of focal users must be inferred from repeated contacts with non-focal users. For a given focal user, we constructed a "friends list" containing all the non-focal users who played with that particular focal user on *two or more distinct days*. We used the number of distinct days, rather than the number of distinct rounds, because the game simply starts a new round when the first round finishes, so multiple mutual rounds within one day may be the result of this mechanism rather than any meaningful social interaction between users.

Throughout this essay, the term "friend" should be interpreted broadly. The users may only be acquainted online (pseudonymously), and may even have met on the game server. Nonetheless, playing the game repeatedly with the same person is a social experience even if the users do not know each other offline. We use "friend" as a catch-all term describing this experience. Many games have built-in social networking features that allow users to explicitly designate other users as friends; such networking data are studied, e.g., by Guo et al. [2018]. Unfortunately, the



(a) Cohort 1.



(b) Cohort 2.

Figure 3.1: Empirical distribution of the degree of users' best friends.

company we partnered with did not keep track of such data for the focal title, and so we chose to infer social behavior directly from the rounds.

Defined in this way, the focal user's friends can be sorted by *degree*, which we define to be the number of distinct days when the focal user played with that friend. Degree varies greatly between friends and also between different friends lists; some focal users have friends whose degrees are in the hundreds. For each focal user, we define the "best friend" as the friend with the highest degree; the empirical distribution of this quantity is shown in Figure 3.1. Comparing the degree of the best friend across all focal users, we find that 25% of these degrees are between 2 and 4, another 25% are between 5 and 10, another 25% are between 11 and 30,

and the final 25% are above 30. We use these quartiles to create "tiers" of social behavior: thus, a focal user is said to be in tier 0 if the degree of the user's best friend is 4 or less; tier 1 if the degree of the user's best friend is between 5 and 10; and so on.

In this way, tiers measure social behavior in a way that can be compared across users. A tier-3 user is more social than a tier-0 user, in the sense that the tier-3 user has a richer social experience involving at least one other person. Although higher-tier users often play more hours in general, this does not have to be the case: a user who plays twice a month, but with the same person every time, will belong to a higher tier than a user who plays every day, but always against random opponents. Similarly, higher-tier users tend to have longer friends lists in general, but this also does not have to be the case: a user who consistently plays with the same person may have a higher tier than a user who has a large number of casual friends. We define the tiers in this way in order to consider social behavior separately from sheer game time: some players simply have more free time than others and will play the game more overall, but their level of engagement is not necessarily higher than that of users who only play occasionally, but derive great enjoyment from the social experience.

Our second measure of social behavior aims to capture such user-specific attributes. For each focal user, we sort the user's friends list by degree and divide it in half. The bottom half is said to be "casual friends," while the top half is said to be "close friends" (again, these terms are meant to be interpreted broadly). Then, we calculate the numbers of casual and close friends with whom the focal user played

during each week; in this way, we account for the length of a user's friends list as a factor in the user's social behavior. Furthermore, for each week, we also calculate the proportions of rounds where the focal user played with casual friends, and a similar proportion for close friends. Note that the distinction between casual and close friends is individual for every player: a very social player is likely to have a large friends list and a high casual/close threshold, whereas a less social player may have a much lower threshold. Nonetheless, the less social player may still derive significant value from his or her close friends: consider two hypothetical friends who rarely have time to play, yet derive their enjoyment of the game primarily from joint sessions (in other words, when they do play, they account for nearly all of each other's rounds).

In this way, our generated attributes seek to identify users who are more social than others, but also to model instances of social behavior by users who are not highly social compared to others. We also include interaction terms between the indicator variables denoting tiers 1-3, and the proportions of rounds played with casual and close friends, to model the relative impact that playing with close friends has on more vs. less social users. Similarly, we include interactions between the tiers and the numbers of casual and close friends that played with the focal user during the week.

3.4 Methodology

Section 3.4.1 presents a model that predicts the simplest possible measure of user engagement, namely whether or not the user played at all in a given time period. We use this model to illustrate the main modeling tools used in our analysis. Section 3.4.2 discusses other variants that use different metrics of engagement as the response variable. Finally, Section 3.4.3 discusses concerns related to endogeneity and model robustness.

3.4.1 Statistical model

Let I be the number of users in a cohort. For week $t = 1, ..., T$, define $y_{i,t} = 1$ if the total play time of user i during week t was strictly greater than zero. Otherwise, let $y_{i,t} = 0$. We formulate a mixed-effect logistic regression model, given by

$$\mathbb{E}\left(y_{i,t} \,|\, b_i\right) = g^{-1}\left(\boldsymbol{x}_{i,t-1}^{\top}\boldsymbol{\beta} + b_i\right), \tag{3.1}$$

where g is the logit link function. The vector $\boldsymbol{x}_{i,t-1}^{\top}$ contains features that describe the behavior of user i during the *previous* week. This includes relevant gameplay and purchase information as described previously, as well as our engineered features describing social behavior. We also include user-specific features; for instance, a particular component $x_{i,t-1,k}$ may equal 1 if the user owned a premium account during week $t-1$. These features also include the number of weeks elapsed since the user was first observed to play, since we expect engagement to naturally decline

over time.

The term b_i is a random effect [Laird and Ware, 1982] that models unobservable variation between users. We make the standard assumption that the variables b_i are independently drawn from a common distribution $\mathcal{N}(0, s^2)$, where s^2 represents the amount of variation, and that the individual data points $y_{i,t}$ are conditionally independent given b_i. There are several arguments in favour of including random effects in the model. First, b_i models characteristics of user i that influence engagement, but are unknown to the service provider (for example, demographic information). The random effect also controls statistical bias due to multiple data points sharing a common source. Second, random effects explicitly model situations where the given set of users represents a sample of a much larger population, which is certainly the case in our application; estimating the population parameters allows for some degree of inference about the population as a whole. Third, the mixed-effect model has only one additional parameter to estimate (the variance s^2), whereas a fixed-effect model would require an additional variable for every user account. The increased size of such a model causes substantial computational difficulties even for moderately-sized datasets. Fourth, by modeling b_i as a random variable, we are expressing the service provider's inherent uncertainty about individual users and their behavioral patterns. This has the effect of "hedging" our estimates and reducing the risk of overconfidently reporting an effect as being significant in cases where the behavior of the data can be explained by unobservable user characteristics.

Given $\boldsymbol{\beta}$ and s, the joint probability of observing $y_{i,t}$ for $t = 1, ..., T$ and

$i = 1, ..., I$ is given by

$$L(\boldsymbol{\beta}, s) = \prod_{i=1}^{I} \int_{-\infty}^{\infty} \prod_{t=1}^{T} \left(\frac{e^{\boldsymbol{x}_{i,t-1}^{\top}\boldsymbol{\beta}+b_i}}{1+e^{\boldsymbol{x}_{i,t-1}^{\top}\boldsymbol{\beta}+b_i}} \right)^{y_{i,t}} \left(\frac{1}{1+e^{\boldsymbol{x}_{i,t-1}^{\top}\boldsymbol{\beta}+b_i}} \right)^{1-y_{i,t}} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{b_i^2}{2s^2}} db_i,$$

where the integral represents the expected the value of a conditional probability given b_i. The maximum-likelihood estimates $(\boldsymbol{\beta}^*, s^*) = \arg\max_{\boldsymbol{\beta},s} L(\boldsymbol{\beta}, s)$ can be computed using standard statistical software. The statistical significance of the results can then be assessed in the usual ways.

We note that, throughout our analysis, we estimate a separate model for each cohort of users. It is possible to combine the cohorts into a single dataset and expand $\boldsymbol{x}$ with additional features and interaction terms to model the differences between cohorts. However, this may cause computational difficulties for estimation due to the increased model size. Furthermore, in our case, we are interested in finding out whether the two types of users are influenced by different sets of factors, which can be addressed by estimating two separate models.

3.4.2 Other measures of engagement

We discuss three additional models that predict different measures of user engagement. On one hand, these models serve as a robustness check for each other and the basic model in Section 3.4.1; on the other hand, they also allow us to establish more nuanced relationships between engagement and social behavior.

Binging model. Let $p_{i,t}$ be the total time that user i spent playing the game during week t (this quantity is also included in the feature vector $\boldsymbol{x}_{i,t}$ in the basic

model). Define $y_{i,t}$ to be 1 if $p_{i,t} \geq q_{i,t}$, where $q_{i,t}$ is the 80th percentile of the set $\{p_{i,t'} : t' = 1, .., t-1\}$. Otherwise, $y_{i,t} = 0$. We then estimate (3.1) using this definition of $y_{i,t}$ as the response.

This model predicts whether user i will "binge," or play for an unusually high length of time, during week t. The precise definition of a binge is individualized, since users have different background levels of engagement and some simply play more than others. Furthermore, the definition depends only on the user's past history up to that week, since our intention is to track whether the user is exhibiting substantially more engagement than was previously typical for him or her. Although the response variable is still related to play time, the proportion of 1s observed is substantially smaller than in the base model from Section 3.4.1.

Survival model. The service provider is particularly interested in understanding when and how users stop playing entirely. We would like to identify any in-game factors that may contribute to dissatisfaction with the service, as well as external factors (such as other purchases, which may "cannibalize" the audience of the present title).

For user i, let $T_i = \min\{0 \leq t \leq T : p_{i,t'} = 0 \text{ for all } t \leq t' \leq T\}$ be the last time within the scope of the study that the user was seen to play. Since the overall time horizon T covers most of the life cycle of the game, most users have dropped out by the end. To estimate the distribution of T_i and relate it to user behavior, we formulate a Cox proportional-hazards model [Lin, 2000]. This model assumes that T_i is continuous, though in practice it is always estimated from discrete data. The

hazard function of T_i, given by

$$\lambda_i(t) = \lim_{dt \to 0} \frac{P(t \leq T_i \leq t + dt \,|\, T_i \geq t)}{dt},$$

can be viewed as a measure of how likely user i is to drop out in the next instant in time after t. The Cox model assumes

$$\lambda_i(t) = \Lambda(t) e^{\boldsymbol{\beta}^\top \boldsymbol{x}_i(t) + b_i}, \tag{3.2}$$

where Λ is a baseline hazard function, $\boldsymbol{x}_i(t)$ is the vector of regression features for user i at time t (note that the features can be time-dependent, as in Fisher and Lin, 1999), and b_i is a user-specific random effect [Sargent, 1998] as before. It is known [Tsiatis and Davidian, 2001] that $\boldsymbol{\beta}$ in (3.2) can be estimated semiparametrically without the need to specify any explicit form for Λ; standard statistical packages are available for this purpose [Thomas and Reyes, 2014]. The estimated coefficients have a standard interpretation: if $\beta_k > 0$, then large positive values of $x_{i,k}(t)$ are correlated with a higher chance of dropping out. Unlike our previous logistic regression models, the number of weeks since the user has started the game is not included among the regression features since the dependence of the hazard rate on t is already modeled implicitly using Λ.

Tweedie model for weekly play time. We also consider a model that predicts the quantity $p_{i,t}$, rather than simply the incidence of playing or binging. The Tweedie

regression model [Tweedie, 1984] makes the distributional assumption

$$p_{i,t} \sim \sum_{n=1}^{N_{i,t}} \hat{p}_{i,t}^{n}, \tag{3.3}$$

where $N_{i,t}$ follows a Poisson distribution, and the random variables $\hat{p}_{i,t}^{n}$, $n = 1, ..., N_{i,t}$ are independently drawn from a common gamma distribution (by convention, $N_{i,t} = 0$ implies $p_{i,t} = 0$). This distribution is then related to the regression features via the link

$$\mathbb{E}\left(p_{i,t}\right)^{q_1} = \boldsymbol{x}_{i,t-1}^{\top}\boldsymbol{\beta}, \qquad Var\left(p_{i,t}\right) = \phi \cdot \mathbb{E}\left(p_{i,t}\right)^{q_2},$$

where q_1, q_2, ϕ are additional parameters chosen through maximum-likelihood estimation. We use standard numerical techniques [Dunn and Smyth, 2005] for estimating the regression coefficients $\boldsymbol{\beta}$. To our knowledge, the statistics literature has not developed a mixed-effect version of the Tweedie model (however, see Section 3.4.3 for a modified version of this model that indirectly incorporates random effects).

The Tweedie model is widely used in the insurance industry [De Jong and Heller, 2008] to model random numbers of claims with random amounts. It has several features that are attractive for our application. First, although the response $p_{i,t}$ is continuous, the model allows $P\left(p_{i,t} = 0\right) > 0$, making it suitable for zero-inflated data where some users may simply not log on during a given week. Second, (3.3) has a natural interpretation in the context of weekly play time: $N_{i,t}$ is the number of gaming sessions for user i during week t, and the terms $\hat{p}_{i,t}^{n}$ represent the lengths of individual sessions.

Regression models for purchases. Finally, we consider a set of models that predict purchase behavior. For both cohorts, we have access to the total number of products (not necessarily from the same franchise as the focus game) purchased by each user throughout the duration of the study. We estimate Poisson regression models [McCullagh and Nelder, 1989] that predict this quantity based on the following user characteristics: total play time (during the entire study); total score earned; average growth rate of score; social behavior (tiers 1-3); total number of friends accumulated by the end of the study; the fractions of the user's total rounds played with casual vs. close friends; whether or not the user bought a premium account for the focus game; whether or not the user owned the previous title in the series; the number of products owned by the user, as well as the total money spent, at the start of the study. In these models, a user represents a single data point, and we do not consider the evolution of the user's behavior over time, since, in the case of cohort 2, we do not know when the users purchased the products.

3.4.3 Model robustness and endogeneity concerns

Because the scope of this essay is restricted to the observational dataset described in Section 3.3, it is not possible to "provably" alleviate all endogeneity concerns. Nonetheless, we took a number of measures to improve the robustness of our results. To begin with, we reiterate that the models in Section 3.4 can be viewed as robustness checks for each other. Indeed, it will be seen in Section 3.5 that all four models produce consistent results. Second, a major source of potential

bias in our models is heterogeneity between users arising from unobserved factors; however, we have made an effort to counteract this by including the random effect term b_i into (3.1) and other models where possible.

One possible concern with the Tweedie model in (3.3), which we use to predict play time, is the issue of sample selection bias [Heckman, 1979]. In the data, we only observe non-zero play times when users decide to play in a given week. These decisions are not randomly selected (in fact, they are not observable), and those players who decide to play on more occasions may also be inclined to play longer. We thus considered a version of (3.3) that includes a selection correction term modeling the probability of playing in a given week. This term is modeled using the probit model $\rho_{i,t} = 1_{\{\rho^*_{i,t}>0\}}$, where

$$\rho^*_{i,t} = \boldsymbol{x}^\top_{i,t-1}\boldsymbol{\beta} + b_i + \varepsilon_{i,t}. \tag{3.4}$$

In (3.4), $\varepsilon_{i,t}$ is the zero-mean, normally-distributed error term, b_i is the user-level random effect, and $\boldsymbol{x}_{i,t-1}$ includes user-level fixed effects such as console, location, tier 1-3, as well as time-dependent features related to purchase history (for cohort 1 users only) and whether or not the user owns a premium account. The other social features are *not* included in (3.4); instead, they are included in (3.3) together with $\rho_{i,t}$. Essentially, eqs. (3.4) and (3.3) now define a two-stage model where the first stage uses all user characteristics that are not related to short-term social behavior to generate a stand-in for the user's decision to play. We estimated this two-stage model as well as the standard model that only uses (3.3), and found that the results

were very similar in both cases. We believe this to be additional evidence of model robustness.

Finally, we discuss the nature of possible causal relationships between social behavior and user engagement (again with the caveat that it is very difficult to provably establish such a relationship based purely on observational data). Potentially, one could speculate that causality works in either direction. The "forward" direction (and the one of primary interest to us) implies that social behavior causes user engagement, for example because positive social experiences may increase users' enjoyment of the game. The "reverse" direction implies that user engagement causes social behavior, for example because spending more time playing the game may lead the user to build up a social network.

In the literature, one possible method for ruling out reverse causality is the difference estimator of Arellano and Bond [1991], which includes past levels of the variables that are suspected to be endogenous as instruments. This method has been widely used in the study of social media; see, e.g., Qiu et al. [2015]. We also considered this approach; however, Mileva [2007] and Roodman [2009] provide arguments that the Arellano-Bond estimator is not necessary in panel data where the panel size T is not very small relative to the number I of panels. In such settings, the correlation of the lagged dependent variable with the error term becomes less significant and the endogeneity issue will have less of an impact on the results over time. Furthermore, because T is fairly large in our dataset, we also found that estimating the Arellano-Bond model became much more computationally burdensome, and the results of the estimation procedure were not reliable.

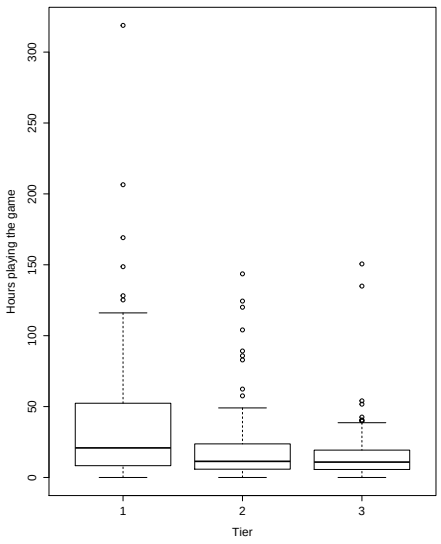
For these reasons, we have chosen to stay with the Tweedie model to predict play time. However, in closing this section, we provide some additional observations regarding the issue of reverse causality. Most of the social features in Section 3.3.3 are time-dependent (e.g., the proportions of rounds played with casual/close friends during a given week) and are meant to reflect short-term social behavior. Such features could hypothetically be influenced by play time (though play time itself is a control in our models). However, the three tier features are treated as user-level *fixed effects*; that is, a tier-3 user is seen as belonging to tier 3 during the entire observation window. We have made this modeling choice because, although the tiers are calculated from the observed data (it takes a certain amount of time for a tier-3 user to play the rounds that will lead us to classify that user into tier 3), we *interpret* them as a measure of users' inherent proclivity toward social behavior. In other words, our view is that some users are simply more social than others; this inclination is reflected in how they play the focal title, but we believe that it was present before they started playing, even though we may not be able to detect it immediately. If this is indeed the case, the effects of tiers 1-3 should not be susceptible to reverse causality because these effects represent behaviors that were in place before the users had any exposure to the focal title.

To provide support for this argument, we calculated the hours that users in tiers 1-3 spent playing in order to meet the conditions for *lower* tiers. Recall from Section 3.3.3 that tier 1 includes those users whose "best friend" has degree between 2-4, while for tier 2 this range is 5-10. However, any tier-2 user will play for some length of time before reaching the tier-2 range, and in order to do so, the user must

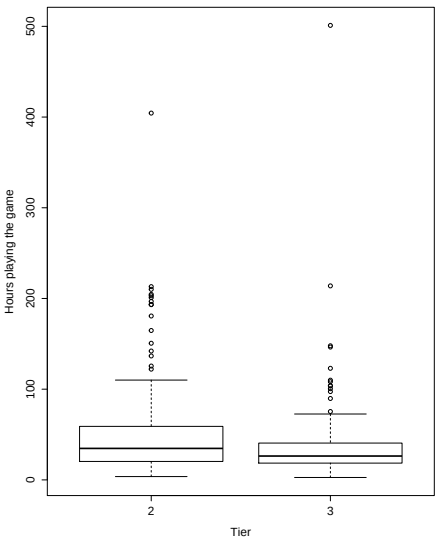
first reach the tier-1 range. We can thus compare the total hours played by users in tiers 1-3 in order to meet the conditions for tier 1 (shown in Figures 3.2(a) and 3.2(c)). Similarly, we can compare the hours played by users in tiers 2-3 in order to meet the conditions for tier 2 (shown in Figures 3.2(b) and 3.2(d)).

If reverse causality were to have a strong effect (i.e., playing more leads to more social behavior), we should expect that higher-tier users take *longer* to reach a given tier. Under reverse causality, such users would require time to build up their social network in the process of playing and make the transition to more social behavioral patterns. However, we see the exact opposite tendency: higher-tier users clearly require *less* time to meet the conditions for lower tiers. This suggests that higher-tier users already have most or all of their social network in place before their experience with the focal title, and so the effect of this existing network on observed play time should be much greater than the other way around. We repeated the analysis of Figure 3.2 with several other measures of time (such as the number of days required to meet the conditions for a tier), but the results were very similar and so we omit them here.

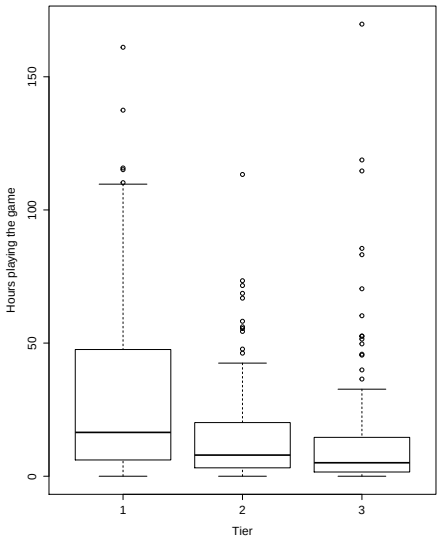
Of course, we cannot conclude from these observations that reverse causality is completely absent. In fact, one can see outliers in Figure 3.2 that do appear to take a long time to reach a lower tier; perhaps these outliers represent those users who did indeed build up their social network over the course of playing the game. However, this phenomenon is clearly the exception rather than the rule: for the vast majority of focal users, their final tier appears to indicate their pre-existing inclination toward social behavior, and thus the effects of the tier 1-3 features in our



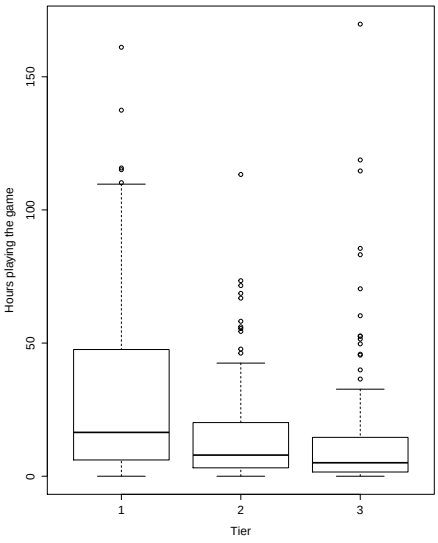
(a) Cohort 1, hrs. to reach Tier 1.



(b) Cohort 1, hrs. to reach Tier 2.



(c) Cohort 2, hrs. to reach Tier 1.



(d) Cohort 2, hrs. to reach Tier 2.

Figure 3.2: Hours spent playing by users in Tiers 1-3 to meet the conditions of lower tiers.

models represent the impact of this inclination on play time, rather than the other way around.

3.5 Results

Section 3.5.1 presents results for the mixed-effect logistic regression model for predicting a user's incidence of play in a given week. Section 3.5.2 presents additional analysis of the other models discussed earlier, covering binging behavior, user churn, and total play time, respectively. Section 3.5.3 discusses user purchase behavior. Finally, Section 3.5.4 discusses the predictive power of the models.

3.5.1 Predicting the incidence of play

We begin with the mixed-effect logistic regression model in (3.1), where the response $y_{i,t}$ represents the incidence of play by user i at any time during week t. Recall that we estimated a separate model for each cohort. Including all of the dummy variables and interaction terms, each model had over 40 features. However, most of the features were not found to be statistically significant for either cohort; due to space considerations, we do not list them here. We also do not report coefficients for features related to the presence of certain DLCs, game modes, maps, or vehicles; while these are important as controls and may be statistically significant, they are less important for our research questions. For instance, we found that the presence of a certain vehicle is negatively correlated with user engagement for cohort 2, but not for cohort 1. Vehicles are quite difficult to control in-game, and it is unsurprising that the more casual users in cohort 2 might not react favourably to them. However, this issue is unrelated to social behavior.

Table 3.1 reports results for relevant features that were found to be significant

Table 3.1: Estimated coefficients for logistic regression model predicting the incidence of play.

	Cohort 1		Cohort 2	
Feature	Coefficient	p-value	Coefficient	p-value
(Intercept)	-2.1579	$<$ 2e-16	-2.2810	$<$ 2e-16
Time since starting game	-0.0252	$<$ 2e-16	-0.0247	$<$ 2e-16
Hours played last week	0.2826	$<$ 2e-16	0.2927	$<$ 2e-16
Skill (score per hour)	0.0014	0.0069	(not significant)	0.8480
Premium account	0.4538	$<$ 2e-16	0.4510	1.37e-15
% rounds played with casual friends	1.1437	0.0001	1.3140	1.36e-07
% rounds played with close friends	1.2667	0.0003	1.6700	1.90e-11
Tier 1 social behavior	0.7537	8.85e-14	0.7646	2.65e-14
Tier 2 social behavior	0.9244	$<$ 2e-16	1.0610	$<$ 2e-16
Tier 3 social behavior	1.6322	$<$ 2e-16	1.5950	$<$ 2e-16
Tier 1 $\times$ % rounds with casual friends	(not significant)	0.2587	-0.6299	0.0299
Tier 2 $\times$ % rounds with casual friends	(not significant)	0.1921	-0.9428	0.0008
Tier 3 $\times$ % rounds with casual friends	-0.8447	0.0069	-0.8533	0.0020
Tier 2 $\times$ % rounds with close friends	(not significant)	0.4496	-0.5296	0.0479
Tier 3 $\times$ % rounds with close friends	(not significant)	0.3427	-0.6475	0.0143
Owned prior titles in the series	0.1297	0.0002	(unavailable)	
Bought next title in the series	-0.9375	$<$ 2e-16	(unavailable)	

at the 0.05 level in at least one of the cohorts. It can be immediately seen that the results, and even the magnitudes of certain coefficients, are quite similar between cohorts. Some of the results are to be expected: for example, the number of weeks elapsed since the user began playing the game has a negative coefficient in both models, implying that engagement decreases as we move closer to the end of the game's life cycle. The effects of the engineered features are more nuanced and require detailed discussion.

One significant difference is immediately evident: the rate at which the user earned points during week $t-1$ (recall that this is our stand-in for the user's current skill level) is strongly significant for cohort 1, but not even approaching significance for cohort 2. This is not a borderline distinction: the p-value is below 0.01 for cohort

1, yet nowhere near 0.05 for cohort 2. As will be shown later, this result is quite consistent throughout all of our models. This answers one of our research questions: early adopters appear to be more engaged by the competitive side of the game, and are more likely to return to it if they perform better. On the other hand, the more casual users in cohort 2 appear to be unaffected (positively or negatively) by their performance.

In marked contrast, social behavior exhibits a strong positive correlation with engagement for both types of users. In both cohorts, the social effect is similar in magnitude, but directionally larger for tier-2 and tier-3 users. Furthermore, both cohorts are always more engaged after playing a greater proportion of their total rounds for the week with friends (both casual and close).

However, different tiers exhibit different *sensitivity* to the proportion of rounds played with friends, as shown by the interaction terms in Table 3.1. Moreover, tiers 2 and 3 (the most social users) are the *least* sensitive to this quantity, although they have higher baseline engagement. Two complementary explanations are possible. First, users who are more social and engaged with the game experience diminishing returns from playing with friends; as they play with more of them, the friends are essentially becoming a commodity. On the other hand, less engaged users with fewer friends overall get more value out of playing with them. Second, our prediction for week t is based on the user's behavior in week $t-1$; users who are more social in general should be less sensitive to the precise proportion played with in just the past week. By contrast, less social users may get much more value out of a single positive social experience. Note also that the sensitivity reduction is generally greater for

casual friends, as one might expect.

The managerial implication suggested by both explanations is that the most immediate improvement in engagement can be obtained by encouraging less social (lower-tiered) users to play socially more often. The marginal gain of social engagement is higher for these users, and they are at greater risk of dropping out to begin with. Even if the user does not move up in tier, playing more rounds with existing friends improves the odds that the player will continue to be engaged in the game. While higher-tiered users also become more engaged after playing socially, the marginal gain is smaller in comparison.

Finally, the last two rows of Table 3.1 relate to purchase information, which is only available for cohort 1. Two features of this type were significant. First, owning prior titles in the series (a measure of brand loyalty) unsurprisingly had a positive effect. Second, owning the next title in the series at time $t-1$ had a negative effect, suggesting that the next title draws committed users away from the current one. For cohort 1, we estimated two different versions of the model, both with and without purchase information; however, for all the features that were common to both models, the estimated coefficients were surprisingly similar (often to the fourth decimal point), and so we do not report both models here.

3.5.2 User engagement: other model variants

We now present results for the models from Section 3.4.2. As before, for cohort 1 we considered versions of the same model with and without purchase information,

Table 3.2: Estimated coefficients for logistic regression model predicting the incidence of binging.

	Cohort 1		Cohort 2	
Feature	Coefficient	p-value	Coefficient	p-value
(Intercept)	-3.9066	$<$ 2e-16	-4.0080	$<$ 2e-16
Time since starting game	-0.0178	$<$ 2e-16	-0.0184	0.0059
Hours played last week	0.1296	$<$ 2e-16	0.1379	$<$ 2e-16
Skill (score per hour)	0.0007	0.0393	(not significant)	0.4906
Premium account	0.3837	2.34e-11	(not significant)	0.5897
No. of close friends played with	0.0441	0.0004	(not significant)	0.7739
% rounds played with casual friends	(not significant)	0.1156	1.5840	2.03e-06
% rounds played with close friends	1.6375	2.55e-06	1.8810	5.94e-08
Tier 1 social behavior	0.3827	8.31e-05	0.3948	0.0002
Tier 2 social behavior	0.3169	0.0015	0.5077	2.77e-06
Tier 3 social behavior	0.4197	2.82e-05	0.4741	7.27e-05
Tier 1 $\times$ No. of close friends played with	-0.0352	0.0068	(not significant)	0.4571
Tier 2 $\times$ No. of close friends played with	-0.0424	0.0007	(not significant)	0.7666
Tier 3 $\times$ No. of close friends played with	-0.0434	0.0005	(not significant)	0.7304
Tier 2 $\times$ % rounds played with casual friends	(not significant)	0.7880	-1.1990	0.0015
Tier 3 $\times$ % rounds played with casual friends	(not significant)	0.2456	-1.0880	0.0030
Tier 1 $\times$ % rounds played with close friends	-0.9658	0.0124	(not significant)	0.1332
Tier 3 $\times$ % rounds played with close friends	-0.8348	0.0211	(not significant)	0.0887
Bought next title in the series	-0.4288	0.0409	(unavailable)	

but the estimated coefficients were so similar for both versions that we only report one set of results. This was the case across all the types of models discussed in Section 3.4.2.

Binging model. Table 3.2 reports results for relevant features that were found to be significantly correlated with the incidence of binging. The overall structure of these results resembles that of Table 3.1: being more social provides a slight baseline increase to the probability of binging (however, the effects from belonging to tier 1-3 are much weaker than in the previous model and the differences between tiers are small), while short-term expressions of social behavior (playing with more friends, or playing more rounds with them) have a positive effect that is reduced for more

social users.

It is interesting to note that this model reveals much more pronounced differences between cohorts 1 and 2 than the previous model. Cohort 1 users appear to derive no benefit at all from interactions with casual friends with regard to their binging behavior. They do see such a benefit from interacting with close friends; however, even this benefit is considerably reduced for more social users. By contrast, cohort 2 users clearly see a benefit from interacting with casual friends (reduced for higher-tier users, as expected), and they derive a considerable benefit from playing with close friends that *does not* appear to diminish as the user's tier increases. In other words, for late adopters that play the game more casually to begin with, short-term social experiences with close friends are more likely to lead to unusually intense short-term engagement, whereas early adopters are more motivated by factors such as skill.

Survival model. Table 3.3 presents results for the Cox survival model from (3.2). All the versions of this model that we considered had very few statistically significant features. Almost all game modes, vehicles, DLCs and maps appeared to have no significant impact on user churn. User skill and ownership of premium accounts did not appear to help retain users, although they were shown to have a positive impact on short-term engagement in the previous models. On the other hand, social behavior was consistently identified as a significant factor. Note that, in this model, positive coefficients imply a higher likelihood of dropping out, while negative coefficients imply the opposite. Thus, for cohort 1 users, a larger number of cumulative purchases is linked to higher churn for this particular title, indicating

Table 3.3: Estimated coefficients for survival model.

	Cohort 1		Cohort 2	
Feature	Coefficient	p-value	Coefficient	p-value
Hours played last week	0.0376	0.0240	0.0801	3.2e-05
Tier 1 social behavior	-0.8381	2.9e-05	-0.7306	1.5e-05
Tier 2 social behavior	-1.1082	8.7e-08	-1.1304	1.7e-10
Tier 3 social behavior	-1.5327	2.6e-13	-1.5480	2.6e-14
Tier 1 × % rounds with casual friends	-1.9559	0.0220	(not significant)	0.64
Tier 3 × % rounds with casual friends	-1.8647	0.0200	(not significant)	0.21
Tier 1 × % rounds with close friends	2.1554	0.0230	(not significant)	0.22
Tier 2 × No. of casual friends played with	(not significant)	0.19	-0.0844	0.0087
Tier 3 × No. of casual friends played with	(not significant)	0.19	-0.0667	0.0020
Tier 1 × No. of close friends played with	-0.0957	0.0480	(not significant)	0.94
Cumulative no. of purchases to date	0.0354	0.0049	(unavailable)	

a degree of cannibalization between games.

Table 3.3 suggests several insights. First, users appear to stay in the game longer if they demonstrate more social behavior in general (i.e., if they belong to higher tiers). However, longer retention is linked to more *casual* behavior: both cohorts experience reduced dropout rates from playing more with casual friends (measured either by the number of such friends or by the proportion of rounds played with them). By contrast, playing more with close friends does not contribute a positive effect anywhere. In fact, for one group of users (tier-1 users in cohort 1), playing with close friends is correlated with *higher* dropout rates. To explain this behavior, consider a user who spends most of his/her in-game time playing with a small group of close friends. Such a user may be more susceptible to trends; when the user's friends drop out and move to a different game, the user is more likely to follow them, since the user's enjoyment of the game is primarily derived from playing it socially. Users may be *more* vulnerable to such behavior if they are early

Table 3.4: Estimated coefficients for Tweedie model.

	Cohort 1		Cohort 2	
Feature	Coefficient	p-value	Coefficient	p-value
(Intercept)	-2.0874	< 2e-16	-2.2960	< 2e-16
Time since starting game	-0.0159	< 2e-16	-0.0164	< 2e-16
Hours played last week	0.0849	< 2e-16	0.0831	< 2e-16
Skill (score per hour)	0.0010	2.28e-07	(not significant)	0.1924
Premium account	0.3971	< 2e-16	0.3485	< 2e-16
No. of close friends played with	0.0414	0.0324	(not significant)	0.4299
% rounds played with casual friends	1.4122	3.74e-05	1.9330	2.07e-12
% rounds played with close friends	2.0906	5.23e-11	2.3540	< 2e-16
Tier 1 social behavior	0.8845	< 2e-16	0.8876	< 2e-16
Tier 2 social behavior	0.9786	< 2e-16	1.0520	< 2e-16
Tier 3 social behavior	1.4428	< 2e-16	1.5600	< 2e-16
Tier 1 × % rounds with casual friends	(not significant)	0.2402	-0.8017	0.0113
Tier 2 × % rounds with casual friends	(not significant)	0.0950	-1.0510	0.0005
Tier 3 × % rounds with casual friends	-0.8426	0.0165	-0.8671	0.0029
Tier 1 × % rounds with close friends	-1.1399	0.0010	-0.9536	0.0016
Tier 2 × % rounds with close friends	-1.1095	0.0009	-1.1220	0.0001
Tier 3 × % rounds with close friends	-1.3284	4.52e-05	-1.5500	2.89e-08
Tier 2 × No. of close friends played with	-0.0396	0.0409	(not significant)	0.4269
Tier 3 × No. of close friends played with	-0.0396	0.0406	(not significant)	0.4768
Owned prior titles in the series	0.0478	0.0037	(unavailable)	
Bought next title in the series	-0.6138	2.13e-08	(unavailable)	
Number of purchases made	-0.0202	5.73e-10	(unavailable)	

adopters (cohort 1), and if they are themselves relatively casual (tier-1). However, if the user has more close friends, losing a few of them to a different game will not be as important (though there will also be no positive effect from playing with them), moderating the churn effect.

Tweedie model. Recall that the Tweedie model from (3.3) aims to predict total weekly play time, rather than simply the incidence of play or binging. The results are given in Table 3.4, and have the standard interpretation that positive coefficients impact play time positively, while the opposite is true for negative coefficients. We considered both this model as well as the more complex two-stage

model described in Section 3.4.3 and found that the results were very similar; due to space considerations, we only report the results of the simpler model here.

Overall the results are consistent with those in Table 3.1: 1) belonging to a higher tier and owning a premium account contributes positively to user engagement; 2) user skill is statistically significant for cohort 1, but not cohort 2; 3) playing more with close friends (as a number or as a proportion of rounds played) is positively correlated with total play time, but much of this effect is canceled out for more social users. Higher-tier users tend to be less sensitive to the effect of playing with both types of friends.

Finally, the purchase information clearly shows the effects of both brand loyalty and cannibalization: owning previous titles in the same series implies that the user is more likely to stay engaged with the newer one, but by the same token, buying the next title in the series makes the user more likely to abandon the current one. Owning more games overall also takes away from the user's engagement in the focus game.

3.5.3 Discussion of purchase behavior

As discussed in Section 3.4.2, we also ran Poisson regression models to predict the number of titles purchased by users during the time frame of the study, as well as whether or not they purchased the next title in the series. Recall that, in these models, there is one data point per user, as for cohort 2 we do not have access to detailed information regarding when the titles were purchased. We thus aggregate

Table 3.5: Estimated coefficients for Poisson model.

	Cohort 1		Cohort 2	
Feature	Coefficient	p-value	Coefficient	p-value
(Intercept)	1.2871	< 2e-16	1.1882	< 2e-16
Total time played	(not significant)	0.4160	-3.2e-7	0.0003
Total score	(not significant)	0.6552	2.1e-8	0.0092
Avg. skill (score per hour)	(not significant)	0.1041	0.0593	2.8e-6
Premium account	0.3315	< 2e-16	0.3189	5.8e-11
Total no. of friends	(not significant)	0.4700	0.0022	0.0002
% rounds played with casual friends	-1.3240	0.0274	(not significant)	0.1461
% rounds played with close friends	1.2890	0.0351	(not significant)	0.9569
Tier 3 social behavior	-0.8893	3.1e-5	(not significant)	0.4131
Tier 2 × % rounds with close friends	1.2921	0.0437	(not significant)	0.5145
Tier 3 × % rounds with casual friends	1.2884	0.0417	1.4880	0.0082
Tier 1 × No. of friends	(not significant)	0.5016	-0.0020	0.0008
Tier 2 × No. of friends	(not significant)	0.4177	-0.0021	0.0003
Tier 3 × No. of friends	(not significant)	0.5007	-0.0022	0.0002
No. products owned at start	0.0468	< 2e-16	0.0470	< 2e-16
Money spent at start	0.0068	2.5e-7	(not significant)	0.1433
Bought next title in the series	0.3833	< 2e-16	(unavailable)	

as many of the user characteristics as possible across weeks; for example, we consider the fraction of the user's total rounds played with casual vs. close friends (as opposed to last week's rounds).

Table 3.5 reports the key results. Unsurprisingly, the volume of purchases made after the beginning of the study is consistently correlated with the volume of prior purchases (likely a reflection of the user's purchasing power), as well as with the ownership of a premium account or the next title in the series (possibly a brand loyalty effect).

The two cohorts exhibit important differences with regard to social behavior. For Cohort 1 (early adopters), purchases are significantly correlated with various expressions of social behavior. Notably, greater dependence on close friends exhibits

a strong positive correlation, whereas a negative correlation is seen for dependence on casual friends. This complements the results of the survival model in Table 3.3, where the signs were reversed: dependence on close friends was positively linked with the chance of dropping out of the focus game. Thus, Cohort 1 appears to be subject to a tradeoff between engagement in the focus game and adoption of other games; early adopters who play with a group of close friends are susceptible to trends, and can more easily move on to adopt other games.

For Cohort 2 (late adopters), we see almost no correlation between social behavior and purchases. There is a small positive effect for having more friends, but it is largely attenuated at the higher tiers. On the other hand, while the user's in-game performance (expressed by score and skill level) did not necessarily translate to greater engagement in the focus game (Table 3.1), it exhibits a strong positive correlation with purchases, suggesting that it may have more wide-ranging impact on the user's overall interest in gaming.

Overall, these results suggest that social behavior does impact sales, but that this impact is primarily confined to early adopters, and may come at the expense of their engagement in other titles. For late adopters, the best opportunity to increase sales may be to improve their satisfaction with games that they are currently playing, by reinforcing their perception of their performance. In this case the most effective improvements would be related to in-game features rather than social ones.

3.5.4 Predictive power of models

We now discuss the predictive power of the models we have developed. This is a separate issue from the interpretation of the coefficients, and can help to evaluate the magnitude of the impact of social behavior on player engagement. For the service provider, these results address the question of whether the user population is predictable, as well as the possible utility of the models in identifying engaged users.

As an illustrative example, we consider the logistic regression model from Section 3.4.1 for predicting the incidence of play. For both cohorts, we use 5-fold cross validation to calculate two widely used performance metrics for predictive models with binary data, namely the AUC, or area under the ROC curve [Smithson and Merkle, 2013], and the lift statistic [Seppänen et al., 2003]. The results are shown in Figure 3.3.

Figures 3.3(a) and 3.3(c) suggest that the predictive power of the model is quite high: the AUC, averaged over the five folds of cross-validation, is 0.908 for cohort 1 and 0.893 for cohort 2. The ROC curves exhibit little variation between folds. The binging model from Section 3.4.2 produces similar results (0.876 for cohort 1, and 0.870 for cohort 2). Moreover, Figures 3.3(b) and 3.3(d) suggest that the models would be effective in identifying the users most likely to be engaged in a given week. For example, suppose that service provider wished to identify the 10% of users with the highest chance of logging on sometime during the next week, perhaps in order to offer them promotions. Then, our models would find this group

3 to 4 times more effectively than would a random sample of 10% of the population.

3.6 Conclusion

We have presented an empirical study of high-resolution player-level data, collected across roughly two years by a major international video game company for one of its flagship first-person shooter titles. From these data, we have engineered features describing various forms of social behavior within that title, and obtained the following insights into the four hypotheses stated in Section 3.2:

1. *User engagement and retention.* We find that social behavior is positively correlated with user engagement across both cohorts. However, this type of engagement is short-term in nature and does not always translate to better retention rates. In fact, for early adopters (cohort 1), greater dependence on close friends is linked to greater churn. More casual play is linked to better retention as well as more intense engagement (binging).

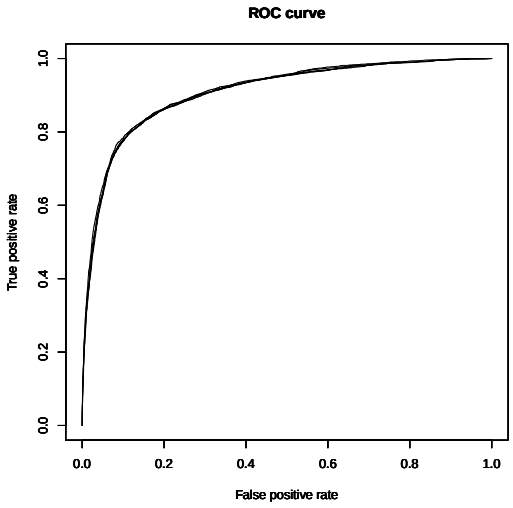
2. *Purchases.* The most consistent predictors of purchases are past purchases. Early adopters do appear to be somewhat more likely to convert their social experience into money spent, but this comes at the cost of their engagement in the focal title. However, identifying such users may be helpful if the company aims to improve engagement in new titles. Later adopters (cohort 2) exhibit a greater link between purchases and past in-game performance.

3. *Early vs. late adopters.* Building on the previous point, we find significant heterogeneities between early and late adopters, which may help to explain

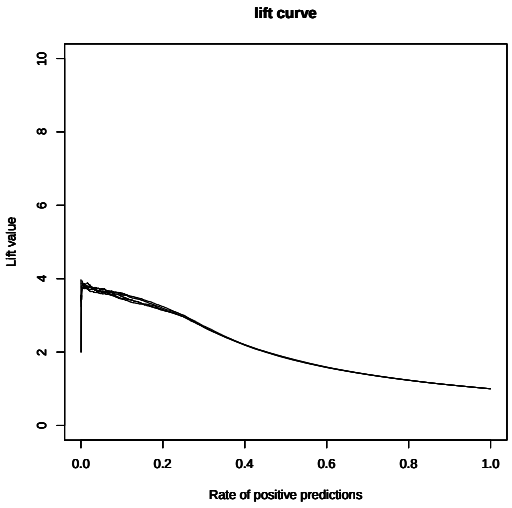
some of the mixed findings in the literature regarding the conversion of engagement to revenue. The two cohorts react differently, not only to the social experience, but also to the competitive side of the game (skill).

4. *Predictability.* The accuracy displayed in Figure 3.3 shows that recent in-game behavior has considerable predictive power for future engagement. The company can reliably use these data to identify users for targeted interventions (e.g., discounts or email offers) to improve retention or, conversely, to increase engagement in newer titles.

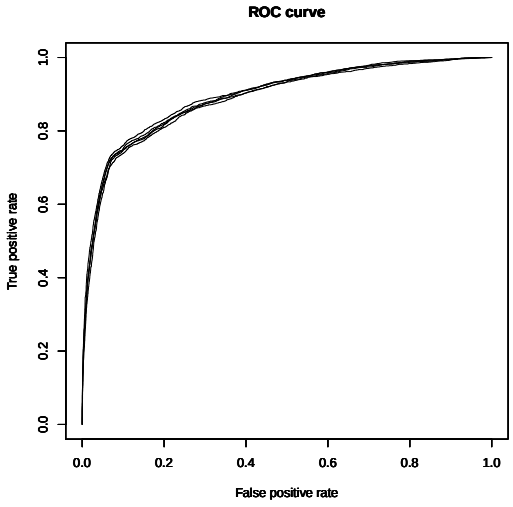
One surprising managerial implication emerging from our research is that playing with casual friends appears to be better for both short- and long-term engagement. This adds an interesting nuance to the service literature on the importance of creating "emotionally engaged" users who are enthusiastic fans of the product [Pine and Gilmore, 1999, Vivek et al., 2012]. Our results show that in video games there are different types of engagement, and greater variety of play may be more valuable to the company than close-knit social behavior. One potential explanation is that greater casual play creates a wider network with less intense links that is more robust to shocks, since the social network can "absorb" a person leaving. By contrast, users who engage mostly with close friends often have a small but very densely connected social network that is more susceptible to sudden changes. Rather than relying on social behavior, a better strategy for game companies may thus be to pursue efforts to create broader communities and promote involvement at all levels of play.



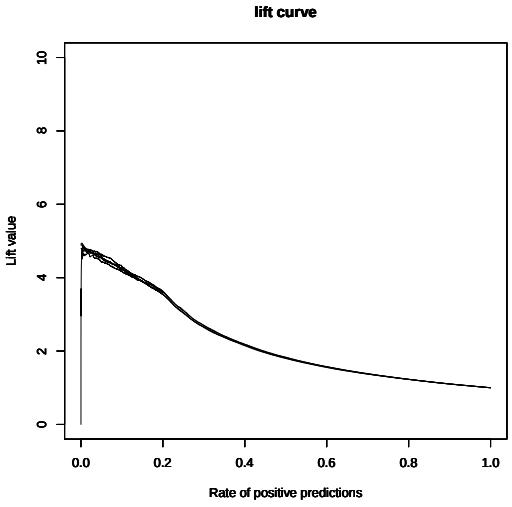
(a) ROC curves, cohort 1.



(b) Lift curves, cohort 1.



(c) ROC curves, cohort 2.



(d) Lift curves, cohort 2.

Figure 3.3: Predictive power of logistic regression model predicting the incidence of play.

Appendix A: Simulation Results for Country 2

In this section, we investigate the generalizability of our results by recalibrating our models from Sections 2.3.2-2.3.5 on data from a second developing country (referred to as Country 2 or C2) where sufficiently good fuel, repair, and maintenance data were available. The fleet in C2 contains multiple vehicle types, with the most data available for types 2 and 3; Type 3 is the same vehicle type used in our analysis of FC (strong transporter, e.g., Land Cruiser or Pajero), while Type 2 represents light vans/minibuses. Thus, we are able to perform a double robustness check: first, we can run new simulations for C2, and second, we can also test whether our conclusions continue to hold on Type 2 vehicles. Unfortunately, C2 does not have enough data to allow detailed demand estimation for more than one mission type (namely Type 2), so we do not study mission type switching (Case 4) in this discussion. However, Cases 1-3 can all be considered.

A.1 Estimation results for cost models

We begin by estimating the depreciation model from (2.3.3) on C2 data. This estimation involves 235 vehicles, of which 7 have zero residual value and are treated as right-censored data. There are 4 mission types and 4 vehicle types in the cross-

sectional data for C2, so the model is

$$\text{Dep\%} = \beta_1 \text{Age} + \sum_{i,j} \beta_{2ij} \log(\text{Odometer}) \times \text{MissionType}_i \times \text{VehicleType}_j + \beta_3 \text{NumAcc},$$

and the results are summarized in Table A.1. The coefficients are overall similar to what we saw in Section 2.3.3, except that age has a stronger positive correlation with depreciation. There is significant heterogeneity in the impact of odometer depending on the vehicle and mission type involved.

	Dependent variable:
	Depreciation Percentage
Age	0.001^{***}
Accident	0.030^{***}
log (Odometer) × Mission1 × VehicleType4	0.044^{***}
log (Odometer) × Mission2 × VehicleType1	0.016^{***}
log (Odometer) × Mission2 × VehicleType2	0.050^{***}
log (Odometer) × Mission2 × VehicleType3	0.032^{***}
log (Odometer) × Mission3 × VehicleType4	0.049^{***}
log (Odometer) × Mission4 × VehicleType1	0.041^{***}
log (Odometer) × Mission4 × VehicleType2	0.061^{***}
log (Odometer) × Mission4 × VehicleType3	0.055^{***}
Observations	235
Log Likelihood	−1,684.091
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.1: Residual value estimation results for C2.

Next, we estimate the fuel cost model using C2 data. We use the same model as in Section 2.3.4, but add dummy variables representing vehicle types since our

previous analysis focused on a single type. The model thus becomes

$$\text{FuelCostPerKm} = \beta_0 + \beta_1 \text{Odometer} + \beta_2 \text{Age} + \sum_j \beta_{3j} \text{MissionType}_j + \sum_k \beta_{4k} \text{VehicleType}_k.$$

Table A.2 shows the results. As in Section 2.3.4, age and odometer are positively correlated with fuel cost.

	Dependent variable:
	Cost per Km
Odometer	3.194×10^{-8} *
Age	5.930×10^{-5} *
Mission2	-0.1284***
Mission4	-0.1268***
VehicleType2	0.0084*
VehicleType3	0.0529***
Constant	0.1955***
Observations	4746
R^2	0.5507
Adjusted R^2	0.5428
Residual Std. Error	0.03132 (df = 4663)
F Statistic	68.87*** (df = 83; 4663)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.2: Fuel cost estimation results for C2.

Next, we estimate the maintenance cost model. We found that vehicle types do not appear to have a significant correlation with maintenance/repair costs. Therefore, we assume that vehicles working on the same mission type follow similar main-

tenance/repair schedules, and the model remains unchanged from Section 2.3.5:

$$\text{CC} = \sum_j \beta_{1j}\text{Age}\times\text{MissionType}_j + \beta_{2j}\text{Odometer}\times\text{MissionType}_j + \beta_{3j}\text{Odometer}^2\times\text{MissionType}_j.$$

The results are shown in Table A.1. Since C2 only has enough data on mission type 2, the MissionType component of the interaction terms is omitted; we see, however, that the coefficients are overall similar to what we saw in Section 2.3.5, with the main difference being that the nonlinear effect of odometer is less pronounced. Thus, if enough data can be gathered to perform estimation reliably, one can recalibrate the cost module of the simulator. If, additionally, monthly mileage data are available, one can calibrate the demand module and repeat the analysis of Section 2.4.

	Dependent variable:
	CumuCost
Age	34.638^{***}
Odometer	0.021^{***}
Odometer^2	1.165×10^{-7} ***
Observations	520
Log Likelihood	−4,074.901
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.3: Maintenance/repair cost estimation results for C2.

A.2 Simulation results for C2, vehicle type 2

We repeated Cases 1-3 from our study of FC on the C2 data with mission type 2 and vehicle type 2. First, let us consider Case 1 (comparison of sales policies), in

which the demand is stable, and five demand levels are considered with an initial fleet size of four vehicles. In other words, the experimental setup is the same, but costs are now calculated according to the C2 models. First, we summarize the results of the base case (no accidents) in Table A.4. The overall pattern is similar to what we observed in the FC study, except that we now exploit vehicles for slightly longer: the sales threshold is held at RIG for demand level 3, and the Myopic policy is preferred at level 2. The impact of odometer on depreciation is sharper for C2 than FC (cf. the coefficients for mission type 2 in A.1), while the spike in maintenance cost for vehicles late in their lifetime is less pronounced, so for demand level 2, it is more economically beneficial to exploit part of the existing fleet more heavily while leaving the rest idle. At the same time, the overall pattern is the same as what we observed for FC: the combination of Myopic and RIG is optimal for lower demand levels, and as the demand increases, we eventually switch to 140/300 and then 100/225 while attempting to balance the load on the fleet.

Demand level	Optimal policies	Cost
1	Myopic, RIG	85127
2	Myopic, RIG	166662
3	Least Odometer , RIG	274870
4	Balance, 140/300k	439609
5	Balance, 100/225k	553761

Table A.4: Optimal assignment/sales policies under 5 stable demand levels.

Next, Table A.5 compares optimal sales and assignment policies for varying accident frequencies, as was done in our study of FC. The results are very similar

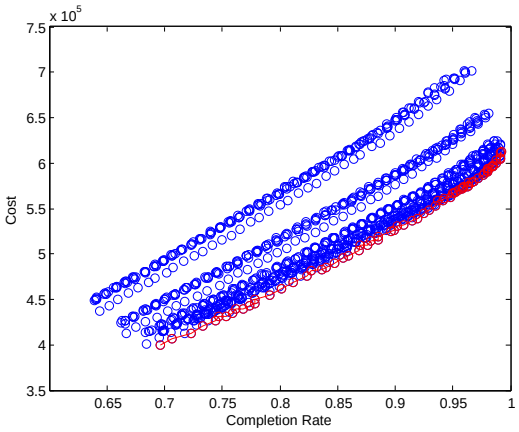
to FC and consistent with the base case: Myopic and RIG are optimal for lower demand, while 140/300 and 100/225 (together with load balancing) are preferred when the demand is high.

Accident Level	Demand Level	Policies	Cost	NumOfAccidents
0	1	Myopic, RIG	85127	0
	2	Myopic, RIG	166662	0
	3	Least Odometer, RIG	274870	0
	4	Balance, 140/300k	439609	0
	5	Balance, 100/225k	553761	0
Baseline	1	Myopic, RIG	86726	1.8
	2	Myopic, RIG	170723	4.54
	3	Least Odometer, RIG	282081	7.87
	4	Balance, 140/300k	445854	7.03
	5	Balance, 100/225k	559037	5.94
2x Baseline	1	Myopic, RIG	88352	3.63
	2	Myopic, RIG	177779	9.38
	3	Least Odometer, RIG	294662	15.57
	4	Balance, 140/300k	452796	14.85
	5	Balance, 100/225k	564491	12.08
4x Baseline	1	Myopic, RIG	91649	7.07
	2	Least Odometer, RIG	187410	15.00
	3	Least Odometer, RIG	318685	25.22
	4	Least Odometer, 140/300k	465182	27.87
	5	Balance, 100/225k	576421	25.51

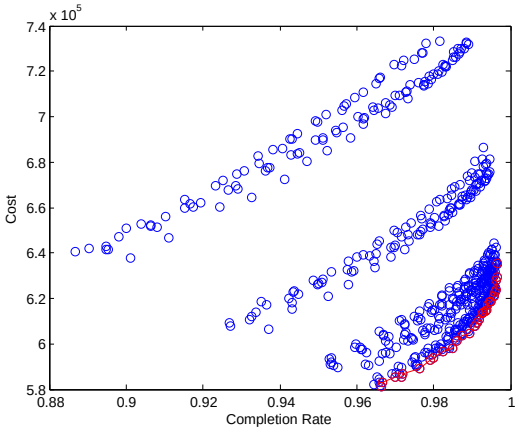
Table A.5: Comparison between different levels of accidents.

Next, we consider Case 2, in which realistic demand data (in this case, for mission type 2) is used to evaluate purchase, assignment and sales policies jointly.

The same reactive purchase policies from 2.4.3 were implemented. The Pareto fronts for both policies are shown in Figure A.1. Similar conclusions hold: for COMP, 55 out of 72 points on the Pareto front use the 100/225 sales policy, while 54 out of 72 use the Balance policy; for UTIL, 12 out of 27 Pareto-optimal points use the 140/300 sales policy while 15 out of 27 use RIG, and the optimal assignment policy for these points is either Least Odometer or Myopic. This is consistent with Case 1, where we saw an inclination toward Myopic and RIG in scenarios with smaller utilization.



(a) COMP policy: $\alpha = \{0.6, 0.61, ..., 0.99\}$.



(b) UTIL policy: $\beta = \{0.8, 0.81, ..., 0.99\}$.

Figure A.1: Cost and completion rate comparison of reactive purchase policies in C2.

Finally, we consider Case 3, which compares centralized procurement with a 6-month lead time vs. decentralized procurement with no lead time but a 50% markup in purchase cost. Table A.6 reports the results of the comparison for stable demand, assuming that the fleet manager has rough knowledge of the demand over the next six months. As for FC, we see that centralized procurement is better for all five demand levels; however, the savings are less pronounced when the demand is high. This is due to the fact that the nonlinear behaviour of the maintenance cost (which causes steep increases late in vehicles' lifetimes) is less pronounced in C2, so it is generally less prohibitive to keep vehicles longer. For this reason, we make fewer purchases under decentralized procurement and so the difference between purchase prices is of less importance. The general observation that centralized procurement is preferable, given the ability to order anticipatively, is still valid.

Demand level	Purchase locally (no lead time, 50% markup)	Order predictively (6-month lead time)	% Cost saved from predictive ordering
1	97413	85127	12.6%
2	185836	166662	10.3%
3	307215	274870	10.5%
4	479895	438445	8.64%
5	601863	550249	8.58%

Table A.6: Comparison of costs between purchasing locally and ordering predictively for stable demand.

Figure A.2 compares the performance of centralized vs. decentralized procurement under realistic demand for C2. The same conclusions apply.

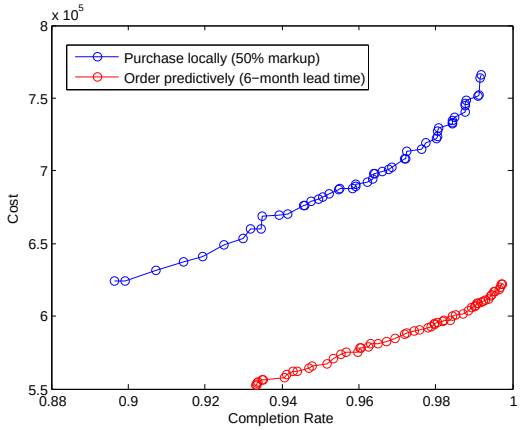


Figure A.2: Comparison of centralized and decentralized purchases.

A.3 Simulation results for C2, vehicle type 3

Finally, we repeat the previous simulations for mission type 2 and vehicle type 3, i.e., we consider a different vehicle type. The cost models are unchanged from Section A.1 as they already considered both vehicle types. Table A.7 reports the results of Case 1; the conclusions are very similar to those in Section A.2.

Demand level	Optimal policies	Cost
1	Myopic, RIG	94665
2	Least Odometer, RIG	184692
3	Least Odometer, RIG	313184
4	Balance, 140/300k	486114
5	Balance, 100/225k	594993

Table A.7: Optimal assignment/sales policies under 5 stable demand levels.

Next, Table A.8 compares optimal sales and assignment policies for varying accident frequencies, with results that are consistent with both FC and the other vehicle type from C2. Myopic and RIG are optimal for lower demand, while 140/300

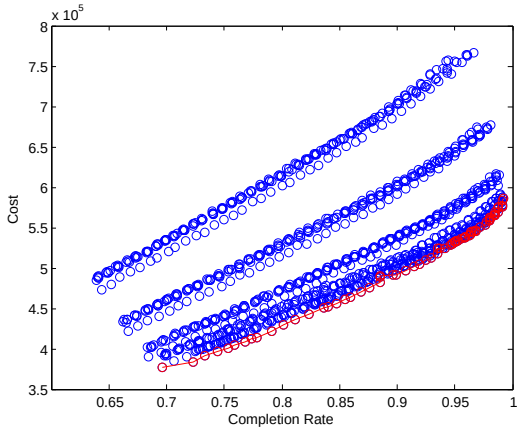
and 100/225 (together with load balancing) are preferred when the demand is high.

Accident Level	Demand Level	Policies	Cost	NumOfAccidents
0	1	Myopic, RIG	94665	0
	2	Least Odometer, RIG	184692	0
	3	Least Odometer, RIG	313184	0
	4	Balance, 140/300k	486114	0
	5	Balance, 100/225k	594993	0
Baseline	1	Myopic, RIG	96288	1.75
	2	Least Odometer, RIG	188548	3.99
	3	Least Odometer, RIG	320058	7.72
	4	Balance, 140/300k	491870	7.00
	5	Balance, 100/225k	600500	6.31
2x Baseline	1	Myopic, RIG	97881	3.54
	2	Least Odometer, RIG	192025	7.90
	3	Least Odometer, RIG	326912	15.73
	4	Balance, 140/300k	498305	14.25
	5	Balance, 100/225k	605989	12.49
4x Baseline	1	Myopic, RIG	101017	7.07
	2	Least Odometer, RIG	199095	15.86
	3	Least Odometer, 140/300k	338646	15.77
	4	Balance, 140/300k	510633	28.13
	5	Balance, 100/225k	616558	24.39

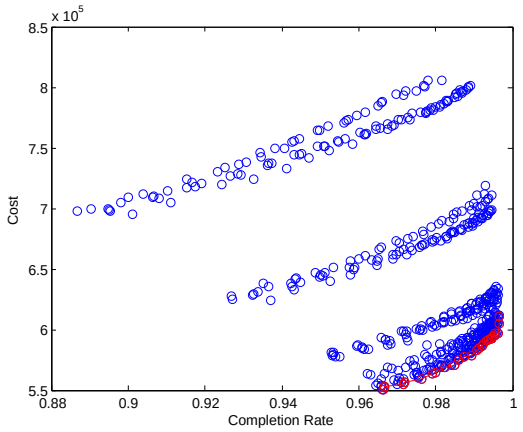
Table A.8: Comparison between different levels of accidents.

In Case 2, A.3 shows that, for COMP, 47 out of 74 Pareto-optimal points use the 100/225 sales policy and 49 out of 74 use the Balance policy; for UTIL, 16 out of 31 Pareto-optimal points use 140/300 and the other 15 use RIG, while 25 out of 31 Pareto-optimal points use the Least Odometer assignment policy. These results

are quite consistent with Section A.2 as well as with the main study of FC.



(a) COMP policy: $\alpha = \{0.6, 0.61, ..., 0.99\}$.



(b) UTIL policy: $\beta = \{0.8, 0.81, ..., 0.99\}$.

Figure A.3: Cost and completion rate comparison of reactive purchase policies in C2.

Table A.9 reports the results of Case 3 for stable demand, while Figure A.4 does the same for realistic demand. These results are quite consistent with Section A.2 as well as with the main study of FC.

Demand level	Purchase locally (no lead time, 50% markup)	Order predictively (6-month lead time)	% Cost saved from predictive ordering
1	106202	94665	10.9%
2	204337	184692	9.6%
3	337549	313184	7.2%
4	533935	484781	9.2%
5	665836	591239	11.2%

Table A.9: Comparison of costs between purchasing locally and ordering predictively for stable demand.

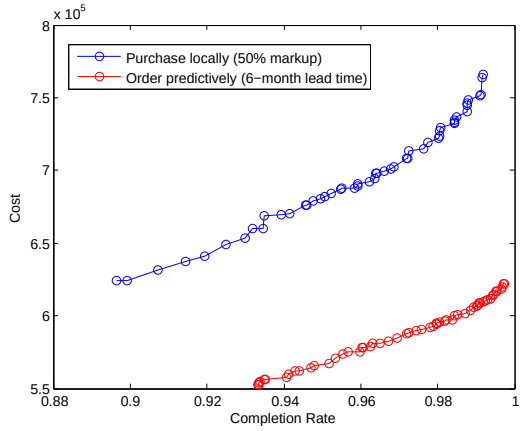


Figure A.4: Comparison of centralized and decentralized purchases.

A.4 Discussion

Overall, the results in Sections A.2-A.3 are similar both to each other and to our earlier results for FC. When vehicle type 2 is used (Section A.2, the results behave as if the demand were "lower," i.e., the patterns that we typically see for demand level 1 persist for demand level 2, and the transition from Myopic to Least Odometer (or from RIG to 140/300) takes place later than for FC. In general, C2 imposes less severe penalties on operational costs for vehicles late in their lifetimes,

but overall vehicles depreciate faster with age; both factors contribute to our tendency to keep type-2 vehicles idle (if possible) or run them into the ground. This tendency is more pronounced for type-2 vehicles since they depreciate even faster than do type-3 vehicles. Thus, the results for type-3 vehicles tend to be in between the results for type 2 and the original results for FC.

These differences illustrate the ability of our simulator to fine-tune its recommendations to the realities of each country and local office. However, the big-picture tendencies observed in this study do not differ from those obtained for FC in any major way.

Bibliography

J. Acimovic and J. Goentzel. Models and metrics to assess humanitarian response capacity. *Journal of Operations Management*, 45:11–29, 2016.

Frank Acito and Vijay Khatri. Business analytics: Why now and what next?, 2014.

N. Altay and R. Pal. Information diffusion among agents: implications for humanitarian operations. *Production and Operations Management*, 23(6):1015–1027, 2014.

R. Y. Arakji and K. R. Lang. Digital consumer networks and producer-consumer collaboration: innovation and product development in the video game industry. *Journal of Management Information Systems*, 24(2):195–219, 2007.

M. Arellano and S. Bond. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297, 1991.

B. Balcik and B. M. Beamon. Facility location in humanitarian relief. *International Journal of Logistics: Research and Applications*, 11(2):101–121, 2008.

B. Balcik, B. M. Beamon, and K. Smilowitz. Last mile distribution in humanitarian relief. *Journal of Intelligent Transportation Systems*, 12(2):51–63, 2008.

Burcu Balcik, Benita M Beamon, Caroline C Krejci, Kyle M Muramatsu, and Magaly Ramirez. Coordination in humanitarian relief chains: Practices, challenges and opportunities. *International Journal of Production Economics*, 126(1):22–34, 2010.

Ravi Bapna and Akhmed Umyarov. Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Management Science*, 61(8):1902–1920, 2015.

Tuncay Bayrak. A review of business analytics: a business enabler or another passing fad. *Procedia-Social and Behavioral Sciences*, 195:230–239, 2015.

B. M. Beamon and S. A. Kotleba. Inventory management support systems for emergency humanitarian relief operations in South Sudan. *The International Journal of Logistics Management*, 17(2):187–212, 2006.

M. Besiou, A. J. Pedraza-Martinez, and L. N. Van Wassenhove. Vehicle Supply Chains in Humanitarian Operations: Decentralization, Operational Mix, and Earmarked Funding. *Production and Operations Management*, 23(11):1950–1965, 2014.

S. Bhattacharya, S. Hasija, and L. N. Van Wassenhove. Designing efficient infrastructural investment and asset transfer mechanisms in humanitarian supply chains. *Production and Operations Management*, 23(9):1511–1521, 2014.

Bloomberg News. Is it game over for video game consoles? 2013. URL http : / / www . bloomberg . com / bw / articles / 2013-11-14 / 2014-outlook-video-game-consoles-game-over-or-new-era. Accessed: 2015-07-14.

R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

Susan A Brown and Viswanath Venkatesh. Bringing non-adopters along: the challenge facing the PC industry. *Communications of the ACM*, 46(4):76–80, 2003.

Eugene L. Bryan. Bestpossible profits : guidebook for forest products companies. 2003.

Eugene L. Bryan and Andrew C. Bielat. The bestpossible enterprise. 2013.

Jason Burby and Shane Atchison. *Actionable Web Analytics: Using Data to Make Smart Business Decisions*. 2007.

Jonathan Burez and Dirk Van den Poel. Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3):4626–4636, 2009.

Raymond P. L. Buse and Thomas Zimmermann. Information needs for software development analytics. In *Proceedings of the 34th International Conference on Software Engineering*, pages 987–996, 2012.

Christian Chabot. Demystifying visual analytics. *IEEE Computer Graphics and Applications*, 29(2):84–87, 2009.

J. Chacko, L. P. Rees, C. W. Zobel, T. R. Rakes, R. S. Russell, and C. T. Ragsdale. Decision support for long-range, community-based planning to mitigate against and recover from potential multiple disasters. *Decision Support Systems*, 87:13–25, 2016.

A. Charles, M. Lauras, L. N. Van Wassenhove, and L. Dupont. Designing an efficient humanitarian supply network. *Journal of Operations Management*, 47-48:58–70, 2016.

C. M. K. Cheung, M. K. O. Lee, and X.-L. Jin. Customer engagement in an online social platform: A conceptual model and scale development. In *Proceedings of the 32nd International Conference on Information Systems*, pages 3105–3112. 2011.

C. M. K. Cheung, X.-L. Shen, Z. W. Y. Lee, and T. K. H. Chan. Promoting sales of online games through customer engagement. *Electronic Commerce Research and Applications*, 14(4):241–250, 2015.

CNN. Why console gaming is dying. 2012. URL http://www.cnn.com/2012/11/09/tech/gaming-gadgets/console-gaming-dead/. Accessed: 2015-07-14.

H. Cole and M. D. Griffiths. Social interactions in massively multiplayer online role-playing gamers. *CyberPsychology & Behavior*, 10(4):575–583, 2007.

M. Consalvo. Using your friends: social mechanics in social games. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, pages 188–195, 2011.

Christopher W Craighead, Kirk R Karwan, and Janis L Miller. The effects of severity of failure and customer loyalty on service recovery strategies. *Production and Operations Management*, 13(4):307–321, 2004.

A. T. Crooks and S. Wise. GIS and agent-based models for humanitarian assistance. *Computers, Environment and Urban Systems*, 41:100–111, 2013.

Thomas H. Davenport and Jeanne G. Harris. The dark side of customer analytics. *Harvard Business Review*, 85(5), 2007.

Thomas H Davenport et al. Competing on analytics. *harvard business review*, 84 (1):98, 2006.

P. De Jong and G. Z. Heller. *Generalized linear models for insurance data*. Cambridge University Press, 2008.

Dursun Delen and Haluk Demirkan. Data, information and analytics as services, 2013.

I. S. Dolinskaya, Z. E. Shi, K. R. Smilowitz, and M. Ross. Decentralized approaches to logistics coordination in humanitarian relief. In T. Doolen and E. Van Aken, editors, *Proceedings of the 2011 Industrial Engineering Research Conference*, 2011.

E. Domahidi, R. Festl, and T. Quandt. To dwell among gamers: Investigating the relationship between social online game use and gaming-related friendships. *Computers in Human Behavior*, 35:107–115, 2014.

Anders Drachen, Rafet Sifa, Christian Bauckhage, and Christian Thurau. Guns, swords and data: Clustering of player behavior in computer games in the wild. In *Proceedings of the 2012 IEEE Conference on Computational Intelligence and Games*, pages 163–170, 2012.

P. K. Dunn and G. K. Smyth. Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15(4):267–280, 2005.

M. Eftekhar and L. N. Van Wassenhove. Fleet management policies for humanitarian organizations: beyond the utilization-residual value trade-off. *Journal of Operations Management*, 44:1–12, 2016.

M. Eftekhar, A. Masini, A. Robotis, and L. N. Van Wassenhove. Vehicle procurement policy for humanitarian development programs. *Production and Operations Management*, 23(6):951–964, 2014.

Ö. Ergun, L. Gui, J. L. Heier Stamm, P. Keskinocak, and J. Swann. Improving humanitarian operations through technology-enabled collaboration. *Production and Operations Management*, 23(6):1002–1014, 2014.

James R. Evans. *Business Analytics: Methods, Models, and Decisions*. 2012.

James R Evans and Carl H Lindner. Business analytics: the next frontier for decision sciences. *Decision Line*, 43(2):4–6, 2012.

L. D. Fisher and D. Y. Lin. Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, 20(1):145–157, 1999.

A.C.M. Fong, S.C. Hui, and G. Jha. Data mining for decision support. *IT Professional*, 4(2):9–17, 2002.

C. Fornell, R. T. Rust, and M. G. Dekimpe. The effect of customer satisfaction on consumer spending growth. *Journal of Marketing Research*, 47(1):28–35, 2010.

James Paul Gee. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1):20–20, 2003.

Nicolas Glady, Bart Baesens, and Christophe Croux. Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197(1):402–411, 2009.

Erica Gralla, Jarrod Goentzel, and Charles Fine. Assessing trade-offs among multiple objectives for humanitarian aid delivery using expert preferences. *Production and Operations Management*, 23(6):978–989, 2014.

C. Guo, X. Chen, C. Zhang, and P. Goes. Social influence in freemium environments: Understanding its effect on willingness to play and willingness to pay in online social games. *Submitted for publication*, 2018. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2686965.

Anders Gustafsson, Michael D. Johnson, and Inger Roos. The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of Marketing*, 69(4):210–218, 2005.

John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymitr Ruta. Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10):2902–2917, 2007.

J. Hamari. Why do people buy virtual goods? Attitude toward virtual good purchases versus game enjoyment. *International Journal of Information Management*, 35(3):299–308, 2015.

J. Hamari and A. Järvinen. Building customer relationship through game mechanics in social games. In M. M. Cruz-Cunha, editor, *Business, Technological and Social Dimensions of Computer Games: Multidisciplinary Developments*, pages 348–365. IGI Global, 2011.

N. Hanner and R. Zarnekow. Purchasing behavior in free to play games: Concepts and empirical validation. In *Proceedings of the 48th Hawaii International Conference on System Sciences*, pages 3326–3335, 2015.

William J. Hauser. Marketing analytics: the evolution of marketing research in the twenty-first century. *Direct Marketing: An International Journal*, 1(1):38–54, 2007.

A. M. Hayashi. When to trust your gut. *Harvard Business Review*, 79(2):59–65, 2001.

J. J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1): 153–161, 1979.

Gregory R Heim and Kingshuk K Sinha. Operational drivers of customer loyalty in electronic retailing: An empirical analysis of electronic food retailers. *Manufacturing & Service Operations Management*, 3(3):264–271, 2001.

Clyde Holsapple, Anita Lee-Post, and Ram Pakath. A unified foundation for business analytics. *Decision Support Systems*, 64:130–141, 2014.

Harrison Hong, Jeffrey D Kubik, and Jeremy C Stein. Social interaction and stock-market participation. *The Journal of Finance*, 59(1):137–163, 2004.

C.-L. Hsu and H.-P. Lu. Consumer behavior in online game communities: A motivational factor perspective. *Computers in Human Behavior*, 23(3):1642–1659, 2007.

Meichun Hsu and Qiming Chen. Scalable data-intensive analytics. In *International Workshop on Business Intelligence for the Real-Time Enterprise*, pages 97–107, 2008.

Ming Hu, Joseph Milner, and Jiahua Wu. Liking and following and the newsvendor: Operations and marketing policies under social influence. *Management Science*, 62(3):867–879, 2016.

Jeff Huang, Thomas Zimmermann, Nachiappan Nagapan, Charles Harrison, and Bruce C. Phillips. Mastering the art of war: How patterns of gameplay influence skill in Halo. In *Proceedings of the 2013 SIGCHI Conference on Human Factors in Computing Systems*, pages 695–704, 2013.

Huffington Post. Mobile gaming is about to overtake consoles. 2014. URL http://www.huffingtonpost.com/2014/12/05/console-gaming_n_6269676.html. Accessed: 2015-07-14.

E. Iakovou, D. Vlachos, C. Keramydas, and D. Partsch. Dual sourcing for mitigating humanitarian supply chain disruptions. *Journal of Humanitarian Logistics and Supply Chain Management*, 4(2):245–264, 2014.

M. Jahre, J. Kembro, T. Rezvanian, O. Ergun, and S. J. Hapnes. Integrating supply chains for emergencies and ongoing operations in UNHCR. *Journal of Operations Management*, 45:57–72, 2016.

J. Jansz and L. Martens. Gaming at a LAN event: the social context of playing video games. *New Media & Society*, 7(3):333–355, 2005.

Jeroen Jansz and Martin Tanis. Appeal of playing online first person shooter games. *CyberPsychology & Behavior*, 10(1):133–136, 2007.

Mikael Jern, Monica Brezzi, and Patrik Lundblad. Geovisual analytics tools for communicating emergency and early warning. *Geographic Information and Cartography for Risk and Crisis Management : Towards Better Solutions*, pages 379–394, 2010.

Hallick Jn. Analytics and the data warehouse. *Health management technology*, 22 (6):24, 2001.

D. Kahneman and G. Klein. Conditions for intuitive expertise: A failure to disagree. *American Psychological Association*, 64(6):515–526, 2009.

E. H. Kaplan, D. L. Craft, and L. M. Wein. Emergency response to a smallpox attack: the case for mass vaccination. *Proceedings of the National Academy of Sciences*, 99(16):10935–10940, 2002.

Avinash Kaushik. *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity*. 2009.

Daniel A. Keim, Gennady L. Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. *Information Visualization*, (4950):154–175, 2008.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

M. Keshvari Fard, M. Eftekhar, and F. Papier. A policy for managing operational assets to minimize deprivation costs. *Submitted for publication*, 2017.

N. Khatri and H. A. Ng. The role of intuition in strategic decision making. *Human Relations*, 53:57–86, 2000.

D. King, P. Delfabbro, and M. Griffiths. Video game structural characteristics: A new psychological taxonomy. *International Journal of Mental Health and Addiction*, 8(1):90–106, 2010.

William R. King. Text analytics: Boon to knowledge management? *Information Systems Management*, 26(1):87–87, 2009.

Jörn Kohlhammer, Daniel A. Keim, Margit Pohl, Giuseppe Santucci, and Gennady L. Andrienko. Solving problems with visual analytics. *Procedia Computer Science*, 7:117–120, 2011.

Vikas Kumar, Luciano Batista, and Roger Maull. The impact of operations performance on customer loyalty. *Service Science*, 3(2):158–171, 2011.

Nathan Kunz, Luk N. Van Wassenhove, Rob McConnell, and Ketil Hov. Centralized vehicle leasing in humanitarian fleet management: the unhcr case. *Journal of Humanitarian Logistics and Supply Chain Management*, 5(3):387–404, 2015.

N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

V. Lehdonvirta. Virtual item sales as a revenue model: identifying attributes that drive purchase decisions. *Electronic Commerce Research*, 9(1-2):97–113, 2009.

Amanda Lenhart, Joseph Kahne, Ellen Middaugh, Alexandra Rankin Macgill, Chris Evans, and Jessica Vitak. Teens, video games, and civics: teens' gaming experiences are diverse and include significant social interaction and civic engagement. *Pew Internet & American Life Project*, 2008.

Matthew Liberatore and Wenhong Luo. Informs and the analytics movement: The view of the membership. *Interfaces*, 41(6):578–589, 2011.

Sharman Lichtenstein, David H.B. Bednall, and Stewart Adam. Marketing research and customer analytics: interfunctional knowledge integration. *International Journal of Technology Marketing*, 3(1):81–96, 2008.

D. Y. Lin. On fitting Cox's proportional hazards models to survey data. *Biometrika*, 87(1):37–47, 2000.

Xueming Luo. How does shopping with others influence impulsive purchasing? *Journal of Consumer Psychology*, 15(4):288–294, 2005.

B. Marr. Really big data at walmart: Real-time insights from their 40+ petabyte data cloud. 2017. URL https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/. Accessed: 2019-05-14.

B. Marr. How much data do we create every day? the mind-blowing stats everyone should read. 2018. URL https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read. Accessed: 2019-05-14.

P. Matl, R. F. Hartl, and T. Vidal. Workload equity in vehicle routing problems: A survey and analysis. *Transportation Science*, 52(2):239–260, 2017.

A. McClintock. The logistics of humanitarian emergencies: notes from the field. *Journal of Contingencies and Crisis Management*, 17(4):295–302, 2009.

Scott D McClurg. Social networks and political participation: The role of social interaction in explaining political participation. *Political Research Quarterly*, 56 (4):449–464, 2003.

J. H. McCoy. Overcoming the challenges of the last mile: A model of Riders for Health. In B. Denton, editor, *Handbook of Healthcare Operations Management*, pages 483–509. Springer, NY, 2012.

J. H. McCoy and H. L. Lee. Using fairness models to improve equity in health delivery fleet management. *Production and Operations Management*, 23(6):965–977, 2014.

Colleen McCue. Data mining and predictive analytics in public safety and security. *IT Professional*, 8(4):12–18, 2006.

P. McCullagh and J. A. Nelder. *Generalized linear models*, volume 37. CRC press, 1989.

Connie McNabb. Descriptive research methodologies. *Retrieved from*, 2012.

K. Meagher and E. G. S. Teo. Two-part tariffs in the online gaming industry: The role of creative destruction and network externalities. *Information Economics and Policy*, 17(4):457–470, 2005.

E. Mileva. Using Arellano-Bond dynamic panel GMM estimators in Stata. Technical report, Fordham University, 2007.

Sunil Mithas, Mayuram S Krishnan, and Claes Fornell. Why do customer relationship management applications affect customer satisfaction? *Journal of Marketing*, 69(4):201–209, 2005.

Diane A Mollenkopf, Elliot Rabinovich, Timothy M Laseter, and Kenneth K Boyer. Managing internet product returns: a focus on effective service operations. *Decision Sciences*, 38(2):215–250, 2007.

G.A. Moore. *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers*. Collins Business Essentials. HarperCollins, 2002. ISBN 9780060517120.

W. Ni, J. Shu, and M. Song. Location and emergency inventory pre-positioning for disaster response operations: Min-max robust model and a case study of Yushu earthquake. *Production and Operations Management*, 27(1):160–183, 2018.

D. Noyes. The top 20 valuable facebook statistics – updated may 2019. 2019. URL https://zephoria.com/top-15-valuable-facebook-statistics/. Accessed: 2019-05-14.

Charles Nyce and A Cpcu. Predictive analytics white paper. *American Institute for CPCU. Insurance Institute of America*, pages 9–10, 2007.

Jerry O'Dwyer and Ryan Renner. The promise of advanced supply chain analytics. *Supply Chain Management Review*, 15(1), 2011.

J. Paavilainen, J. Hamari, J. Stenros, and J. Kinnunen. Social network games: Players' perspectives. *Simulation & Gaming*, 44(6):794–820, 2013.

B.-W. Park and K. C. Lee. Exploring the value of purchasing online game items. *Computers in Human Behavior*, 27(6):2178–2185, 2011.

A. J. Pedraza-Martinez and L. N. Van Wassenhove. Vehicle replacement in the International Committee of the Red Cross. *Production and Operations Management*, 22(2):365–376, 2013.

A. J. Pedraza Martinez, Orla Stapleton, and Luk N. Van Wassenhove. Field vehicle fleet management in humanitarian operations: a case-based approach. *Journal of Operations Management*, 29(5):404–421, 2011.

A. J. Pedraza-Martinez, S. Hasija, and L. N. Van Wassenhove. Fleet management coordination in decentralized humanitarian operations. Technical report, INSEAD, 2013.

Jorge Peña and Jeffrey T Hancock. An analysis of socioemotional and task communication in online multiplayer video games. *Communication Research*, 33(1): 92–109, 2006.

B Joseph Pine and James H Gilmore. *The experience economy: Work is theatre & every business a stage*. Harvard Business Press, 1999.

W. B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality (2nd ed.)*. John Wiley and Sons, New York, 2011.

Madeleine E Pullman and Michael A Gross. Ability of experience design elements to elicit emotions and loyalty behaviors. *Decision Sciences*, 35(3):551–578, 2004.

L. Qiu, Q. Tang, and A. B. Whinston. Two formulas for success in social media: Learning and network effects. *Journal of Management Information Systems*, 32 (4):78–108, 2015.

M. S. Rauner. Using simulation for AIDS policy modeling: benefits for HIV/AIDS prevention policy makers in Vienna, Austria. *Health Care Management Science*, 5(2):121–134, 2002.

C. G. Rawls and M. A. Turnquist. Pre-positioning and dynamic delivery planning for short-term response following a natural disaster. *Socio-Economic Planning Sciences*, 42(1):46–54, 2012.

Bonnie Ray, Chid Apte, Kevin McAuliffe, Lea Deleris, and Eric Cope. Ibm research report harnessing uncertainty: The future of risk analytics. 2008.

Frederick F Reichheld. Loyalty-based management. *Harvard Business Review*, 71 (2):64–73, 1992.

Anne Robinson, Jack Levis, and Gary Bennett. Informs to officially join analytics movement. *OR/MS Today*, 37(5):59, 2010.

Everett M Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.

D. Roodman. How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1):86–136, 2009.

N. Sahebjamnia, S. A. Torabi, and S. A. Mansouri. A hybrid decision support system for managing humanitarian relief chains. *Decision Support Systems*, 95: 12–26, 2017.

J. Salmeron and A. Apte. Stochastic optimization for natural disaster asset prepositioning. *Production and Operations Management*, 19(5):561–574, 2010.

D. J. Sargent. A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics*, 54(4):1486–1497, 1998.

Satheesh Seenivasan, K Sudhir, and Debabrata Talukdar. Do store brands aid store loyalty? *Management Science*, 62(3):802–816, 2016.

Richard W. Selby. Analytics-driven dashboards enable leading indicators for requirements and designs of large-scale systems. *IEEE Software*, 26(1):41–49, 2009.

Jouni K Seppänen, Ella Bingham, and Heikki Mannila. A simple algorithm for topic identification in 0–1 data. In N. Lavrac, D. Gamberger, L. Todorovski, and H. Blockeel, editors, *Knowledge Discovery in Databases: PKDD 2003*, pages 423–434. Springer, 2003.

John L Sherry, Kristen Lucas, Bradley S Greenberg, and Ken Lachlan. Video game uses and gratifications as predictors of use and game preference. *Playing video games: Motives, responses, and consequences*, pages 213–224, 2006.

Eric Siegel. *Predictive analytics: The power to predict who will click, buy, lie, or die*. John Wiley & Sons, 2013.

Hugo P Simao, Jeff Day, Abraham P George, Ted Gifford, John Nienow, and Warren B Powell. An approximate dynamic programming algorithm for large-scale fleet management: A case application. *Transportation Science*, 43(2):178–197, 2009.

Sumit Sircar. Business intelligence in the business curriculum. *Communications of the Association for Information Systems*, 24(1):17, 2009.

Mikhail Smelyanskiy. Challenges of mapping financial analytics to many-core architecture. In *2008 Workshop on High Performance Computational Finance*, pages 1–1. IEEE, 2008.

M. Smithson and E. C. Merkle. *Generalized linear models for categorical and continuous limited dependent variables*. CRC Press, 2013.

G. K. Smyth and B. Jørgensen. Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *ASTIN Bulletin*, 32(1):143–157, 2002.

J. O. B. Soh and B. C. Y. Tan. Mobile gaming. *Communications of the ACM*, 51 (3):35–39, 2008.

G Spais and C Veloutsou. Marketing analytics: managing incomplete information in consumer markets and the contribution of mathematics to the accountability of marketing decisions. *South European Review of Business Finance and Accounting*, 3(1):127–150, 2005.

Srini S Srinivasan, Rolph Anderson, and Kishore Ponnavolu. Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of Retailing*, 78(1):41–50, 2002.

Thomas R Stewart and Claude McMillan. Descriptive and prescriptive models for judgment and decision making: implications for knowledge engineering. In *Expert judgment and expert systems*, pages 305–320. Springer, 1987.

H. Takeuchi, L.V. Subramaniam, T. Nasukawa, S. Roy, and S. Balakrishnan. A conversation-mining system for gathering insights to improve agent productivity. In *The 9th IEEE International Conference on E-Commerce Technology and The 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007)*, pages 465–468, 2007.

J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM International Conference on Knowledge Discovery and Data Mining*, pages 807–816, 2009.

David Taniar. *Progressive Methods in Data Warehousing and Business Intelligence: Concepts and Competitive Analytics*. 2009.

F. Tazelaar and C. Snijdes. Operational risk assessments by supply chain professionals: Process and performance. *Journal of Operations Management*, 31(1-2): 37–51, 2013.

R. Terlutter and M. L. Capella. The gamification of advertising: analysis and research directions of in-game advertising, advergames, and advertising in social network games. *Journal of Advertising*, 42(2-3):95–112, 2013.

The Entertainment Software Association. 2014 essential facts: The entertainment software association. 2014. URL http://www.theesa.com/wp-content/uploads/2014/10/ESA_EF_2014.pdf. Accessed: 2015-07-14.

The Entertainment Software Association. Industry fact: The entertainment software association. 2015. URL http://www.theesa.com/about-esa/industry-facts/. Accessed: 2015-07-13.

L. Thomas and E. M. Reyes. Tutorial: survival estimation for Cox regression models with time-varying coefficients using SAS and R. *Journal of Statistical Software*, 61(1):1:1–1:23, 2014.

Chunhua Tian, Rongzeng Cao, Hao Zhang, Feng Li, Wei Ding, and Bonnie Ray. Service analytics framework for web-delivered services. In *2008 IEEE International Conference on Service Operations and Logistics, and Informatics*, volume 1, pages 635–640. IEEE, 2008.

Brian M. Tomaszewski, Anthony C. Robinson, Chris Weaver, Michael Stryker, and Alan M. MacEachren. Geovisual analytics and crisis management. 2007.

F. Toyasaki, E. Arikan, I. Falagara Sigala, and L. Silbermayr. Disaster relief inventory management: Horizontal cooperation between humanitarian organizations. *Production and Operations Management*, 6(26):1221–1237, 2017.

Peter Trkman, Kevin McCormack, Marcos Paulo Valadares De Oliveira, and Marcelo Bronzo Ladeira. The impact of business analytics on supply chain performance. *Decision Support Systems*, 49(3):318–327, 2010.

A. A. Tsiatis and M. Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88 (2):447–458, 2001.

John Turner, Alan Scheller-Wolf, and Sridhar Tayur. OR practice-scheduling of dynamic in-game advertising. *Operations Research*, 59(1):1–16, 2011.

M. C. K. Tweedie. An index which distinguishes between some important exponential families. In *Statistics: Applications and new directions – Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, pages 579–604, 1984.

M. Vanajakumari, S. Kumar, and S. Gupta. An integrated logistics model for predictable disasters. *Production and Operations Management*, 25(5):791–811, 2016.

Shiri D Vivek, Sharon E Beatty, and Robert M Morgan. Customer engagement: Exploring customer relationships beyond purchase. *Journal of Marketing Theory and Practice*, 20(2):122–146, 2012.

Chris Voss, Aleda V Roth, and Richard B Chase. Experience, service operations strategy, and services as destinations: foundations and exploratory investigation. *Production and Operations Management*, 17(3):247–266, 2008.

D. Weibel, B. Wissmath, S. Habegger, Y. Steiner, and R. Groner. Playing online games against computer-vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior*, 24(5):2274–2291, 2008.

D. Westwood and M. D. Griffiths. The role of structural characteristics in videogame play motivation: A Q-methodology study. *Cyberpsychology, Behavior, and Social Networking*, 13(5):581–585, 2010.

D. Y. Wohn, C. Lampe, R. Wash, N. Ellison, and J. Vitak. The "S" in social network games: Initiating, maintaining, and enhancing relationships. In *Proceedings of the 44th Hawaii International Conference on System Sciences*, pages 2894–2903, 2011.

L. Yang, S. Dimitrov, and B. Mantin. Forecasting sales of new virtual goods with the Elo rating system. *Journal of Revenue and Pricing Management*, 13(6):457–469, 2014.

S. Ye and S. F. Wu. Measuring message propagation and social influence on Twitter.com. In L. Bolc, M. Makowski, and A. Wierzbicki, editors, *Proceedings of the 2nd International Conference on Social Informatics*, pages 216–231. Springer, 2010.

N. Yee. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence*, 15(3):309–329, 2006.

H. Peyton Young. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, 99(5):1899–1924, 2009.

S. A. Zenios, L. M. Wein, and G. M. Chertow. Evidence-based organ allocation. *The American Journal of Medicine*, 107(1):52–61, 1999.

Y. Zhang. Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Statistics and Computing*, 23(6):743–757, 2013.

F. Zhu and X. Zhang. The influence of online consumer reviews on the demand for experience goods: The case of video games. *Proceedings of the 27th International Conference on Information Systems*, pages 367–381, 2006.