8-9-2014

Three Essays on Consumer Product Returns

Guangzhi Shang
University of South Carolina - Columbia

Three Essays on Consumer Product Returns

by

Guangzhi Shang

Bachelor of Business Administration
The Hong Kong Polytechnic University 2010

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Business Administration

Darla Moore School of Business

University of South Carolina

2014

Accepted by:

Michael R. Galbreth, Major Professor

Mark E. Ferguson, Committee Member

Pelin Pekgün, Committee Member

Bikram P. Ghosh, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

Dedication

I dedicate this dissertation to my parents,

Dingyu Sun and Yuzhuo Shang,

who have always loved and supported me.

I could not be who I am today without their most generous encouragement.

I also dedicate this dissertation to my grandparents,

Yanhua Hui, Shuwen Shang, Cuihua Jiang, and Longjia Sun,

who give me the warmest smiles that lighten up my life,

who still keep the virtuous habit of giving me pocket money.

Acknowledgments

I give my first thanks to my advisor, dissertation chair, and coauthor, Dr. Mike Galbreth. I feel extremely lucky that my interaction with Mike started very early in my PhD career. Mike taught me step-by-step about how to develop an interesting idea into a research project. His caring and patient personalities are precious treasures for me because I often find myself obsessed with experimenting with thoughts that do not materialize into concrete ideas. I'm deeply indebted to him for having the kind of freedom that a doctoral student can rarely enjoy. I am extremely happy that this flexibility helped me take the most difficult yet essential course for a young researcher, in my opinion. That is, how to learn things outside classroom and on demand. I have to also thank him for giving me endless encouragements, especially during some of the low times.

My next big thanks goes to Dr. Mark Ferguson, my dissertation committee member, mentor, and program director. I wish I could have the opportunity to work with Mark even earlier. He gave me many inspirations of how to spice up academic research with more practicality. I went to my first practitioners' conference with him, where I found out nerdy research can also be easy-going. Being the PhD program director, Mark provided me with the most generous travel funding, which was the life blood for a student who often needs to spend beyond his budget.

I would also like to thank my other dissertation committee members, Dr. Bikram Ghosh and Dr. Pelin Pekgün. I learned many analytical skills from Bikram during working on a project that he was very generous to have me involved. Pelin has the most friendly personality for a student. She is also a truly constructive collaborator, an attribute that I certainly need more development in my career. My dissertation could not have come this far without the support from all of you. Thanks again!

Let me also give my gratitude to Dr. Manoj Malhotra for his guidance on two other projects and his help as a recommender. I need also to thank Dr. Sanjay Ahire for his generosity of providing help in the early stage of this dissertation. I also want to mention Dr. Stephen Finger, who have taught my industrial organization class and born with my endless questions on econometrics.

I would like also to thank my coauthors on various projects, not previously mentioned, Dr. Joan Donohue, Dr. Timothy Fry, and Dr. Robert Ployhart. I also want to acknowledge those professors who have either directly taught me or indirectly inspired me.

My previous officemates and good friends, Dr. Cidgem Ataseven, Dr. Ashley Metcalf, and Dr. Mariana Nicholae, have given me so much fun during my PhD study. My current officemates and good friends, Erin McKie, Minseok Park, Cherry Singhal, Deepa Wani, and Övünç Yılmaz, have provided me with some thoughts on my dissertation and cared for my well-being in the program. Julia Witherspoon, our department secretary, and Scott Ranges, our business PhD program coordinator, have answered my numerous questions regarding the administrative duties. A big thanks to them as well.

I want to also thank all my friends at University of South Carolina, who influenced me in some way. Although it's impossible to list all of them, those come to mind include Nicholas Bailey, Ruiyuan Chen, Wen Chen, Yin Fu, Hye Sun Kang, Chris Ling, Pulkit Nigam, He Wang, Yijia Zhao, and Dr. Xiaolan Zheng.

The last and most special thanks go to my girlfriend Shengfang Sun, who I shared highs, lows, laugh, sorrow, and everything else.

Abstract

Return policies (aka. money-back-guarantee policies) are frequently offered by retailers such that consumers can bring back the purchased products that do not fit their needs. On the benefit side, consumers perceive these policies as a risk-mitigation service and are willing to pay more for an otherwise identical product. All else being equal, being lenient in its return policy should allow a retailer to enjoy higher sales, which leads to increased revenue. On the cost side, a generous return policy will induce more returns and possibly also late returns. In 2007 alone, U.S. manufacturers and retailers paid more than $100 billion to process consumer returns. For electronic products, this cost is estimated at $16.7 billion per year, which represents 6% of an average electronics manufacturer' revenue and 3% of an average retailer's total sales. The return rate in the retail sector ranges from 10% to 20%. To effectively manage returns, a retailers should concurrently optimize its price and restocking fee so that it could enjoy the most benefit from offering a return policy. At the same time, it should also seek to better forecast the quantity of incoming returns and make efforts to reduces the number of futures returns.

Despite the magnitude of the returns management problem, academic research on this topic is still relatively scant. Through three separate studies, this dissertation delves into returns management from a multitude of perspectives. The first study explores a retailer's optimal return policy when a fraction of the market is consist of opportunistic consumers, who might purchase with the explicit intention of returning. The second study demonstrates how retailers might use their point-of-sale data to better forecast future returns, uncover the duration of consumers' product trial, and

identify products that are more likely to be returned. The last study introduces a novel approach for retailers to quantify the incremental willingness-to-pay that consumers assign to money-back-guarantee policies in the on-line environment.

Table of Contents

List of Tables

List of Figures

Chapter 1

Overview

Consumer product returns represent a substantial challenge and a size-able expense for retailers and manufacturers. In 2007 alone, U.S. manufacturers and retailers paid more than $100 billion to process consumer returns Blanchard (2007). For electronic products, the cost of returns in the U.S. has been estimated at $16.7 billion per year, representing 6% of revenue for an average consumer electronics manufacturer and 3% of total sales for an average retailer (Douthit et al., 2011). According to National Retail Federation (2011a), the return rate in the retail sector ranges from 10% to 20%. Despite such magnitude, research on the management of consumer returns is relatively scant. This dissertation attempts to delve into three critical and intriguing issues for managing returns. The key insights are previewed in following three paragraphs.

Most retailers offer refunds to consumers who, after a trial period, return a product that they find does not fit their needs. There is evidence from practice that some consumers are willing to use this return option opportunistically, essentially "renting" a product by using it during the trial period and then returning it. Restocking fees (partial refunds) can be used to combat this behavior. However, such fees might be viewed negatively by consumers in cases where the return is due to a true lack of fit. We derive optimal pricing and return policies that explicitly consider both the extent of opportunism (how many consumers consider such behavior) and the benefit of opportunism (the attractiveness of the renting option). Our analysis reveals several new insights for retailers regarding pricing and return policies when opportunism is present in the marketplace. We also examine the profit impact of changes in the extent or benefit of opportunism, finding that this impact is not always intuitive, and that increases in either of these two opportunism constructs can actually increase profits in some cases. In an extension, we provide additional insights for when the salvage value of returned items is linked to how much utility has been extracted by the opportunistic consumer.

Almost all prescriptive practices for lowering the cost of processing returns require a forecast of how many consumer returns will arrive in each period. Despite the fact that retail transactions data is readily available and has been shown to provide significantly better demand forecasts for new product sales, this data has not been explored for forecasting consumer returns. Using a data set provided by a major U.S. retail chain and consisting of 20,801 transactions of 2,483 electronic products, we develop an econometric model that simultaneously explains the consumer's experience duration and return probability, which is in turn used for predicting return quantity in a given time period. This approach yields 20% to 40% lower forecast errors than the benchmark time-series models, and the performance gain is sustained even with a very parsimonious set of explanatory variables. Our econometric model also identifies promising managerial actions for lowering the cost of consumer returns such as better targeting of buyer assistance programs and choosing different return time windows for specific product families.

While Money-Back-Guarantee (MBG) return policies are common in practice, the academic literature provides little guidance on the value, if any, that consumers assign to such a policy. For on-line retailers, both shipping charges and seller reputation may influence the perceived value of MBGs. A non-refundable forward shipping charge is likely to be perceived by consumers as an implicit restocking fee and hence, decrease the value of an MBG policy. On the other hand, good customer feedback may signal that MBGs will be handled easily, thus increasing their value in the eyes of consumers. To investigate the value of MBGs in the on-line context, we collect the final selling prices of identical products offered by different vendors (with different return policies) on eBay auctions. Our data set includes 2946 transactions across 86 consumer electronic products. Endogeneity of the MBG decision by sellers is addressed using a maximum likelihood estimator based on the error correlation structure. We estimate that offering an MBG policy for on-line consumer electronics increases a customer's valuation of a product by an average of 5.2% and show that this increase is substantially influenced by both seller reputation and shipping charges. Our results provide an empirical measure of the value of an MBG that can be weighed against the cost implications of providing such a policy.

Chapter 2

Optimal Retail Return Policies with Consumer Opportunism

2.1 Introduction and Literature

Returns are viewed as necessary by many retailers since it is often impossible for a consumer to fully resolve all uncertainty regarding the fit of the item to her needs before a purchase is made (Nelson 1970). In-store assistance, on-line reviews, and other resources can provide some information, but a trial period is often needed to assess fit. Indeed, Ferguson et al. (2006) explain that products are often returned with no functional or cosmetic defect, but rather because a mismatch between product attributes and consumer preferences is revealed during the trial period. Reasons for mismatch can include installation difficulty, product performance issues, or remorse (Kumar et al., 2002). Given the role of the trial period in resolving fit uncertainty, it is generally accepted that consumers should have the opportunity to return a non-defective retail purchase for a refund (possibly net of a restocking fee) if, after further evaluation, they find that it does not meet their needs. However, consumer returns are not driven exclusively by product mismatch with consumer needs – some consumers take advantage of lenient return policies by purchasing items with the explicit intent of returning them after a period of use, essentially “renting” the items. Familiar examples of such opportunistic behavior include buying a digital camcorder to use on a single vacation or buying a cocktail dress to wear to one event. This type of behavior is well-known among retailers. Best Buy identified such consumers as “devil” customers (McWilliams, 2004), while studies of apparel retailers have dubbed the behavior “wardrobe borrowing” (Wood, 2001). According to a recent survey by the National Retail Federation (2011b), 61.4% of retailers experience such opportunistic returns.

Most analytical models of consumer returns assume that *no* consumers will consider opportunistic behavior (e.g. Su, 2009; Shulman et al., 2011; Hsiao and Chen, 2012; Akçay et al., 2013). On the other hand, a few studies have assumed that *all* consumers will consider opportunistic behavior (Chu et al., 1998; Hess et al., 1996). We suggest that the best representation of reality lies between these two extremes. Specifically, since opportunistic behavior is not consistent with the intention of the returns policy, it may be viewed as socially unacceptable by some consumers (indeed, casual conversations about this research have elicited responses ranging from open admissions of opportunistic behavior to expressions of outrage at the idea). Willingness to consider such actions is a function of one’s psychological traits such as self-concept (Mazar and Ariely, 2006) and symbolic self-completion (Rosenbaum and Kuntze, 2005). Laboratory evidence has shown that individuals who emphasize the alignment between self-image and social desirability will not engage in opportunistic behavior (Mazar et al., 2008). Thus, a model of returns behavior should reflect the fact that some consumers reject opportunistic behavior while others include it in their consideration set. Accordingly, in this study we assume that there exists an exogenous fraction of consumers who will consider opportunistic returns behavior.

We use the analytical framework of Su (2009) (see also Akçay et al., 2013) to model a retail context in which returns might occur due to a low realized valuation of the product once uncertainty has been eliminated after purchase. We augment this framework to accommodate the coexistence of two consumer segments - an “opportunistic” segment (willing to consider renting items) and an “ordinary” segment (does not consider the renting option). The impact of opportunism on the retailer is conceptualized as having two components: the size of the opportunistic segment (which we refer to as the *extent of opportunism*), and the fraction of product value that can be extracted by opportunistic consumers during the trial period (the *benefit of opportunism*). In the following sections we examine the impact of these two opportunism constructs on the retailer's optimal pricing strategy and profits.

2.2 Model

We consider a monopoly retailer selling to a market consisting of some mix of opportunistic and ordinary consumers. We first delineate how each segment makes its purchase and subsequent keep/return decisions. We then determine the resultant demand from each segment and derive total retailer profits.

Consumers

The products of interest for our analysis are experience goods (Nelson, 1970). At the time of purchase, consumers face uncertainty regarding whether an experience good will match their needs. This uncertainty will be resolved after product trial. Following the approach of Su (2009) and Akçay et al. (2013), we assume that consumers have the same *ex ante* evaluation of the product, v_0, and their *ex post* adjustments to this initial evaluation, ε, are random draws from a common distribution. Specifically, we let $v_0 = \frac{1}{2}$ and ε uniformly distributed between $-\frac{1}{2}$ and $\frac{1}{2}$. Put it differently, consumers' true evaluation of the product, $v = v_0 + \varepsilon$, is uniformly distributed[1] between 0 and 1.

As described in the Introduction, consumers in our model are categorized into two groups based on their willingness to opportunistically use the return policy to

[1]The assumption that a consumer's post-trial product evaluation follows a uniform distribution is consistent with previous analytical models of consumer returns (e.g. Akçay et al., 2013; Shulman et al., 2011).

rent a product. Ordinary consumers use the product trial period for its intended purpose — to assess fit. They do not consider renting, as they find this behavior to be inconsistent with social norms and thus unacceptable. Opportunistic consumers do not find renting unacceptable, either because it is accepted by their social peers or because they have a low regard for social desirability (Mazar and Ariely, 2006; Mazar et al., 2008). Note that opportunistic consumers do not simply use the trial period to asses fit - instead, they use the product intensively during the trial period, and hence their returns might be in worse condition than returns made by ordinary consumers. We denote size of segment of opportunistic consumers, i.e. the *extent of opportunism*, as γ. It follows that rest of the market, $1-\gamma$, is made up of ordinary consumers.

Each consumer makes two decisions: first, she decides whether or not to buy the product; secondly, after realizing her true valuation of the product, she decides whether to keep or return it. Assuming a price p and a restocking fee f, an ordinary consumer who has purchased will keep the product if her realized product evaluation, v, is at least as high as the refund, $p-f$. If the opposite is true ($v < p-f$), she will return the product and claim the refund. Thus, her expected net utility from making the purchase is $U_{or} = E\max(v, p-f) - p$. U_{or} can be simplified as follows:

$$U_{or} = E\max(v, p-f) - p = \int_{p-f}^{1} v dv + \int_{0}^{p-f} (p-f) dv - p = \frac{1+(p-f)^2-2p}{2} \quad (2.1)$$

Opportunistic consumers consider abusing the return policy such that they partially consume the product during the trial period and pay the restocking fee. Essentially, this abusive behavior is analogous to renting from the retailer, where rental period is the return time window and rental charge is the restocking fee. Let β denote the fraction of product value consumed before return, $\beta \in [0,1]$. It follows that the rent option gives opportunistic consumers $\beta v + p - f$ utility. That is, the sum of the consumed product value, βv, and the refund, $p-f$. We refer to β as the *benefit of opportunism* and suggest that it is largely product-specific. Consider, for example, a tuxedo versus a casual jacket. The former is typically used on rare occasions, while the latter is intended for everyday wear. In other words, a larger proportion of the value of a tuxedo can be consumed in just a few uses, as opposed to the much slower consumption of a casual jacket. Thus, *ceteris paribus*, the tuxedo has a much higher β than the casual jacket (a similar example of the product-specific nature of β was observed in the experimental study of Hjort and Lantz (2012)). The same logic applies for many other product categories – camcorders vs. music players, sunglasses vs. eyeglasses. etc. – in which some items might be needed for use on specific, rare occasions (such as a camcorder or sunglasses on a vacation) while others provide recurring benefits over a longer period of time.

For the opportunistic segment, a consumer who has purchased the product compares the utilities of renting, $\beta v + p - f$, and keeping, v, to decide whether to make an opportunistic return. The two options have the same utility when $v = \frac{p-f}{1-\beta}$. When β is large, i.e. $\frac{p-f}{1-\beta} > 1$, it is possible that opportunistic consumers will always return. In this case, their expected net utility from purchasing the product U_{op} is simply $E(\beta v + p - f) - p$. If β is small and hence $\frac{p-f}{1-\beta} \leq 1$, their expected net utility U_{op} is $E \max(v, \beta v + p - f) - p$ instead. U_{op} can be simplified as follows:

$$U_{op} = \begin{cases} E \max(v, \beta v + p - f) - p = \frac{1-\beta+(f-p)^2-2p(1-\beta)}{2(1-\beta)} & \text{for } \frac{p-f}{1-\beta} \leq 1, \\ E(\beta v + p - f) - p = \frac{\beta}{2} - f & \text{for } \frac{p-f}{1-\beta} > 1. \end{cases} \tag{2.2}$$

We restrict our attention to $\gamma \in [0, \frac{1}{2}]$, which essentially states that there are more ordinary consumers than opportunistic consumers in the market[2]. If a product is returned by an ordinary consumer, the retailer earns restocking fee f plus salvage value s_1 net unit product cost c, i.e. $f + s_1 - c$. If returned by an opportunistic consumer, the retailer earns $f + s_2 - c$ instead. $s_1 > s_2$ captures the fact

[2]National Retail Federation (2011a) reports that fraudulent returns, which nest abusive returns as a subset, represent around 8.5% of total returns. This implies that the population of consumers that consider return policy abuse is not likely to be very large, making this a reasonable restriction on γ.

that opportunistic returns are in worse condition due to excessive wear and tear. In addition, we assume that $\frac{1}{2} > c > s_1$, which implies that the retailer's unit product cost is less than consumers' average product evaluation but larger than unit salvage value. Lastly, when a buyer retains her purchase, the retailer earns $p - c$. Denote the retailer's unit profit from ordinary consumers as π_{or}. Given the respective probabilities of keeping the product, $1 - (p - f)$, and returning it, $p - f$, we have $\pi_{or} = (p-c)[1-(p-f)]+(f+s_1-c)(p-f)$. Denote the unit profit from opportunistic consumers as π_{op1} for $\frac{p-f}{1-\beta} < 1$ and π_{op2} for $\frac{p-f}{1-\beta} > 1$. Recall that when $\frac{p-f}{1-\beta} < 1$, some opportunistic consumers return, and when $\frac{p-f}{1-\beta} > 1$, all opportunistic consumers return. Thus we have $\pi_{op1} = \left(1 - \frac{p-f}{1-\beta}\right)(p-c)+\frac{p-f}{1-\beta}(f+s_2-c)$ and $\pi_{op2} = f+s_2-c$. Normalizing the market size to 1, the retailer's profit function is as follows.

$$\pi = \begin{cases} (1-\gamma)\pi_{or} + \gamma\pi_{op1} & \text{for } \frac{p-f}{1-\beta} \leq 1, \\ (1-\gamma)\pi_{or} + \gamma\pi_{op2} & \text{for } \frac{p-f}{1-\beta} > 1. \end{cases} \tag{2.3}$$

The retailer's problem is to choose the price p and restocking fee f that maximize (2.3) subject to the constraints that both segments of consumers participate in the market ($U_{or} \geq 0, U_{op} \geq 0$).[3] The timing of the game is as follows:

1. The retailer chooses the price p and restocking fee f.
2. Consumers make their purchases and realize their true valuations v.
3. Ordinary consumers keep the product if $v > p - f$ and return it otherwise; opportunistic consumers keep the product if $v > \beta v + p - f$ and return otherwise.

2.3 Results

Solving the retailer's optimization problem yields the following lemma:

[3]In Appendix A, we show that it is always profit maximizing for the retailer to serve both segments as long as a mild condition on unit product cost , $c < Min[\frac{1-s_1+2s_1^2}{2}, \frac{2s_2+s_1^2(-1+s_2(2+s_2))}{4s_2}]$, holds.

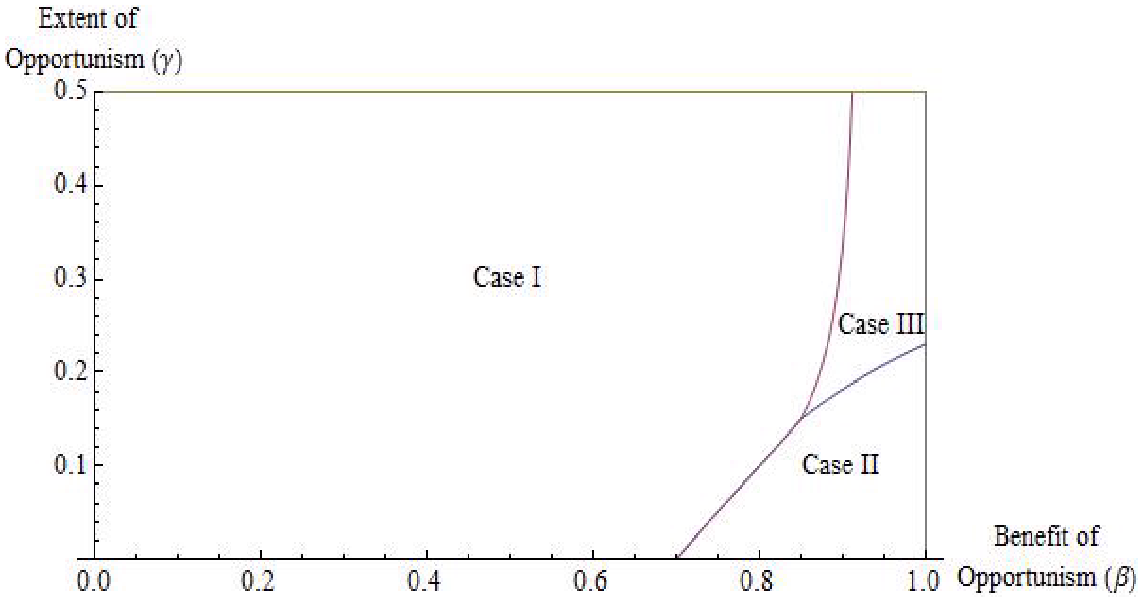

Figure 2.1 Retailer's Optimal Return Policy Regions

Lemma 2.1. *The retailer's optimal price (p^*) and restocking fee (f^*) depend on the extent (γ) and the benefit (β) of opportunism, which together define three regions of the solution space. Table A.2 in Appendix A summarizes p^*, f^*, and optimal profit (π^*), as well as the boundary conditions for each region. Figure 2.3 provides a visual illustration of the three optimal policy regions ($c = 0.35$, $s_1 = 0.3$, and $s_2 = 0.15$ in that figure).*

Note that when there is no opportunistic consumer ($\gamma = 0$), our optimal price ($p^* = \frac{1+s_1^2}{2}$) and restocking fee ($f^* = \frac{(1-s_1)^2}{2}$) are consistent with previous literature (Akçay et al., 2013; Su, 2009) given the uniform distribution of v. Next, we provide rationale for why the retailer should adopt three distinctive types of return policies.

First, consider the difference between Case I and the other two cases. The benefit of opportunism, β, captures the opportunistic consumers' incentive to rent the product. When β is low, it is relatively easy to induce opportunistic consumers to keep the product, since the benefit of renting is relatively low. As β increases, the appeal of keeping the purchased item, as opposed to renting and returning it, is reduced (recall that the probability of keep for the opportunistic segment is $1 - \frac{p-f}{1-\beta}$). Consequently, the opportunistic segment will always rent when β is sufficiently high - this occurs in Cases II and III in the Figure. The difference between Cases II and III can be understood as follows. Since opportunistic consumers always rent in these regions, the only way for the retailer to earn more profit from the opportunistic segment is through increasing restocking fee, which is evident in $\pi_{op2} = f + s_2 - c$. in Case II (high β, low γ), such a restocking fee increase is possible for the retailer. However, as γ increases, an increasingly large restocking fee is needed to address the prevalence of opportunism. At some point, a γ threshold is reached - the sum of restocking fee and salvage value has become very high relative to price - and it is no longer optimal to continue to increase the restocking fee. This defines Case III (high β, high γ) - as long as the retailer is in this region, its price and restocking fee are unaffected by γ.

Impact of the Extent of Opportunism

In this section, we study the profit maximizing response to changes in the extent of opportunism.

Proposition 2.1. *(a) Optimal price is non-increasing in the extent of opportunism.*

(b) Optimal restocking fee is non-decreasing in the extent of opportunism.

(c) Moreover, $\left|\frac{\partial p^*}{\partial \gamma}\right| \leq \left|\frac{\partial f^*}{\partial \gamma}\right|$.

We discuss parts (a) and (b) of the Proposition in terms of the regions of the solution space (Cases I-III) shown in Figure 2.3. First, consider restocking fee. In Case I and Case II, the retailer should, as expected, always increase restocking fee in response to more opportunistic consumers. For reasons explained earlier, the retailer does not adjust restocking fee when both the benefit and extent of opportunism are high (Case III). In terms of pricing, whenever the retailer increases restocking fee to

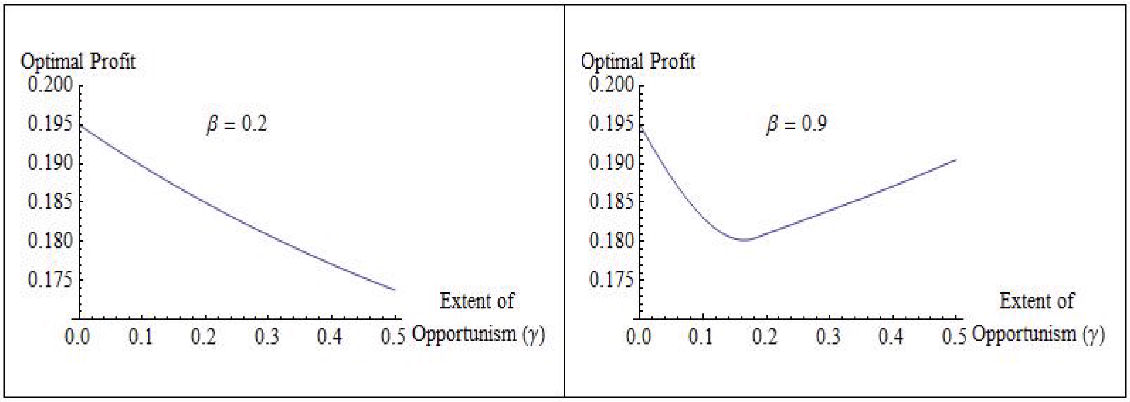

Figure 2.2 Impact of the Extent of Opportunism on Optimal Profit

extract more profit from the opportunistic segment, it needs to decrease price at the same time in order to offset this cost for ordinary consumers.

The intuition for part (c) follows from this idea of an offsetting price decrease along with the fact that, at optimality, an ordinary consumer's probability of keeping the product is always larger than the probability of returning it. Thus, the negative utility impact of increased restocking fee can be offset by a price decrease of a lesser magnitude (since the probability of incurring the price is larger than the probability of incurring the restocking fee).

Proposition 2.2. *(a) When the benefit of opportunism is low ($\beta < \widehat{\beta}$), optimal profit is decreasing in the extent of opportunism, γ.*

(b) When the benefit of opportunism is high ($\beta > \widehat{\beta}$), optimal profit is U-shaped in the extent of opportunism, γ. $\widehat{\beta} = \frac{\sqrt{(s_1-s_2)(9s_1-s_2)}+s_2}{2s_1} - \frac{1}{2}$.

Figure 2.3 provides a visual illustration of the above proposition ($c = 0.35$, $s_1 = 0.3$, and $s_2 = 0.15$ in that figure). To more clearly explain this finding, we decompose the profit impact of the extent of opportunism into three parts, $\frac{\partial \pi^*}{\partial \gamma} = \frac{\partial[(1-\gamma)\pi^*_{or}+\gamma\pi^*_{op}]}{\partial \gamma} = (1-\gamma)\frac{\partial \pi^*_{or}}{\partial \gamma} + \gamma\frac{\partial \pi^*_{op}}{\partial \gamma} + (\pi^*_{op} - \pi^*_{or})$. The first two parts measure the indirect effects of γ transmitted by the segment-specific unit profits, while the last part measures the direct effect of replacing ordinary consumers with opportunistic ones. As γ increases from zero, the retailer gradually shifts the focus of its pricing scheme away from the ordinary segment, evidenced by raising the restocking fee. Therefore, it is intuitive to see that the unit profit it earns from the ordinary segment is decreasing ($\frac{\partial \pi^*_{or}}{\partial \gamma} < 0$), while its unit profit from the opportunistic segment is increasing ($\frac{\partial \pi^*_{op}}{\partial \gamma} > 0$). Interestingly, we can show that the weighted average of these two counteracting unit-profit effects, $(1-\gamma)\frac{\partial \pi^*_{or}}{\partial \gamma} + \gamma \frac{\partial \pi^*_{op}}{\partial \gamma}$, is always zero. It follows that the final effect of γ – a certain number of ordinary consumers is replaced by the same number of opportunistic consumers – determines overall effect of γ on optimal profit.

Since the unit profit from ordinary consumers (π^*_{or}) is always decreasing in the extent of opportunism and that from the opportunistic consumers (π^*_{op}) is always increasing, $\pi^*_{op} - \pi^*_{or}$ increases in γ. That is, switching an opportunistic consumer with an ordinary one is more costly to the retailer when the extent of opportunism is small. As γ increases, $\pi^*_{op} - \pi^*_{or}$ continues to raise. However, whether this negative effect can change its sign and turn to a positive one depends on the profitability of the opportunistic consumers. When the benefit of opportunism (β) is high, opportunistic consumers are willing to pay a high price to rent the item, which implies that the retailer is able to profit more from the returns through a high restocking fee. Indeed, we see that optimal profit can increase with the the opportunistic segment size only when β is high.

To summarize, opportunistic consumers are always detrimental to the retailer's profit when renting is relatively less attractive to them (low β). However, when these consumers can extract more benefit from renting (high β) and hence become more profitable to the retailer, an increase in the opportunistic segment has a counter-intuitive U-shaped effect on retailer's profit. In other words, an increase in the extent of opportunism can increase profits in some cases.

Corollary 2.1. *When optimal profit is decreasing in the extent of opportunism, it decreases at a decreasing rate. That is,* $\frac{\partial^2 \pi^*}{\partial^2 \gamma} > 0$ *given* $\frac{\partial \pi^*}{\partial \gamma} < 0$.

The above finding makes an interesting practical implication for retailers. Consider the technologies available for identifying individuals who make return claims at abnormally frequent rates, such as the Verify-2 by Retail Equation Inc. (Speights and Hilinski, 2005). Employing such tracking technologies, retailers can marginally reduce the number of opportunistic consumers, γ. Now, assume a retailer's is in the region where its profit is negatively affected by γ. Corollary 2.1 provides guidance into when the marginal benefit of adopting the technology is highest. Specifically, we find that a reduction in the opportunistic segment size provides the largest payoff when size of the segment is small. Therefore, when informed by this result, retailers should adopt a more proactive approach toward combating abusive returns. In other words, the best timing for investing in the returns tracking technology is, counter-intuitively, not when the opportunistic consumer population is considerable but when it is still limited.

Impact of the Benefit of Opportunism

We now turn our attention to the second aspect of consumer opportunism - the benefit of opportunism, β. Note that retailers might have some limited influence over β through adjusting the time duration of the return window – longer windows provide more opportunity to benefit from an item before returning it. Thus, β is analogous to γ in the previous section in that, while neither is a firm's decision variable, both might be influenced at the margin (β by adjusting the return time window, γ by implementing consumer tracking technologies). This limited endogeneity makes the following comparative statics particularly relevant.

Proposition 2.3. *The marginal impact of the benefit of opportunism (β) on profit, $\frac{\partial \pi^*}{\partial \beta}$, can be either zero or positive or negative:*

(a) Optimal profit is unaffected by β for $\beta \in [1-s_1, 1]$ and $\gamma \in [0, Min[\widehat{\gamma}, \frac{\beta+s_1-1}{2\beta+s_1-1}]]$.

(b) Optimal profit is increasing in β for $\beta \in [\frac{s_1-s_2}{s_1}, 1]$ and $\gamma \in [\underline{\gamma}, \frac{1}{2}]$.

(c) Optimal profit is decreasing in β otherwise. $\underline{\gamma} = Max[\frac{(s_1-s_2)(-1+\beta)}{s_2-s_1(1-\beta)-2s_2\beta}, \frac{\beta+s_1-1}{2\beta+s_1-1}]$.

In comparing Proposition 2.3 with 2.2, it is interesting to note that, while both β and γ can be considered as measures of the potential negative impact of opportunism, they affect profit in different ways. Specifically, an increase in the benefit of opportunism, β, has a more nuanced impact – it could be either beneficial or irrelevant or detrimental to retailer's profit. Since β to a certain extent is controlled by the time length of return window, Proposition 2.3 offers interesting insights into when and how a longer return window exerts an indirect effect on retailer's profit.

When the extent of opportunism is relatively low (part (a) of Proposition 2.3), the retailer tailors its return policy more towards the ordinary segment through a low restocking fee. This low restocking fee, if coupled with a high benefit of opportunism β, induces the opportunistic consumers to always rent. As long as γ is low, it is not optimal for the retailer to combat this renting through a higher restocking fee, which would discourage the ordinary consumers (who make up most of the market) from buying. As a result, the retailer simply maintains a "static" return policy in this region (no changes to p or f) as β increases. This explains, in part (a) of the above proposition, why β has no impact on optimal profit when β is high and γ is low.

Next, we consider part (b), where both γ and β are high - i.e. many opportunistic consumers deriving a large utility from renting. In this context, the retailer can tailor its return policy more towards the opportunistic segment, which is relatively large and has a high valuation of the rental option. As β increases further, renting becomes even more appealing, and the retailer can aggressively raise restocking fee f and enjoy a higher unit profit from the opportunistic segment. The higher restocking fee does deter ordinary consumers, but the increased profits from the opportunistic segment more than offset this loss, and the net result is that profits increase.

Lastly, increases in the benefit that consumers receive from opportunistic behavior never improve firm profits when the extent of opportunism is small. Again, a higher

β increases the frequency of abusive behavior in the opportunistic segment. If the restocking fee is also lenient, the retailer cannot recover the lost of net sales due to an increased volume of returns. This is exactly what happened in part (c) of the above proposition.

2.4 Extension: Linking Salvage Value to the Benefit of Opportunism

In the previous section, we make the reasonable assumption that an item returned by an opportunistic consumer will have a lower salvage value than one returned by an ordinary consumer. In this section, we refine this aspect of the model by explicitly linking β to salvage value. Recall that ordinary consumers simply use the trial period to assess the fit of the product to their needs, while opportunistic consumers use the product intensively during the trial. By definition, higher-β products can be used more intensively by opportunistic consumers during the trial, and thus might show more signs of wear-and-tear than lower-β products, in turn leading to a lower salvage value (lower s_2). As mentioned earlier, that a longer time window increases the opportunity to extract value during the trial period (higher β). In this section we consider the possibility that it might also lead to a lower salvage value of opportunistic returns.

Our model can be changed to reflect this by simply redefining $s_1 = s$ and $s_2 = s - \beta\delta$. With this setup we retain model parsimony while capturing the idea that a higher β increases the quality differential between the opportunistic returns and the ordinary returns. At the same time, s_2 is still strictly smaller than s_1, which preserves the expectation that renting in general reduces salvage value relative to product evaluation by an ordinary consumer. Details of the analysis of this extension is in Appendix E. We find that, while all of our previous insights regarding the extent of opportunism (γ) remain valid, the profit impact of the benefit of opportunism (β) changes, as detailed in the following Proposition.

Proposition 2.4. *When salvage value is linked to β, the marginal impact of β on profit, $\frac{\partial \pi^*}{\partial \beta}$, can be either positive or negative:*

(a) Optimal profit is increasing in β for $\beta \in [\frac{2\delta}{s+\delta}, 1]$ and $\gamma \in [\frac{\delta(1-\beta)}{\beta(s-\delta)}, \frac{1}{2}]$.

(b) Optimal profit is decreasing in β otherwise.

Proposition 2.4 differs from Proposition 2.3 when γ is low and β is high (this context is included in part (b) of Proposition 2.4). When salvage value is linked to the benefit of opportunism, β exerts a negative impact on profit. This contrasts with the finding of no profit impact of β in this context with the main model (part (a) of Proposition 2.3). Recall that when the opportunistic segment size is small, the retailer implements a low restocking fee. When this low γ is coupled with a high incentive to rent (high β), it is optimal (in both the main model and this extension) for the retailer to, in essence "write off" the opportunistic consumers and maintain a "static" return policy with respect to β (optimal price and restocking fee do not change with β). Thus, the expected profit from an opportunistic consumer in the main model is $f^* + s_2 - c$, while it is $f^* + s - \beta\delta - c$ in this extension. Therefore, when γ is low and β is high, further increments in β *will* have a negative impact on firm profit when salvage value is linked to the benefit of opportunism.

2.5 Conclusion

This chapter contributes to the retail operations literature by incorporating the behavior of opportunistic consumers into a retailer's optimal return policy decisions. By incorporating two opportunism-related constructs, extent and benefit of opportunism, into our analytical framework, we provide several counter-intuitive and novel managerial insights. We first demonstrate that the monopolistic retailer's optimal response to an increased number of opportunistic consumers is not always to increase the restocking fee. When the extent of opportunism is low but the benefit of opportunism is high, it is in fact better to maintain a "static" return policy. We also show that when a restocking fee is necessary, it should always be accompanied by a price decrease of a smaller magnitude.

My analysis of the profit impact of the extent of opportunism confirms the common expectation that an increase in consumer opportunism will hurt firm profits, but only when the benefit of opportunism is low. By examining the second order effects, we refine this conventional wisdom by showing that an increase in the extent of opportunism has the most detrimental impact on profits when the segment of opportunistic consumers is currently small. More interestingly, we also show that, if the retail context involves both a high benefit and a high extent of opportunism, the retailer's optimal profit can actually *increase* with the size of the opportunistic segment. This implies that actively working to reduce opportunism might not increase profits in certain business contexts. Lastly, our analysis of impact of changes in the benefit of opportunism reveals that its profit impact can be zero, positive, or negative, and we define the conditions under which each of these is the case. These results suggest that there are there are cases in which the firm might actually prefer to increase the benefit of opportunism (e.g. by adjusting the length of the trial period).

Chapter 3

Using Transactions Data to Improve Consumer Returns Forecasting and Retailer Practices

3.1 Introduction

In 2010, electronics retailer Best Buy eliminated its 15% restocking fee for consumer electronics. By offering a more lenient return policy, Best Buy was aligning itself with the majority of U.S. retailers. Walmart, for example, has long offered a 90 day full refund policy for most of its products. One consequence of such lenient policies is that *consumer returns*, i.e. items that are returned to the retailer for a full refund during the allowable time window, represent a significant cost in the supply chain[1]. They also present significant operational challenges for both OEMs and retailers, including how to process and transport returned items, how to staff returns centers, and how to recover value from returned products via new or existing sales channels.

In the Operations Management area, previous research on consumer returns has addressed the retailer's inventory replenishment policy (Ketzenberg and Zuidwijk, 2009), a retailer-manufacturer incentive alignment program designed to reduce the number of consumer returns (Ferguson et al., 2006), the OEM's consumer return allocation strategy between selling as refurbished and using for warranty demand (Pince et al., 2013), and network design for consumer returns (Guide et al., 2006). The models presented in these papers all share a common element: the requirement of

[1] According to a study by Douthit et al. (2011), the cost of consumer returns is between 5% - 6% of total revenue for an average Original Equipment Manufacturers (OEMs) in 2011.

a forecast of the flow of consumer returns. Toktay et al. (2003) report that many companies rely on simple heuristics for these forecasts, such as multiplying the projected sales by the historical return rate of a product. Our conversations with over fifteen OEMs during the 2013 Consumer Returns Annual Conference reveal that this remains the most common choice[2]. The returns forecasting literature also provides little guidance for the consumer returns problem, as it is primarily focused on forecasting the flow of end-of-use/end-of-life returns, for which an OEM might use past sales history to project returns at the end of the product's useful life. For consumer returns, however, retailers often have a potentially powerful resource for predictive modeling – transaction-level data, which typically contains detailed product and price information and, potentially, each individual consumer's previous purchase and return behavior. Despite the fact that retail transactions data is frequently shared with OEMs for forecasting new product sales (Aviv, 2007), we are not aware of it being applied to returns forecasting.

Our first contribution in this study is that we explore the power of transaction-level retail data to improve the accuracy of consumer returns forecasting. In doing so, we make a contribution to the theory of returns forecasting by addressing two econometric challenges specific to the behavior of consumers in a retail return context. First, consumers differ in how certain they are about their valuations of a product before they make a purchase. During the trial period (i.e. return time window), each consumer sufficiently experiences the product and adjusts her valuation accordingly, and the scale of this adjustment is likely to be larger for those with more initial uncertainty. This heterogeneous nature of trial uncertainty makes incorporating non-constant error variance necessary. Secondly, consumers also differ in how long it takes them to sufficiently experience a given product. Of course, the retailer cannot observe

[2]In fact, none of the OEMs we talked to suggest that their returns forecasting is conducted with retail transactions data. A list of the OEMs and retailers who attended this conference is available at www.wbresearch.com/consumerreturnsusa/home.aspx.

this duration for the products that are not returned, so the only available data on trial duration is for the products that are returned. Any estimation based solely on this "selected" sample introduces bias. We address both of these methodological challenges by constructing an econometric model that simultaneously accounts for non-constant error variance and sample selection bias.

Our second contribution is that we use our econometric results to provide guidance for retailers' consumer returns management. Some retailers have attempted to actively influence the volume of consumer returns through buyer assistance programs and/or by adjusting the return time window. The goal of the former is to reduce the probability of a return by helping customers locate the most suitable product, while the latter is meant to reduce both the number and average age of returns. For buyer assistance programs, the cost justification depends heavily on: 1) where a product is in its life-cycle and thus how familiar consumers are likely to be with it, and 2) how product assortment causes consumers to be less certain about their product choices. Our model enables us to empirically investigate the affect of these two factors on the probability that a product will be returned, and thus to suggest some guidance regarding the focus of buyer assistance. For the return time window, retailers often set a single policy uniformly across all returns, e.g. 30, 60, or 90 days after the day of purchase. This rule-of-thumb approach, while simple to understand by consumers, ignores the fact that some products take longer to evaluate than others. Using our econometric model we empirically investigate the relationship between product type and the length of time required for customers to determine their post-purchase utility.

Using data provided by a major U.S. consumer electronics retailer, we apply our econometric model to 20,801 transactions of 2,483 products. A summary of our main results is as follows. First, our causal model based forecasting approach is, on average, 20% to 40% more accurate than a variety of benchmark models that do not factor in the retailers' transaction data. This improvement in accuracy is largely sustained even when only month and category dummy variables are included, which constitutes a very parsimonious and widely available set of explanatory variables. Second, we provide guidance for buyer assistance programs by showing how factors such as product variety and product maturity are correlated with the probability of a product being returned. These findings enable retailers to better allocate scarce resources within their buyer assistance programs. Third, we provide guidance for setting the return time windows by deriving a set of reference policies based on our finding that different types of products require different lengths of time for consumers to sufficiently experience them and make an assessment of fit. Finally, we empirically validate the assumption made in Shulman et al. (2009, 2011) that trial uncertainty is convexly decreasing in product variety.

3.2 Literature Review

Our study is closely related to two emerging literature streams, managing consumer returns and product returns forecasting. In the first stream, most previous studies have an analytical focus on deriving the optimal return policies under different circumstances. Examples include Akçay et al. (2013), Shulman et al. (2009, 2011), and Su (2009), as well as the model presented in the last chapter. The current study empirically tests one key assumption made in this literature – that more product variety leads to lower trial uncertainty. On the empirical side, Anderson et al. (2009) contrasts an econometric model that incorporates consumer's buy-return sequential decisions to the conventional purchase incident model. They demonstrate how the model could be used to estimate the optional value of a full refund policy. Petersen and Kumar (2009) hypothesize and empirically confirm that consumers' past return volume links positively to their future purchase volume. In a different context, Griffis et al. (2012) show that consumers who returned in the past tend to order more frequently, put more items in an order, and buy more expensive items than those who never returned. Our research extends this line of discourse by demonstrating that past return behavior is a also good predictor of future return behavior. Thus, customers with past returns may not be as profitable as previously conjectured.

The return forecasting literature has a predominant emphasis on *end-of-use* returns. Toktay et al. (2003) provide a comprehensive review of these studies. While Toktay et al. (2000) examine the case of Kodak's single-use cameras, Clottey et al. (2012) expand on this work by considering alternative distributions for the return lag. Li et al. (2011) apply count regression techniques to forecast trade-in returns in a business-to-business context. Despite this extensive literature on end-of-use returns forecasting, to our knowledge the problem of forecasting *consumer returns* has not been addressed in the literature. In contrast to OEMs, who typically have only aggregated period-level sale and return data available, retailers often collect transaction-level data through Point-of-Sale (POS) technologies and loyalty programs, resulting in a rich data set for returns forecasting. While not their main focus, Hess and Mayhew (1997) is the only work that shows potential of utilizing transactions data to predict consumer returns flow. Specifically, they use price and product category to estimate, for each transaction, the probability of a return and the time lag until a return. We take the additional step to aggregate these two pieces of information into a forecast of consumer returns flow (discussed in detail in Section 3.5). Furthermore, our econometric model can be viewed as a more flexible model that includes Hess and Mayhew (1997) as a special case, and we show the forecasting capability of both approaches. While our full econometric model generates 20% to 40% more accurate forecasts than traditional time-series forecasting techniques, their approach captures around two thirds of this accuracy gain.

3.3 The Econometric Model

Consider a consumer who purchases an experience product from a retailer, which is recorded by the retailer as transaction i. Before purchasing the product, assume the consumer's utility of the product is $\mu_i + v_i$, where μ_i is common knowledge to both the consumer and the retailer, while v_i is the consumer's private knowledge, not observed by the retailer. The product trial, i.e. the post-purchase product evaluation by the consumer, is modeled as follows. After y_i days of experiencing the product, the consumer adjusts her utility by τ_i such that her post-trial utility is $d_i = \mu_i + v_i + \tau_i$.

There are multiple reasons that the trial uncertainty, or the variance of the trial experience, $V(\tau_i)$, might vary across transactions. First $V(\tau_i)$ should be larger when a product is in the early stages of its life-cycle, since publicly available product information such as consumer reviews becomes more abundant as a product matures. Second, $V(\tau_i)$ should be larger when there are fewer competing products, since a lack of alternatives makes it more difficult for a consumer to ascertain whether a product is the best fit for her needs (Shulman et al. 2009, 2011). Third, $V(\tau_i)$ should be larger when transactions occur during the holiday season (November and December), since products purchased during this time are more likely to be gifts and evaluating what others might like is more difficult than choosing the best product for oneself. Fourth, there could be category-specific variations in trial uncertainty, since products in different categories (televisions, cameras, computers, etc.) are quite different in terms of how easily and accurately their fit can be assessed. We elaborate on the above points further in Section 3.4 and test each one empirically in Section 3.6.

Note that both v_i and τ_i are not observable to the retailer so they will collapse into a single error term in estimation. We denote the collapsed error by ε_i and assume it follows a normal distribution with first and second moments equal to 0 and σ_i^2. The transaction specific variance, σ_i^2, allows us to explicitly model the heterogeneity in product trial uncertainty. So, d_i can be rewritten as $d_i = \mu_i + \varepsilon_i$. Let r_i be the cost of a return, which includes both the restocking fee and any perceived return hassle. If $d_i - r_i > 0$, the consumer keeps the purchased item. If $d_i - r_i < 0$, she returns it to the retailer. Hence, whether or not a product is returned can be modeled using a heteroskedastic probit regression. Since only full refunds appear in our data, r_i degenerates into a constant term that captures perceived hassle of return. In estimation, this constant collapses with the constant in μ_i, so we drop r_i to simplify notation[3]. The retailer records the consumer's return behavior through a binary variable I_i, $I_i = 1$ for returns and $I_i = 0$ for non-returns.

Next, we turn to the duration of experience, y_i. We make the common assumption that it follows a Weibull distribution[4]. Let $y_i = \lambda_i u_i$, where u_i is a standard Weibull distributed error term with duration dependence parameter ζ and λ_i is a scale function that allows experience duration to depend on exogenous factors such as product category and seasonality. This setup of y_i is referred to as Accelerated Failure Time (AFT) in the survival analysis literature. Strictly speaking, y_i is the consumer's private knowledge, but the retailer may approximate y_i using the time lag between purchase and return.

Since experience duration is observed only when an item is returned, estimating the parameters in λ_i using only return transactions may result in biased estimates. This issue is called non-random sample selection bias (Heckman, 1979), which arises due to the correlation between ε_i and u_i. We do not have *a priori* expectation on the directionality of this bias because there are reasonable arguments for both positive and negative correlation. For example, consumers who face "buyer's remorse" might be more likely to return the product and also be more likely to end the product trial early (a positive correlation between ε_i and u_i). On the other hand, a perfectionist

[3]Price is already included in μ_i. A restocking fee variable could also be included in μ_i if this fee varies across transactions.

[4]Clottey et al. (2012) apply an exponential distribution, which nests within Weibull as a special case, while Toktay et al. (2000) use a discrete analog of the exponential distribution.

might spend a long time testing the product, while her perfectionism might also cause even small mismatches between her expectations and the product's functionality to result in a product return (a negative correlation between ε_i and u_i). Thus, to obtain consistent estimates, we simultaneously model y_i and d_i and, more importantly, explicitly incorporate their conditional correlation.

Note that the conventional setup of a sample selection model that utilizes a bivariate normal distribution is not appropriate for our context because we need one marginal distribution of the bivariate distribution to be Weibull. Previous research of non-normal selection models, such as the pioneering work by Lee (1983), uses copula theory to construct flexible bivariate distributions. We adopt a simple form of the Farlie-Gumbel-Morgenstern copula family (Kotz et al., 2004; Smith, 2003). The bivariate probability density function (PDF) is as follows:

$$f_{y,d}(y_i, d_i) = f_y(y_i) f_d(d_i)\{1 + \theta[2F_y(y_i) - 1][2F_d(d_i) - 1]\} \tag{3.1}$$

The marginal PDFs are $f_y(y_i) = (\frac{\zeta}{y_i})(\frac{y_i}{\lambda_i})^{\zeta} e^{-(\frac{y_i}{\lambda_i})^{\zeta}}$ and $f_d(d_i) = \frac{1}{\sigma_i}\phi(\frac{d_i - u_i}{\sigma_i})$. The marginal Cumulative Distribution Functions (CDFs) are $F_y(y_i) = 1 - e^{-(\frac{y_i}{\lambda_i})^{\zeta}}$ and $F_d(d_i) = \Phi(\frac{d_i - u_i}{\sigma_i})$ where θ measures the correlation between y_i and d_i and falls in the interval $-1 < \theta < 1$.

The above specification has three appealing features. First, θ has the same directional interpretation as the more familiar Pearson's correlation coefficient, ρ. That is, a positive θ means that realizations of y_i and d_i on the same side of their respective medians will appear more frequently, whereas a negative θ indicates a higher chance of y_i and d_i appearing on the opposite side of their respective medians. In addition, the transformation from θ to ρ is fairly straightforward. $\rho = \theta \frac{\int F_y(y_i)[1 - F_y(y_i)]dy_i \int F_d(d_i)[1 - F_d(d_i)]dd_i}{\sqrt{Var(y_i)Var(d_i)}}$. Second, selection models derived from the FGM copula family have been reported to have a reasonable speed of convergence in estimation (Boehmke et al., 2006; Prieger, 2002). Third, when applied to duration data, using FGM copula to connect a Weibull and a normal distribution has been shown to outperform Lee (1983) approach, where a transformed bivariate normal distribution is used (Prieger, 2002).

With the bivariate distribution on hand, we can now construct the likelihood function. The commonly used two-stage estimator presented in (Heckman, 1979) is not applicable to our problem because the regression equation for y_i is not linear in λ_i (Lee, 1983). Instead, we develop a Full Information Maximum Likelihood (FIML) estimator. The retailer confronts two types of data: (1) an item is not returned ($I_i = 0$) and (2) an item is returned ($I_i = 1$) y_i days after purchase. The likelihood of observing the first type is given by

$$\Pr(I_i = 0) = \Pr(d_i > 0) = \Pr(\varepsilon_i > -\mu_i) = \Phi(\frac{\mu_i}{\sigma_i}) \tag{3.2}$$

The likelihood of observing the second type is obtained by integrating the joint density function $f_{y,d}(y_i, d_i)$ over the relevant range of d_i.

$$\begin{aligned} \Pr(I_i = 1, y_i) &= \Pr(\varepsilon_i < -\mu_i, y_i) = \int_{-\infty}^{0} f_{y,d}(y_i, d_i) dd_i \\ &= f_y(y_i)\Phi(-\frac{\mu_i}{\sigma_i})\{1 + \theta[1 - \Phi(-\frac{\mu_i}{\sigma_i})][2F_y(y_i) - 1]\} \\ &= (\frac{\zeta}{y_i})(\frac{y_i}{\lambda_i})^{\zeta} e^{-(\frac{y_i}{\lambda_i})^{\zeta}} \Phi(-\frac{\mu_i}{\sigma_i})\{1 + \theta\Phi(\frac{\mu_i}{\sigma_i})[1 - 2e^{-(\frac{y_i}{\lambda_i})^{\zeta}}]\} \end{aligned} \tag{3.3}$$

Taking Equation 3.2 and 3.3 together, we are able to describe the log likelihood of observing the whole sample as follows:

$$\ln L = \sum_{i=1}^{n} \{(1 - I_i)\ln[\Pr(I_i = 0)] + I_i \ln[\Pr(I_i = 1, y_i)]\} \tag{3.4}$$

The FIML estimator produced by maximizing Equation 3.4 is asymptotically consistent. Its finite sample performance is examined through a Monte Carlo simulation, which is available upon request. Within the framework of our simulation, we find the estimator to have little bias and to produce estimates considerably closer to the true parameters than a model not accounting for sample selectivity.

3.4 Data and Variables

As an effort to promote quality empirical research, the Institute for Operations Research and the Management Sciences (INFORMS) encourages scholars to publish unique and comprehensive data sets. The outcomes of this effort include a hotel booking data set (Bodea et al., 2009) and a multiechelon supply chain data set (Willems, 2008) published in *Manufacturing & Service Operations Management*, as well as a durable consumer electronics data set in *Marketing Science* (Ni et al., 2012) . Four features of the last data set make it especially appealing for applying our econometric model. First, studies of experience goods are often conducted in the context of consumer electronics (see Gu et al. (2012) and Obloj and Capron (2011) for examples). Second, our model requires transaction-level data, and Ni et al. (2012) suggest that theirs is, by far, the most comprehensive data set of consumer electronics transactions available. Third, researchers are specifically encouraged to work on product return issues with the Ni et al. (2012) data[5]. Lastly, and perhaps most importantly, the retailer in this data set employs a extremely generous time window policy for returns – the longest return lag in the data is 251 days, and there are several others that took place more than 150 days after purchase. Such leniency in the return time window permits estimating consumers' experience duration without truncation issues.

The original data contains 173,262 transactions provided by a major U.S. electronics chain, the identity of which, for confidentiality reasons, is not provided. All returns at this retailer are accepted open-box and given a full refund. The transactions in this data set took place between December 1998 and November 2004. We have limited our attention to the subset of the original data that is applicable to our context. Information goods (CDs, DVDs, books, etc.), accessories (SD cards, batteries, cords, etc.), and services were excluded since these are fundamentally different from

[5]More information could be found at the journal's website `www.informs.org/Community/ISMS` and in Ni et al. (2012).

durable electronics. Records for which the product description is missing/ambiguous are also excluded, as are records that do not indicate if/when the purchased item was returned. Finally, when this data set was gathered, on-line transactions were a very small fraction (less than 1%) of this retailer's sales. Therefore, we are not in a position to investigate differences between on-line and in-store return behavior, and so the on-line transactions are eliminated as well. This leaves us with 20,638 transactions across 2,481 different products.

The products in our data can be broadly divided into the six categories in Table 3.1. Overall return rates fluctuate around 10%, with the audio speaker category being the highest (15.15%) and major appliances being the lowest (5.58%). The longest average return lag is for digital video systems (16.85 days), while the shortest is for major appliances (7.41 days). In the following section we discuss how each variable in our model is coded.

Table 3.1 Descriptive Statistics of Product Categories

Category	**Return Lag**	**Return Rate**	**Products**	**Obs.**
Audio Speaker	14.28	15.15%	301	1558
Imaging Equipment[1]	11.85	10.98%	381	2085
Major Appliance[2]	7.41	5.58%	160	394
Computer[3]	10.41	8.03%	325	3499
Television	11.92	7.17%	746	5899
Digital Video System[4]	16.85	12.19%	568	7203

[1]Imaging equipment include camcorders, cameras, camera lenses, etc.

[2]Major appliances include kitchen appliances, washers and dryers.

[3]Computers include laptop, desktop, monitor, printer, scanner, etc.

[4]Digital Video System includes HD boxes, DVD players, etc.

Dependent Variables

The two dependent variables in Equation(3.4) are I_i, a dummy variable for return, and y_i, experience duration. Coding for I_i is straightforward: $I_i = 1$ for the returns and $I_i = 0$ for the non-returns. Because y_i is visible only in a sample of returned transactions, its empirical manifest is essentially return lag – the time difference between purchase and return. The histogram of return lag shows proximity to the shape of an exponential or Weibull distribution. In addition, the Nelson-Aalen cumulative hazard plot (Kalbfleisch and Prentice, 2002) exhibits a descending slope as y_i increases, contradicting the exponential distribution's constant slope pattern. This provides further "informal" evidence for the use of the more general Weibull distribution rather than the exponential. A formal statistical test is presented in Table 3.6 (ζ is significantly different from 1).

Independent Variables

The selection regression $d_i = \mu_i + \varepsilon_i$ calls for two sets of predictors, one for the mean μ_i and the other for the standard deviation σ_i of the error term. As discussed in Section 3.3, we expect that σ_i might differ across transactions. Modeling of the categorical and monthly effects is straightforward using dummy variables. The effects of product maturity and variety, however, require further discussion given their theoretical importance.

Product maturity ($Maturity_i$), or a product's length of existence on the market, is a measure of the product's available public information. As Nelson (1970) pointed out, all products are located somewhere on the search-experience spectrum. A particular product's position depends on the consumers' sources of information acquisition. If external sources such as advertising, expert opinion, and word of mouth provide vivid descriptions of the product, consumers tend to acquire information by "searching" these sources. When this information is available, a consumer is able to learn more about the product prior to purchase, so her idiosyncratic uncertainty at the time of purchase is lower. On the contrary, if public information regarding a product is relatively scarce or hard to access, consumers must rely on their own experience to determine its value, and thus idiosyncratic uncertainty is higher. One important factor that shifts a product's position on the search-experience spectrum is its length of existence. For example, comparing a newly introduced television model with one that has been on the market for several months, the amount of publicly available information (e.g. expert reviews, consumer reviews, defect rates, etc.) is much more abundant for the latter (Kalish, 1985). The accumulation of publicly available product information over time is even more apparent with the popularization of consumer forums and on-line recommendation systems (Dimoka et al., 2012). In general, as a product becomes more mature in the market, more information will be disclosed through various sources (Ghosh and Galbreth, 2013) and hence its value could be more precisely gauged without hands-on "experience". Therefore, consumers' trial uncertainty of a more mature product is expected to be lower. $Maturity_i$ is operationalized by differencing the time (in months) of the focal transaction and the first transaction for the same product. For example, 12 transactions are documented in our data set for a certain digital camera. If the 5th transaction was recorded on July 2002 and the first transaction of this camera was observed on February 2002, the maturity at the time of the 5th transaction is 5 months. Since we expect that the uncertainty reduction effect of maturity is stronger for newer products, a squared term, $(Maturity_i)^2$, is also included in σ_i.

The number of competing alternatives, $(Variety_i)$, also impacts trial uncertainty since comparing products with heterogeneous design attributes could help consumers locate the one that better matches their preferences. With more similar products being offered, customers can more easily assess their fit, which in turn reduces their reliance on experience. However, this relationship between product variety and trial uncertainty is not likely to be linear. As demonstrated by Kahn (1998) and Lehmann (1998), it becomes increasingly hard for consumers to realize the incremental benefit that each additional product represents. This convexly negative relationship between product variety and trial uncertainty is assumed in the analytical models of Shulman et al. (2009, 2011). Our analysis enables us to test this assumption empirically. Since our data includes a detailed product categorization scheme, we are able to define competing products. For example, products within the television category are grouped by screen size and features. For transaction i, $Variety_i$ is the number of all other products sold in the same subcategory in the same month as transaction i. To account for the non-linear effect of $Variety_i$, we also include its squared term, $(Variety_i)^2$. Following the standard treatment of conditional heteroskedasticity (White, 1980), we parameterize σ_i as an exponential function of the above covariates.

$$\begin{aligned} \sigma_i &= \exp[\delta_1 Maturity_i + \delta_2 Variety_i + \delta_3 (Maturity_i)^2 + \delta_4 (Variety_i)^2 \\ &\quad + \sum \delta Category_i + \sum \delta Month_i] \end{aligned} \tag{3.5}$$

In the mean function, μ_i, we include a list of potential predictors for return probability. Among them, the number of returns made in the past ($Past\_return_i$) is of theoretical interest. It has been shown in prior studies that past return behavior correlates with both the dollar amount and the frequency of future purchases (Griffis et al., 2012; Kumar et al., 2002). In contrast, our interest is whether past return behavior can predict future return behavior. Chircu and Mahajan (2006) classified the costs involved in buyer-seller interaction into price-type costs, time-type costs, and psychological-type costs. Consumers who have never made a return may feel hesitant to make the first return, incurring a psychological-type cost. Inexperienced consumers might also inherently possess higher time-type costs. For example, the higher opportunity cost of making a return trip could have filtered them into the inexperienced category initially. In sum, we believe consumers without return experience perceive higher costs of making a return, which implies a stronger intention to keep the purchased item. Thus, return experience should positively predict return probability. For the consumer who made transaction i, we count the number of products she has returned in the past, denoted $Past\_return_i$.

Other predictors of return probability included in the model are $Popularity_i$, $Other\_purchase_i$, $Past\_purchase_i$, $Past\_visit_i$, Avg_price_i, and $Purchase\_price_i$. Since consumers may assign a higher value to a more popular product, making it less likely to be returned, we include $Popularity_i$, calculated as the product's average number of monthly transactions. Since buying more than needed may lead to remorse and thus a higher likelihood of return, we include $Other\_purchase_i$, calculated as the total dollar value of other items in the same receipt. Since past purchases might reflect a consumer's loyalty to the retailer, which might in turn affect the intention to return, we include $Past\_purchase_i$, calculated as the aggregate value of items purchased by the consumer in the past six months. Since frequent visitors have more planned trips to the retailer and hence have lower opportunity cost of return, we include $Past\_visit_i$, calculated as the consumer's frequency of transactions in the past six months. Since Hess et al. (1996) found that more expensive items are more likely to be returned, we include Avg_price_i, calculated as the mean of all transactions of a certain product. Finally, to capture the impact of within-product price fluctuation, we include $Purchase\_price_i$, calculated as the percentage difference between the price of transaction i and Avg_price_i for the same product. $Other\_purchase_i$, $Past\_purchase_i$, and Avg_price_i are scaled by 1000 for easier presentation of the coefficients.

Finally, to reliably estimate the δ coefficients in Equation(3.5), we include the same variables in μ_i. In summary, the mean utility function is parameterized as follows.

$$
\begin{aligned}
\mu_i \quad = \quad & \gamma_0 + \gamma_1 Popularity_i + \gamma_2 Maturity_i + \gamma_3 Variety_i + \gamma_4 (Maturity_i)^2 \\
& + \gamma_5 (Variety_i)^2 + \gamma_6 Other\_purchase + \gamma_7 Past\_purchase + \gamma_8 Past\_visit \\
& + \gamma_9 Past\_return + \gamma_{10} Avg\_price + \gamma_{11} Purchase\_price \\
& + \sum \gamma Category_i + \sum \gamma Month_i
\end{aligned} \tag{3.6}
$$

The duration regression $y_i = \lambda_i u_i$ calls for one set of predictors for the scale parameter λ_i. Potential predictors of duration include month (e.g. since Holiday purchases might be gifts and thus not be evaluated as quickly after purchase) and category (e.g. since the effort to evaluate a product might differ by category). In addition, popular and/or mature products might have more abundant public information available, reducing the role of hands-on experience in product evaluation. Finally, since more variety means a larger pool of back-up choices, the decision to terminate the product trial might be easier. In summary, $Month_i$, $Category_i$, $Popularity_i$, $Maturity_i$, and $Variety_i$ are included, and λ_i is parameterized as follows.

$$\begin{aligned}\lambda_i &= \exp(\beta_0 + \beta_1 Popularity_i + \beta_2 Maturity_i + \beta_3 Variety_i \\ &\quad + \sum \beta Category_i + \sum \beta Month_i) \qquad (3.7)\end{aligned}$$

Descriptive statistics for the continuous independent variables are provided in Table 3.2. Methodological studies of sample selection models (see Puhani (2000) and Vella (1998) for reviews) have advised that model identification can be improved by including exogenous predictors in the selection equation that do not belong to the outcome equation. Our specifications of μ_i and λ_i closely follow this advice as the former contains all variables in the latter plus additional covariates. Among those additional covariates, Avg_price_i is shown by Hess and Mayhew (1997) to affect the choice of return but not the time of return.

3.5 Return Forecasting

Due to heterogeneity in experience duration, the total number of returns (R_t) in a given time period t could be partitioned into two parts, those attributed to sales from previous time periods ($R_{p,t}$) and those attributed to sales in the current time period ($R_{c,t}$). That is, $R_t = R_{p,t} + R_{c,t}$. The common practice in industry as well as the focus of previous literature is to use period-level sales and/or return data to directly

Table 3.2 Descriptive Statistics of Independent Variables

Variable	Mean	Std. Dev.	Min.	Max.
Other_purchase	0.35	0.69	0	12.56
Popularity	2.44	3.94	0	56
Maturity	4.87	5.03	0	45
Variety	16.84	10.43	1	62
Past_purchase	0.27	0.76	0	20.23
Past_visit	0.72	1.36	0	22
Past_return	0.68	1.71	0	28
Avg_price	0.29	0.28	0.02	5.12
Purchase_price	0	0.1	-0.88	1.05
N		20638		

predict R_t. In contrast, we exploit the rich information contained in transaction-level retailer data that is available through the retailer's POS systems. However, at the end of period $t-1$, the retailer only has transactional information on periods prior to period t, which implies that our econometric model can only predict $R_{p,t}$. To make a meaningful comparison between our forecasting approach and the existing ones, there are two options. The first and more straightforward option is to compare the forecasts of $R_{p,t}$. However, this ignores the fact that the quantity of interest in practice is R_t. The second option compares R_t predicted by a benchmark model with the sum of $R_{p,t}$ predicted by our econometric model and $R_{c,t}$ predicted by the same benchmark. As a result, this option requires a set of pair-wise comparisons, the number of which is the same as the number of benchmarks. To ensure a close tie to practice, we proceed with the second option.

Before introducing the specifics of our benchmark and focal models, we discuss how period-level variables are generated from the original data set. Past research and our interactions with practitioners indicate that firms perform return forecasting on a weekly to a monthly basis. Thus, the duration of a period is specified as either one week, two weeks, or one month. We calculate sales quantity, Q_t, along with R_t, $R_{p,t}$, and $R_{c,t}$ for each period. The complete sample of our data has a time span of six years, and we reserve the last two years as the prediction sample. The exact number of periods for each specification along with other descriptive statistics are provided in Table 3.3.

Table 3.3 Description of Period-Level Data

Length of t	Total Periods (T)	Predict Sample	Avg. Q_t	Avg. R_t	Avg. $\frac{R_{p,t}}{R_t}$
one week	313 periods	104 periods	65.93	6.61	68.05%
two weeks	156 periods	52 periods	131.97	13.26	52.44%
one month	72 periods	24 periods	286.64	28.72	35.67%

A concern with the approach described above is that there is only one replication for each length of time period we study. To test the robustness of our results, we create four additional replication samples. First, the entire six-year data set is randomly split into four parts, each containing roughly one fourth of the original data. Second, we omit the parts one at a time to construct four subsamples: Subsample 1 – Subsample 4. Third, we calculate Q_t, R_t, $R_{p,t}$, and $R_{c,t}$ for every subsample, where t is weekly, biweekly, and monthly. Descriptive statistics are presented in Table 3.4. As expected, Q_t and R_t are roughly three-fourth of those in Table 3.3.

Table 3.4 Description of Period-Level Data in Replication Samples

	Subsample 1			Subsample 2		
Length of t	Avg Q_t	Avg R_t	Avg $\frac{R_{p,t}}{R_t}$	Avg Q_t	Avg R_t	Avg $\frac{R_{p,t}}{R_t}$
one week	49.51	4.96	68.06%	49.49	5.02	68.73%
two weeks	99.10	9.96	52.61%	99.07	10.06	53.33%
one month	215.26	21.56	35.31%	215.18	21.79	34.88%
	Subsample 3			**Subsample 4**		
Length of t	Avg Q_t	Avg R_t	Avg $\frac{R_{p,t}}{R_t}$	Avg Q_t	Avg R_t	Avg $\frac{R_{p,t}}{R_t}$
one week	49.23	4.85	68.03%	49.55	5.01	67.39%
two weeks	98.56	9.72	52.29%	98.56	10.04	51.52%
one month	214.07	21.06	36.21%	215.40	21.76	36.19%

Next, we introduce the setup of the benchmark models and the forecast produced by our econometric model. To simplify exposition, we assume that the length of a period is one week. Full extension to the biweekly and monthly cases is provided in the Appendix F. The benchmark models incorporate a seasonality component in the monthly case.

Benchmark 1: Baseline Model. The first benchmark does not employ any formal statistical modeling. Since R_t is essentially a fraction of the total sales (Q_t) in period t, it can be estimated as the product of the average past return rate and the sales forecast ($\widehat{Q_t}$). Because forecasting demand is not the focus here, we replace $\widehat{Q_t}$ with Q_t:

$$\widehat{R_t^{Baseline}} = (\frac{1}{t-1}\sum_{j=1}^{t-1}\frac{R_{t-j}}{Q_{t-j}}) \times Q_t \tag{3.8}$$

The above specification gives the baseline model a considerable advantage. Therefore, its *de facto* forecasting accuracy should be much lower than what we show.

Benchmark 2: Exponential Smoothing Model. Because we do not see an obvious upward or downward trend in R_t, a simple exponential smoothing model is used. The second benchmark is as follows.

$$\widehat{R_{p,t}^{Smoothing}} = \alpha R_{t-1} + (1-\alpha)\widehat{R_{t-1}^{Smoothing}} \tag{3.9}$$

where α is the smoothing parameter. The starting value of the smoothing series, $\widehat{R_1^{Smoothing}}$, is set to the actual returns of period 1, R_1.

Benchmark 3: ARIMA Model. Before fitting an ARIMA(p,d,q) model to our data, we first need to determine the appropriate number of parameters. As no trend is detected, d is set to zero. Analyzing the correlogram and partial correlogram, we count the number of significant moving average lags (5) and autoregressive lags (1). Then, we fit 5×1 different ARIMA models. A Bayesian Information Criteria (BIC) is recorded for each model and the model with the lowest BIC value is chosen, in our case ARIMA(1,0,1).

$$\widehat{R_t^{ARIMA}} = \alpha_0 + \alpha_1 R_{t-1} + \alpha_2 e_{t-1} + e_t \tag{3.10}$$

Benchmark 4: Lagged Sales Model. In this benchmark, we adopt the technique from Toktay et al. (2000) – regressing current returns on past sales. The rationale is that a given proportion, α_j, of transactions made j periods prior to t will be returned in period t. Thus, $\alpha_j Q_{t-j}$ is period $t-j$'s contribution to R_t. A central question for this approach is how far to look back, i.e. how large j should be. The answer to this question marks the contextual difference between our study and the literature. The empirical context of Toktay et al. (2000) is a remanufacturer collecting *end-of-use* returns, which have an estimated mean return lag of 2 months. Additionally, (De Brito and Dekker, 2003, p.237) suggest that returns continue to take place even two years after sales for some industrial materials. In comparison, we are examining *consumer returns* received by a retailer, which typically have a short return lag. While the mean return lag in our data is 2 weeks, its 99th percentile is 126 days or 18 weeks. Therefore, it is enough to include Q_{t-1} through Q_{t-18} as regressors. Adding an intercept, the lagged sales model is as follows.

$$\widehat{R_t^{LagSales}} = \alpha_0 + \sum_{j=1}^{18} \alpha_j Q_{t-j} + e_t \tag{3.11}$$

When forecasting, we drop the insignificant α_j coefficients to avoid poor forecasts due to over-fitting. The above model estimates at most nineteen α_js and the variance of e_t. The total consumption of twenty degrees of freedom is reasonable considering the size of our estimation sample[6]. Thus, we are exempted from making distributional assumptions on the α coefficients as in Toktay et al. (2000).

Benchmark 5: ARIMAX Model. We combine the two benchmarks above to formulate a lagged sales model with ARMA(1,1) error structure. Such temporal regression belongs to the ARIMAX model family in the time series literature.

$$\widehat{R_t^{ARIMAX}} = \alpha_0 + \sum_{j=1}^{18} \alpha_j Q_{t-j} + \kappa_t, \textit{ where } \kappa_t = \alpha_{19}\kappa_{t-1} + \alpha_{20} e_{t-1} + e_t \tag{3.12}$$

[6]To make the first prediction, we have $313 - 100 = 213$ data points. The lagged sales go back 18 periods, which reduces 18 more observations. Each subsequent prediction will have one more observation available as historical data.

Similar to (3.11), we drop the insignificant α_j coefficients to avoid poor forecasts due to over-fitting.

The above five benchmarks exploit period-level data to different degrees. While the baseline model produces a naïve forecast simply using the historical average, the ARIMAX model incorporates past return and sales data in a fairly sophisticated fashion. These five benchmarks represent the forecasting methodology either used in the industry or available in the literature. They also are representative of what OEMs can reasonably do to forecast product returns in the absence of having transactions data from the retailer. Next, we derive a return forecast using our econometric model.

Focal Model: Partial I, Partial II, and Full. As discussed earlier, our econometric model predicts $R_{p,t}$, the number of returns attributed to the previous periods' sales. To make a tight connection to the literature and ensure practicality for retailers, we parameterize our econometric model in three ways, which leads to three versions of $\widehat{R_{p,t}}$.

As discussed in the literature review, the empirical model of Hess and Mayhew (1997) may be used to forecast the flow of product returns. For a given purchase, Hess and Mayhew (1997) predict both the probability that a product will be returned (using price and product category[7] as explanatory variables) and the time lag until a return occurs (using price as an explanatory variable). However, they do not consider heteroskedasticity and do not correct for sample selection bias. To closely match their specification, we include product category dummies in λ_i and transaction prices in both λ_i and μ_i, while constraining σ_i to 1 and θ to 0. This version of $\widehat{R_{p,t}}$ is denoted as *Partial I*.

We acknowledge that the wealth of transactional data used in our full model might not always be available, either because it is not collected by the retailer or because

[7]Hess and Mayhew (1997) classified products into three broad types – low fit, medium fit, and high fit. Since the exact classification criteria are not discussed, we instead use dummy variables for product category.

the forecasting is being done by an OEM with whom all of the retailer's transactional data is not shared. To address this practical concern, we present two versions of $\widehat{R_{p,t}}$, *Partial II* and *Full*. In the *Partial II* setup, only category and month dummies are included in λ_i, μ_i, and σ_i. In the *Full* setup, a complete parameterization of λ_i, μ_i, and σ_i as in (3.5), (3.6), and (3.7) is used. While *Partial II* serves to examine the forecasting capability of our econometric model with minimal data input, the *Full* setup provides insights into the additional predictive value of other variables.

Next, the three versions of $\widehat{R_{p,t}}$ are joined with the $R_{c,t}$ forecasts estimated by the five benchmarks, such that $\widehat{R_t} = \widehat{R_{p,t}} + \widehat{R_{c,t}}$, to construct pair-wise comparisons. Replacing R_t with $R_{c,t}$ in (3.8) - (3.12), we obtain $\widehat{R_{c,t}^{Baseline}}$, $\widehat{R_{c,t}^{Smoothing}}$, $\widehat{R_{c,t}^{ARIMA}}$, $\widehat{R_{c,t}^{LagSales}}$, and $\widehat{R_{c,t}^{ARIMAX}}$. Then, we calculate $\widehat{R_{p,t}}$ using our econometric model in five steps.

1. Denote for each transaction the date of sale as date_i^{sale} and for each period the starting date and ending date as date_t^{start} and date_t^{end}.

2. Run the maximum likelihood estimation in (3.4) using all the transactions made on or before 126 days prior to date_t^{start}. Omitting the final 126 days ensures that the retailer has return lag data available on all the transactions used for estimation.

3. Among the transactions made between 126 days prior to date_t^{start} and date_t^{start}, denote by N_t the total number of those that have not been returned yet.

4. Calculate for each transaction in N_t the probability of the product being returned between date_t^{start} and date_t^{end} given that it is not returned before date_t^{start}. For transaction i, this probability is $\Pr(d_i < 0, a < y_i < b | d_i > 0 \cup y_i > a)$,

where $a = \text{date}_t^{start} - \text{date}_i^{sale}$ and $b = \text{date}_t^{end} - \text{date}_i^{sale}$. Using (3.1), we have

$$\begin{aligned} & \Pr(d_i < 0, a < y_i < b | d_i > 0 \cup y_i > a) \\ = & \frac{\Pr(d_i < 0, y_i < b) - \Pr(d_i < 0, y_i < a)}{1 - \Pr(d_i < 0, y_i < a)} \\ = & \frac{\int_{-\infty}^{b} \int_{-\infty}^{0} f_{y,d}(y_i, d_i) dd_i dy_i - \int_{-\infty}^{a} \int_{-\infty}^{0} f_{y,d}(y_i, d_i) dd_i dy_i}{1 - \int_{-\infty}^{a} \int_{-\infty}^{0} f_{y,d}(y_i, d_i) dd_i dy_i} \end{aligned} \tag{3.13}$$

5. Calculate the expected number of returns, $\widehat{R_{p,t}}$, in period t. Because each transaction in N_t has a non-zero chance of being returned in period t and these probabilities are not equal, the total number of returns follows a Poisson-binomial distribution. Its mean is the sum of N_t Bernoulli probabilities, or

$$\widehat{R_{p,t}} = \sum_{i=1}^{N_t} \Pr(d_i < 0, a < y_i < b | d_i > 0 \cup y_i > a) \tag{3.14}$$

All models, focal and benchmark, are estimated after each time period to allow for incorporation of the latest sale, return, and transaction information. We employ two prediction accuracy measures for evaluating the relative forecasting capability of different models. The mean absolute deviation (MAD) reflects the average bias of forecasts:

$$MAD = \frac{1}{T} \sum_{t=1}^{T} |\widehat{R_t} - R_t| \tag{3.15}$$

while the mean squared error (MSE) penalizes the large errors more severely:

$$MSE = \frac{1}{T} \sum_{t=1}^{T} (\widehat{R_t} - R_t)^2 \tag{3.16}$$

All benchmark models are estimated using STATA's existing routines, "tssmooth" for exponential smoothing, "arima" for ARIMA and ARIMAX, and "regress" for lagged sales. The econometric model is programed with STATA's ml language. A summary of the pair-wise comparisons between the benchmark and focal models is presented in Table 3.5. The detailed results of each forecasting approach are available upon request.

Table 3.5 Average Forecast Accuracy Improvements

	Focal Model					
	Partial I		Partial II		Full	
Benchmark	MAD	MSE	MAD	MSE	MAD	MSE
Baseline	13%	36%	20%	41%	19%	41%
Smoothing	18%	29%	23%	36%	26%	41%
ARIMA	14%	26%	20%	33%	23%	38%
Lagged Sales	13%	17%	20%	27%	21%	30%
ARIMAX	3%	10%	10%	19%	15%	25%

The numbers show average improvements made by focal models over benchmarks across 15 cases.

Table 3.5 shows clear evidence for the improved performance of our model relative to the benchmark models. Recall that we have varied time length per period and sample composition to construct 15 cases for comparison. All the benchmarks together only outperform our model in one of these 15 instances. Specifically, in Subsample 4 of the one-month-per-period case, the ARIMAX model has a lower forecasting error. Moreover, the accuracy gain from our model is substantial. Comparing with the various benchmarks, our model *consistently* achieves 20% lower MAD and 35% lower MSE. While all models exhibit diminishing forecasting accuracy as the prediction horizon increases from one week to one month, our model sustains an impressive MAD and MSE performance improvement. In fact, our model's MSE performance relative to the benchmarks does not change substantially. This is an especially strong result given that only 35% of the returns in a given month come from the previous months' sales (Table 3.3). Furthermore, the Partial II and Full specifications perform significantly better than the Partial I specification. Therefore, we conclude that, by exploiting the transaction-level data, our econometric model shows encouraging performance improvements over the existing approaches. Retailers as well as their upstream OEMs who conduct operational planning on a weekly or a monthly basis may achieve considerable forecasting accuracy gains using our proposed method.

Surprisingly, performance of the parsimonious Partial II model is only marginally lower than that of the Full model. For all 15 comparisons, using a succinct set of variables (category and month dummies) is not substantially inferior to using the complete variable set. Furthermore, the homogeneity between Partial II and Full is robust to both MAD and MSE measures. This result is encouraging for retailers who only track minimal consumer behavior data, as the product and time data required for the Partial model can be collected from any modern Point-of-Sale (POS) system. Thus, the Partial II model essentially offers considerable advantage over the benchmarks with no additional data collection cost beyond a current POS system (i.e. no need for a customer loyalty program). From a methodological perspective, the performance proximity of Partial II and Full models also demonstrates that our transaction-level modeling approach, not the specific choice of variables, contributes the most to forecasting accuracy.

Among the benchmarks, the ARIMAX model appears to have the best overall performance followed by the lagged sales model[8]. While the lagged sales model has performance nearly identical to the ARIMAX model in the one week case, the latter's advantage becomes more obvious as the time horizon increases. Recall that both models exploit the correlation between past sales and current returns. Therefore, if only period-level sales and return data are available (e.g. for an OEM who does not have access to the retailer's POS data), regressing returns on past sales is still the best choice. This observation complements Toktay et al. (2000) in suggesting that a lagged sales model or an ARIMAX model is the appropriate method for predicting not only *end-of-use returns* but also *consumer returns*, provided that no transaction-level data is available.

[8]The baseline model is excluded from this comparison between benchmarks, since it uses actual sales data instead of a forecast, which offers unfair advantage.

Table 3.6 Estimation Results

	Naive Model ($N = 2010$)		Full Model ($N = 20047$)					
	Duration: Scale $\ln(\lambda_i)$		Duration: Scale $\ln(\lambda_i)$		Selection: Mean μ_i		Selection: Sigma $\ln(\sigma_i)$	
Variables	Estimate	Robust SE	Estimate	Robust SE	Estimate	Robust SE	Estimate	Robust SE
Month Dummies	Base is January. February - October are omitted for succinct exposition.							
November	0.417**	(0.129)	0.426***	(0.124)	0.0645	(0.0614)	0.299**	(0.106)
December	0.450***	(0.101)	0.464***	(0.0933)	0.0917	(0.0585)	0.397***	(0.0762)
Category Dummies	Base is audio speaker systems.							
Imaging	-0.0620	(0.150)	-0.0435	(0.137)	-0.554***	(0.121)	-1.252**	(0.444)
Majors	-0.544*	(0.261)	-0.426	(0.270)	0.507	(1.081)	0.0616	(0.794)
Computer	-0.307*	(0.152)	-0.286*	(0.134)	0.0291	(0.284)	-0.269	(0.346)
Television	-0.204	(0.136)	-0.160	(0.122)	-0.457***	(0.139)	-1.112**	(0.345)
Digital Video System	0.163	(0.139)	0.147	(0.125)	0.146	(0.343)	0.165	(0.358)
Popularity	0.0176+	(0.00918)	0.0175+	(0.00936)	-0.000218	(0.00179)		
Maturity	-0.00810	(0.00600)	-0.00822	(0.00564)	-0.00463	(0.00394)	-0.0321***	(0.00850)
Variety	-0.00994*	(0.00413)	-0.00836*	(0.00394)	-0.0131*	(0.00561)	-0.0372***	(0.00771)
(Maturity)^2					0.00000703	(0.000162)	0.000952*	(0.000376)
(Variety)^2					0.000289*	(0.000125)	0.000669***	(0.000144)
Other_purchases					-0.128**	(0.0448)		
Past_purchase					0.0298	(0.0207)		
Past_visit					0.0150*	(0.00712)		
Past_return					-0.0245*	(0.0108)		
Avg_price					0.0589+	(0.0340)		
Purchase_price					0.117	(0.0771)		
Constant	2.526***	(0.156)	3.020***	(0.140)	0.932***	(0.0728)		
Druation Dependence ζ	0.839***	(0.0149)	0.906***	(0.0147)				
Error Correlation θ			0.855***	(0.0245)				

+p<0.1 *p<0.05 **p<0.01 ***<0.001. Robust standard errors assume transactions clustered within products.

3.6 Managerial Implications

Through improving the accuracy of returns forecasting, retailers can better align their strategic and tactical responses with the increasing volume of consumer returns by better matching supply with demand. In addition to this "passive" approach, some retailers also "actively" search for ways to influence returns (Toktay et al., 2003), which include investing in buyer assistance programs and adjusting the return time window. In the following, we use our econometric model to derive managerial implications for these more proactive approaches. We also discuss the implications of our empirical findings to previous consumer returns studies.

The richness of our data affords us to construct an extensive list of independent variables. While some of these, such as $Maturity_i$, $Variety_i$, and $Past\_return_i$, are theoretically interesting, others are more related to the operations of the retailer and therefore can suggest new insights for retail managers. Since one central feature of our model is sample selection, we compare the model with sample selection (full model) to the one without (naïve model[9]), with both models fitted to data excluding the last 126 days to avoid the right truncation issue. The naïve model is estimated using Stata's "streg" routine. Furthermore, due to the transaction-within-product structure of our data, we present the cluster robust standard errors for all parameter estimates. Results are presented in Table 3.6.

Return Time Window

When determining the appropriate return time window, a retailer considers numerous factors including returns processing, value depreciation, and the time window's impact on demand. A very lenient policy, such as the one represented in our data, does not

[9] The naïve model assumes consumers who return their purchases take the same amount of time to evaluate products as those who do not return. Therefore, it estimates only the return lag regression and generalizes to the whole sample.

put any constraint on trial duration and thus, *ceteris paribus*, it should attract the highest demand. On the other hand, this leniency might also induce more frequent returns and/or returns that have lost a significant portion of their value (especially true for consumer electronics) by the time a product is eventually returned. While our goal in this study is not to derive an optimal return window, which would require assumptions regarding the retailer's trade-offs as well as how consumers react to different time windows, we do provide some potentially valuable insights into this complex managerial decision. For example, our results provide guidance on whether a given time window is likely to affect consumers' trial experience – i.e., whether the time window is too short to allow the consumer to fully assess the product's fit or too long such that the window can be shortened with minimal impact on consumer demand. As an example, we consider a time window, η, that corresponds to the 95th percentile of experience duration for a typical consumer[10]. That is, $\Pr(y < \eta|\overline{\lambda}, \widehat{\zeta}) = 1 - e^{-(\frac{\eta}{\lambda})^{\widehat{\zeta}}} = 0.95$. Solving for η, we have $\eta = \overline{\lambda}(\ln 20)^{\frac{1}{\widehat{\zeta}}}$. Because experience duration has been postulated to differ significantly across product categories and in the holiday season, we estimate η for each category-season combination. Results are presented in Table 3.7. We acknowledge that, in practice, an easily communicated return policy (e.g. a uniform 90-day window) has the advantage of being easy to understand. However, our results suggest that the mild customizations demonstrated in Table 3.7 could help retailers better manage the trade-offs when setting return time windows.

Return Probability

In Section 3.4, we provided reasons why product maturity, variety, and the buyer's past return experience might affect her current return decision. Table 3.6 shows

[10]The typical consumer has all covariates equal to their sample averages. This approach is sometimes called evaluating margin at means.

Table 3.7 Time Windows Equivalent to 95th Percentile of Experience Duration

Category	Non-holiday (days)	Holiday (days)
Audio Speaker	60.95	93.41
Imaging Equipment	59.25	90.81
Major Appliance	38.62	59.19
Computer	46.17	70.76
Television	53.03	81.26
Digital Video System	72.47	111.06

Non-holiday: Jan. through Oct.; Holiday: Nov. and Dec.

statistical significance of these effects based on the Wald test, which is consistent under the more formal likelihood ratio test. To provide more specific insights, we also estimate the marginal effects of $Maturity_i$, $Variety_i$, and $Past\_return_i$ on return probability (this estimation is needed since their coefficients do not have the same interpretation as in a linear regression model). Consider product maturity for illustrating the estimation procedure. Assume we want to know the amount of decrease in return probability after a new product has been on the shelf for six months. For transaction i, $\Pr(d_i < 0|Maturity_i = 0) - \Pr(d_i < 0|Maturity_i = 6)$ is the quantity of interest. However, this quantity is heterogeneous across the sample. Thus, we compute an average as follows,

$$\Delta_{Maturity} = \frac{1}{N}\sum_{i=1}^{N}\Phi(-\frac{\widehat{\mu}_i}{\widehat{\sigma}_i}|Maturity_i = 0) - \frac{1}{N}\sum_{i=1}^{N}\Phi(-\frac{\widehat{\mu}_i}{\widehat{\sigma}_i}|Maturity_i = 6)$$

where $\Delta_{Maturity}$ is the Average Marginal Effect (AME) of product maturity. Point estimates and standard errors of return probability evaluated at $Maturity_i = 0$ and $Maturity_i = 6$ along with $\Delta_{Maturity}$ are shown in Table 3.8 under the column headings of estimate, S.E., and AME, respectively. We note that the specific values of maturity (i.e. 0, 6, and 12) are arbitrarily chosen to facilitate interpretations. The AMEs of $Variety_i$ and $Past\_return_i$ are also provided in Table 3.8[11]. These results are used

[11]Examining the cumulative distribution of $Variety_i$, we see that it reaches the first quartile at around 5, second at around 15, and the third at around 25. Thus, we use 5, 15 and 25 as

together with the coefficients in Table 3.6 to draw the managerial insights discussed next.

Table 3.8 Average Marginal Effects

	Return Probability		
	Estimate	S.E.	AME
Maturity			
New	0.114	0.004	
6 months	0.092	0.003	-0.022(-19.41%)
12 months	0.085	0.004	-0.029(-25.36%)
Variety			
5	0.124	0.005	
15	0.102	0.002	-0.022(-17.70%)
25	0.088	0.004	-0.036(-29.32%)
Past_return			
None	0.093	0.002	
Once	0.103	0.002	0.010(10.21%)
Twice	0.113	0.003	0.020(21.60%)

Standard errors are derived using Delta method.

The negative sign of $Maturity_i$ and positive sign of $(Maturity_i)^2$ in the skedastic function $\ln(\sigma_i)$ (see Table 3.6) offer support for our argument that a consumer's trial uncertainty is convexly decreasing in product maturity. That is, while experience uncertainty will continue to drop throughout a product's life cycle, the steepest decline occurs in the first few periods. Furthermore, through its link with trial uncertainty, maturity also negatively relates to return probability. This gives a quantitative idea of how public information gradually replaces the role of hands-on experience in product evaluation. This finding has interesting implications for recent work on processing product returns. For example, in a recent study, Pince et al. (2013) examined how an OEM should deploy returned items between warranty demand and refurbished products. They find that the optimal strategy is sensitive to the return rate. Therefore, taking the maturity-return relationship into account, the OEM should adjust its

representative values. Regarding $Past\_return_i$, 90% of the transactions are made by consumer with twice or fewer return instances in the past. As a result, we use 0, 1, and 2 as representative values.

allocation strategy as the focal product ages. In another recent paper, Ketzenberg and Zuidwijk (2009) incorporated product returns into existing inventory models assuming a constant return rate over time. Our results suggest that the incorporation of a decreasing return rate would be an interesting extension of this work.

Remark 3.1. Product maturity has a negative relationship with both consumer's trial uncertainty and return probability.

Regarding product variety[12], we observe a similar convexly decreasing relationship with trial uncertainty – a negative sign for $Variety_i$ and a positive sign for $(Variety_i)^2$ in $\ln(\sigma_i)$ (see Table 6). In the multi-product retail return models of Shulman et al. (2009, 2011), consumers are assumed to be more certain about their product evaluation when more alternatives exist. The above result offers empirical support for this modeling assumption. Interestingly, our analysis also shows that more variety leads to lower return rates – for example, a product with 25 competing alternatives is 29.32% less likely to be returned than one with only 5 alternatives. This indicates another potential benefit of more product variety that should be included in product assortment decisions.

Remark 3.2. Product variety has a negative relationship with both consumer's trial uncertainty and return probability.

Next, we use the above two remarks to derive managerial insights for "buyer assistance" programs. In practice, retailers can implement buyer assistance programs to help consumers resolve their uncertainties about products as much as possible before making a purchase (Douthit et al., 2011). For example, sales personnel might spend extra time with consumers and listen to their needs thoroughly before recommending a product (Ferguson et al., 2006). In addition, more samples may be displayed on the

[12]Our data does not contain any truly innovative product categories, such as tablet computers. Therefore, we do not expect product variety and maturity to be correlated, which is also supported by the −0.006 Pearson correlation coefficient between the two variables.

floor and the full array of similar products might be arranged more effectively on the shelf (Ofek et al., 2011). The hope in offering these programs is that the cost of these programs will be offset by a drop in return rates (as well as perhaps an increase in sales and customer sanctification). Our empirical findings suggest that these buyer assistance investments provide larger benefits to newer products as well as categories that offer fewer similar choices.[13]

Finally, our AME analysis suggests that a consumer's past return experience exerts a strong impact on future return tendency. Switching from zero past returns to only one past return increases the probability of a future return by 10.2%, from 0.093 to 0.103, and further increases in return instances lead to additional increases in the return probability. While returns are commonly regarded as a cost center, retailers have also been advised to not overlook their positive link to long-term profits. Both Griffis et al. (2012) and Petersen and Kumar (2009) show a positive correlation between past return behavior and future purchases. We suggest that the impact of consumers' return experience on retailer profits might be less straightforward – return-experienced consumers might make more purchases, but they also tend to negate a higher proportion of those revenues by returning more.

Remark 3.3. A consumer's past return instances is positively related to her future return probability.

3.7 Conclusion

In this study we have developed an econometric model that enables retailers to leverage transaction-level data to forecast consumer returns by simultaneously predicting whether and when a purchased item will be returned. When tested on a holdout sample of returns data from a large consumer electronics retailer, our approach generates

[13]We remark that this finding should be treated with some caution, since our data did not include information on the *current* focus of any buyer assistance programs at the retailer, and thus endogeneity issues cannot be ruled out.

on average 20% to 40% lower forecast error than a benchmark set of time-series methods that only use data currently available to the OEMs. Improved forecast accuracy is very important to OEMs who operate refurbishing facilities because it allows them to better match supply with demand for parts inventory and staffing decisions. In addition to this practical contribution of a superior forecasting algorithm, our model is also useful for the research community. Since many models of retail returns processing use return forecasts as a key input, our substantial improvement in accuracy over benchmark methods makes our model an important contribution to the retail operations literature.

Interpreting the estimation results of our econometric model allows us to also provide guidance for retailers regarding how to: 1) target buyer assistance programs to items with the highest unresolved consumer uncertainty, and 2) set return time windows that are sufficient to capture the experience duration of most consumers without being overly lenient. Finally, we provide empirical support for the hypotheses that product maturity, variety, and previous return history all have a significant impact on return probability.

Chapter 4

Estimating the Value of a Money-Back-Guarantee Policy in On-line Retailing

4.1 Introduction and Literature

Money-Back-Guarantee (MBG) return policies are common for retailers of experience goods, such as consumer electronics and fashion apparel, that require some "hands-on experience" by consumers to fully evaluate whether an item will fit their needs. Given the ubiquity of MBG policies, managing consumer returns in recent years has emerged as one of the most challenging issues for retailers. The National Retail Federation (2011a) reports that the average consumer return rate in the retail sector is 8.5% while for consumer electronics this figure is between 10% and 20%, with returns processing costs running as high as 3% of revenues (Douthit et al., 2011). While these returns are costly, consumers value the opportunity to return an item that does not fit their needs. Thus, the basic trade-off for a retailer is between the benefit of consumers' increased product valuation from the MBG and the cost of processing returns[1]. While much is known regarding the costs of returns, the benefits are more difficult to quantify, and the academic research is largely silent on this point. In this study, we address this research gap by empirically measuring the additional

[1]While some OEMs may provide full refunds to retailers for consumer returns, the extent of the refund is typically negotiated for each OEM and, regardless of the refund amount, the retailer always incurs some of the processing costs. There are several recent examples of retailers having to tighten their return policies because of escalating costs. See for example Grind (2013).

price that on-line customers are willing to pay for a product with an MBG versus an identical product without an MBG. Thus, we provide insights into the value of an MBG in an on-line context.

On-line channels exhibit some unique characteristics that may lead to return policies being viewed differently than at traditional stores. First, consider shipping fees. To be equivalent to a true MBG policy at a traditional store, an on-line store should offer free forward shipping, free return shipping, and a refund of the entire product price. An examination of the MBG policies of major retailers' on-line stores and manufacturers' direct channels, however, reveals that return shipping is often the responsibility of consumers[2] and any positive forward shipping charge is not refunded upon return (see Table 4.1). Since a return shipping charge is paid for by the consumer rather than being set by the retailer, it is less interesting from a managerial perspective and can be viewed as an exogenous cost of returning an item that is relatively constant across all consumers. Non-refundable forward shipping charges, however, vary widely across sellers and products, and are relatively easy for a retailer to adjust[3]. This leads to interesting research questions such as whether consumers perceive forward shipping as a "latent restocking fee" and how this affects their value of an MBG policy. Thus, we do not consider the heterogeneity in return shipping cost, but instead focus on the effect of *forward shipping* cost on the value of an MBG.

A second uniqueness of on-line retailing is the lack of a physical interaction, which makes it difficult for consumers to observe how a seller fulfills the MBG promise. In such a context, consumers often rely on the feedback of other customers (e.g. reviews of a seller) as a proxy for the service orientation of the seller. It follows that such

[2]For apparel products, return shipping is sometimes covered by the retailer. Zappos and Urban Outfitters are two well-known examples. Furthermore, if consumers receive a wrong item, return shipping is typically covered by the retailer.

[3]The actual forward shipping charge *incurred* by a retailer is primarily driven by its agreement with the shipping carrier. However, the shipping fee it *charges* to the consumers is at its own discretion.

reputational information may impact the value of an MBG, but the magnitude of this impact is not well-understood.

Table 4.1 Money-Back-Guarantee Policies of On-line Consumer Electronics Retailers

Company	Refund for product price	Refund for forward shipping (if not free)	Who pays for return shipping
Panel A[1]: online stores of major retailers			
Wal-Mart	full refund	no refund	consumer
Target	full refund	no refund	retailer
Amazon[2]	full refund	no refund	consumer
Best Buy	full refund	not clear	consumer
Sears	maybe 85%[3]	no refund	consumer
Panel B: direct channel of representative manufacturers			
Dell	maybe 85%[3]	no refund	consumer
HP	full refund	no refund	consumer
Sony	maybe 85%[3]	no refund	consumer
Lenovo	85%	no refund	consumer
Shure	full refund	no refund	consumer
Panel C: our data source			
eBay[4]	mostly full refund	no refund	mostly consumer

[1] We include the top five consumer electronic retailers on the "2013 Top 100 Retailer" chart published by National Retail Federation. Since Amazon is on the chart, it is included although it does not have a brick-and-mortar store.

[2] Most of the Amazon marketplace sellers use the same return policy as Amazon.

[3] The retailer in general offers full refund on product price. However, it might charge up to 15% depending on the condition of the returned item.

[4] In principle, sellers on eBay could specify their own MBG terms. Our observation is that most sellers fully refund the product price but do not refund shipping charges.

In addition to the practical need of retailers to quantify the benefit of MBGs, there is also a theoretical interest in deriving the optimal MBG policies under different market and product conditions. Analytical studies of consumer returns such as Akçay et al. (2013), McWilliams (2012), Shulman et al. (2009, 2010, 2011) and Su (2009) assume that consumers are willing to pay more for a product if it comes with an MBG policy. Our study complements this research stream by providing an approach to estimate this increment in consumers' willingness-to-pay (WTP). To our knowledge, the only existing study that empirically estimates the value of an MBG policy is Anderson et al. (2009), who examine a mail-order catalog retailer of apparel products. Our study provides three new contributions to this literature. First, we explore the on-line retailing context, including the unique roles of forward shipping and seller reputation in determining the value of MBGs. Second, both our approach to the estimation problem and the type of data analyzed are quite distinct from Anderson et al. (2009), because our focus is on consumer electronics while theirs was on apparel products. The fact that identical consumer electronic products are frequently sold by a variety of retailers facilitates our third contribution. That is, we actually observe purchases from retailers with and without MBG policies rather than using a counterfactual analysis as in Anderson et al. (2009).

In practice, the value of an MBG policy is hard to quantify because most retailers are consistently lenient in accepting returns and hence one does not directly observe the consequences of a no-MBG policy. Anderson et al. (2009) tackle the estimation problem by fitting a structural model to a catalog retailer's panel data, such that while all consumers in the data set made purchase and return decisions in the presence of an MBG policy, their potential response to the absence of an MBG policy is approximated by a counterfactual analysis. We introduce an alternative approach for the estimation problem by exploiting some of the uniqueness of the transactions on eBay.[4] The advantage of the eBay platform is that multiple sellers offer identical products, some with an MBG and some without. Furthermore, for those having an

[4]The approaches presented in Anderson et al. (2009) and the current study are suitable for different products, depending on the retailer's current assortment. For example, if a product has been carried by a retailer for some time, Anderson et al. (2009) has the advantage of estimating the value of MBG from the retailer's own data. Our approach, in contrast, can be used for any product, provided that it is widely offered on eBay. In situations where both approaches are applicable, one could be used to corroborate the findings from the other.

MBG policy, forward shipping charges vary widely, from free to as high as 50% of the total price paid. Therefore, our data collected from eBay captures large variations in both the existence and the leniency of a retailer's MBG policy, which lays the foundation for identifying the value of an MBG and uncovering how forward shipping fees affect this value. In addition, the heterogeneous reputation of eBay sellers also allows us to study the relationship between e-reputation and MBG value. Our approach enables us to explore whether the on-line reputation of a retailer is used by consumers to infer the smoothness of its return handling process and hence influence the value of its MBG policy. One key econometric challenge in our study involves accounting for a seller's endogenous choice of whether or not to offer an MBG policy. Intuitively, the sellers who are more likely to offer MBGs should also be the ones who expect to yield a better payoff from doing so. In Section 5 we discuss why the commonly used instrumental variable approach fails to correct for this endogeneity and propose the use of a Full Information Maximum Likelihood (FIML) estimator. We show that failure to account for this endogeneity issue results in an estimate of average MBG value that is substantially higher (36%) than our FIML estimates.

Based on 2946 eBay transactions of 86 products, our empirical analysis reveals several interesting results. First, ignoring the endogeneity of the MBG policy choice, we estimate that a seller with free forward shipping and an average reputation could expect a 7.01% increase in a consumer's WTP for the product if it switches from no-MBG to MBG. However, this increment is reduced to 5.16% once endogeneity is accounted for. Second, the value of an MBG policy for on-line consumer electronics is, in general, between 1% and 7% of the product value. Third, shipping charges tend to rapidly erode the value of an MBG policy. For example, we find that if 20% or more of the total price paid for a product is attributed to shipping, then the value of an MBG is statistically indifferent from zero. Fourth, we show that positive and negative seller reputations have separate and opposing effects on the value of an MBG policy and that the sellers with the largest number of reviews are not necessarily the firms that benefit the most from offering MBGs. Lastly, we show how our estimation results can be used to construct an optimization model that maximizes a retailer's expected per-unit profit by varying the forward shipping charge.

4.2 Decomposing the Value of an MBG policy

A consumers' WTP for a product is typically expressed as an aggregate monetary value. Given that our goal is to estimate the value of an MBG, it is useful to first discuss the individual components of WTP. When evaluating a product, consumers take both its physical and service attributes into account, which together define the product's total value. Following the hedonic pricing literature, the contribution of each attribute in forming this aggregate value is additively separable (Rosen, 1974). In other words, if we take away a service attribute, such as the MBG policy, consumer's WTP for the product decreases by a quantity that is equal to the value of the MBG. This additive separability idea has been applied in a variety of areas to understand consumers' valuation of warranty service (Chu and Chintagunta, 2009), remanufactured products (Subramanian and Subramanyam, 2012), brand names (Randall et al., 1998), and arts (Renneboog and Spaenjers, 2013). In the area of consumer returns, although it is reported that over 70% of consumers actively evaluate a retailer's MBG policy before purchase (Hsiao and Chen, 2012), guidance on how to estimate the value of MBGs is very limited.

Let V_1 denote the total value of a product with an MBG and V_0 denote the value of the exact same product without an MBG. Given the additive separability assumption, $V_1 - V_0$ corresponds to the value of an MBG, denoted by V. If we have data on both V_1 and V_0, estimating V should be achievable through an appropriate empirical framework. We show in Section 4.3 that such data can be collected from eBay and, in Section 4.4, that this empirical framework should account for a seller's endogenous decision of whether to offer an MBG. Note that V captures the impact of an MBG on product valuation, not on demand directly. Next, we hypothesize why V is contingent on both the forward shipping fee and the seller's reputation in an on-line retail context. While confirming the statistical significance of a hypothesis is important, our main focus is on the quantification of the effect sizes, which form the basis of our theoretical and managerial implications.

The Effect of Forward Shipping Charge

We begin by considering the case where no MBG is offered with a product, which is the focus of previous shipping fee studies. We observe that both analytical (e.g. Leng and Becerril-Arreola, 2010) and empirical (e.g. Lewis, 2006) works on shipping fees often construct a consumer's utility function as the difference between an exogenously determined product valuation and the total price paid. Therefore, the shipping fee alters a consumer's net utility from purchasing a product through the total price paid for that product, but does nothing to her V_0. That is, controlling for shipping service type (e.g. standard versus expedited) and observable seller characteristics, a higher shipping charge does not alter a consumer's valuation for the product, V_0. This is because the shipping charge itself does not carry any implication for the physical and service attributes of the product that a consumer uses to calculate her V_0. A testable empirical implication of the above discussion is that the shipping charge is not a significant predictor for the total monetary amount a consumer is willing to pay for a product (conditional on seller and transactional attributes). In Section 4.5, we show that our analysis supports this argument, providing consistency between our data and existing theories.

Next, we consider the case when an MBG is offered. Now, consumers have the opportunity to return their purchased items after experiencing them. When a return is deemed necessary, a consumer typically receives a refund for the total price paid minus the forward shipping charge. For example, if a digital camera is purchased for \$300 with a \$50 forward shipping charge, the refundable amount is \$300, not \$350. As discussed in the Introduction, this is a common practice for the online retailers of consumer electronics, including the sellers on eBay[5]. Given that most buyers of consumer electronics (a prototypical category of experience goods and our focus in this study) have a non-trivial probability of returning the purchased item, the calculation of product valuation V_1 needs to include the non-refundable shipping charge. The higher the shipping charge, the more loss a consumer will potentially incur, which in turn reduces V_1. Put differently, the forward shipping charge should be considered by a rational consumer as an "implicit restocking fee." To summarize, while an MBG policy grants consumers the opportunity to return a disliked item and should thus have a positive value ($V > 0$), the size of this value is likely to diminish with the amount of the forward shipping charge, as expressed by the following hypothesis.

Hypothesis 1. *The forward shipping charge has a negative impact on the value of an MBG for a product.*

Note that a higher shipping charge may also has a positive effect (from the retailer's perspective) of reducing the likelihood of a product being returned (since less of the total price paid is refunded to the consumer). In other words, a higher shipping charge simultaneously reduces the WTP for a product (bad for the seller) and reduces the probability of a return in the case of a misfit (good for the seller). We show how our estimation results may be used to optimize this trade-off in Section 4.5.

[5]The only instance in which an eBay seller is held responsible for refunding the forward shipping is when a consumer receives an item not matching the description on the product listing page. In this case, the buyer can file a claim with eBay and the seller is obligated to refund both the base product price and the forward shipping charge.

The Effect of Seller Reputation

Sellers on e-marketplaces such as Amazon, eBay, and Rakuten often have very similar store designs that are based on the templates provided by their retail platforms. In addition, product listing pages are usually standardized – the only customizable section is the product description area, where sellers can add text and pictures[6]. Nevertheless, consumers' WTP for identical products sold in such homogeneous environments exhibit significant variation (Standifird, 2001). The previous literature suggests that the seller reputation indices posted by the marketplaces provide consumers with an effective differentiator between various sellers' service qualities (e.g. Houser and Wooders (2006) and Resnick et al. (2006)). For example, eBay publishes the number of reviews (positive, neutral, and negative) a seller receives in a past period of time (three, six, and twelve months). These measures are used by consumers to infer a seller's speed of order processing, carefulness of packaging, friendliness of communication, and other service competencies. In addition, previous research (e.g. Dimoka et al. (2012), Obloj and Capron (2011), and Standifird (2001)) also demonstrates that positive and negative feedback for a seller have separate effects. In the context of eBay, this implies that more positive seller reviews increase consumers' WTP, while more negative or neutral reviews[7] decrease it. We show in Section 4.5 that our data supports this argument.

Using our notation, the existing theories offer explanations for why V_1 and V_0 would change with a seller's reputation indices. However, they do not provide insights or evidence into the relationship between reputation and the value of offering an MBG, V. The effect of the shipping charge on V focuses on the objective aspect of an MBG

[6]EBay offers the option to sellers of paying extra for a premium display, which is called eBay store. We control for a seller's participation in eBay store in our model.

[7]We are consistent with previous literature in grouping neutral and negative reviews together to represent the negative reputation of a seller (see Dimoka et al. (2012), Obloj and Capron (2011), and Subramanian and Subramanyam (2012)).

– the amount of money a seller promises to refund upon return. For consumers to enjoy the benefits of an MBG, the fulfillment of this promise is as important as the promise itself.

In practice, sellers have a fair degree of freedom in executing the return and refund process. First, a seller may attempt to deny the returns that are purely due to buyer remorse. Second, if original packaging cannot be fully recovered or item labels are removed, a seller may be less willing to process the returns. Third, a seller may be very slow in actually issuing a refund.[8] Such actions can create a considerable level of return hassle for the consumer. We expect that consumers will use the reputation indices to infer how a seller will exercise its return processing freedom. If a consumer believes that a seller will process returns more smoothly than a competing seller, her valuation of this seller's MBG should be higher. Thus, we propose opposing effects of positive and negative seller reputation as follows.

Hypothesis 2a. *The value of an online seller's MBG policy increases with its positive reputation.*

Hypothesis 2b. *The value of an online seller's MBG policy decreases with its negative reputation.*

In addition to our theoretical contribution to the e-reputation literature, the above hypotheses also offer practical implications to eBay sellers. Based on the distribution of reputation indices, we can first segment these sellers into groups and then estimate the group-specific values for V. Depending on the relative effect sizes of the positive and negative reputation, sellers with more reviews might not enjoy a higher MBG value. This analysis is presented in Section 4.5.

[8] The eBay community forum (accessible from `community.ebay.com`) shows buyer complaints regarding all these aspects of the return handling process.

4.3 Data and Variables

To estimate the value of an MBG policy and test our hypotheses, we collect transactional data from eBay auctions of consumer electronic products. We believe that eBay is an ideal data source for our study for three reasons. First, for identical products, we are able to observe transactions both with and without an MBG policy, which is a result of the fact that some sellers on eBay offer MBGs while others do not. Second, most eBay sellers impose a flat rate shipping fee (location independent). Therefore, we can observe the actual shipping cost paid by the buyer, which varies across sellers. Third, sellers on eBay can list a product either in the bidding format or in the fixed-price format. We only use the former, which resembles a second-price English auction, and final prices paid in these bidding transactions have been extensively used to measure consumers' valuation of a product (Bajari and Hortaçsu (2004), Houser and Wooders (2006), Obloj and Capron (2011), Resnick et al. (2006) and Standifird (2001)).

Our focus on consumer electronics has some practical advantages. eBay sellers' MBG policies for these products match very closely with the three characteristics of a typical MBG policy in on-line retailing – while product price is refunded, forward and return shipping charges are not. In addition, many electronic products have a reasonable transaction frequency on eBay, which increases the chance of observing both MBG and no-MBG transactions. Finally, as compared with other product categories such as apparel, there are few private labels in consumer electronics, which allows for comparisons between retailers selling the exact same products.

To construct our sample, we followed a two-step data collection procedure – product selection and transaction collection. For product selection, we started with eBay's existing categorization scheme for consumer electronics[9]. The categories with low

[9] Accessible at `www.ebay.com/electronics`.

transaction frequency, such as digital clocks and electronic kettles, were filtered out to avoid data sparsity. We also excluded products that have complicated configurations such as laptops and desktops to avoid matching issues[10]. This high-level screening process resulted in 16 categories of consumer electronic products (see Table 4.2). For each category, we then used eBay's product search engine to list the top 100 products by popularity and randomly selected 10 products from this list. This product selection approach allows us to mitigate the risk of constructing a biased sample and ensures a reasonable degree of transaction frequency. Our product selection process generated $16 \times 10 = 160$ products. Next, we collected bidding transactions for these products. Based on their conditions, items on eBay are categorized by new, new other (mostly open box items), refurbished, used, and for-parts. We collected completed bidding transactions only of new items during the first quarter of 2013 regarding these 160 products.

The following steps were used to clean the data to make it applicable for our analysis. First, we excluded the transactions that have a location-dependent shipping charge, since the exact shipping fee paid by the buyer is not visible to us for these transactions. Second, we eliminated the transactions that have a "buy it now" option, since the format of these transactions is a mixture of fixed-price and auction. Third, the small fraction of transactions that have an explicit restocking fee were excluded. Fourth, free return shipping transactions (also a very small fraction) were also excluded. Fifth, to best estimate V, we required products to have transactions both with and without an MBG. Thus, we excluded the products that only have one type of transactions. Lastly, we removed a small number of potential outliers, which is discussed in the next section. The screened sample contains 2,946 transactions of 86 products.

[10] For example, virtually all laptop models can be further divided based on hard disk, processor, and RAM, which makes matching the exact same model difficult.

Dependent Variables

For each transaction i, we define the consumer's valuation of the product, v_i, as the *total price* paid in an eBay auction. If there is a shipping charge, v_i is the sum of the base product price and the shipping charge. As discussed earlier, this operationalization of product valuation or WTP is widely used in the literature. We denote the MBG transactions by $I_i = 1$. The typical eBay terminology for an MBG policy is "returns accepted, X days money back, and buyer pays return shipping". The default "X" value when a seller lists a product is 14, although it could be changed to 7, 30 and 60. The no-MBG transactions are denoted by $I_i = 0$. If there is no MBG, the eBay terminology is simply "no returns accepted". In our sample, the proportions of no returns, and 7, 14, and 30 days money back are 76.04%, 0.1%, 23.52%, and 0.34%, respectively. No 60-day MBG policies were observed. Since the vast majority of the MBG transactions has a 14-day return period (the default), we do not study the impact of return period. However, we perform a robustness check on our estimation results by removing the 7 and 30 days MBG transactions (Section 4.5).

Table 4.2 Difference in Consumer Valuation Between MBG and No-MBG Transactions

Product category	Avg. differential ($)	Product category	Avg. differential ($)
Audio System	4.08	Headphone	6.36
Calculator	8.18	Keyboard	4.65
Camcorder	12.07	Modem	2.89
Camera	10.84	MP3 Player	6.99
Camera Lens	10.88	Printer	5.04
Vacuum Cleaner	16.20	Scanner	5.58
DVD Player	2.48	Tablet	21.31
GPS	10.41	Webcam	2.83

In Table 4.2, we report the impact of an MBG policy on product valuation in its simplest form – the difference in v_i between MBG and no-MBG transactions. Specifically, we calculate $\overline{v}|_{I=1} - \overline{v}|_{I=0}$ for each specific product and average this differential by category. This simple calculation suggests that MBG policies are valued positively by consumers in our data. It also suggests that the dollar value of an MBG is likely to vary widely across products, probably due to their heterogeneous price levels. This observation implies that in order to pool products together for analysis, we should use a normalized version of v_i as our dependent variable. For transaction i of a certain product, we define the dependent variable as $y_i = \frac{v_i}{\overline{v}|_{I=0}} \times 100$, where $\overline{v}|_{I=0}$ is the average v_i of no-MBG transactions for this product. That is, we scale the transaction price by the average price of no-MBG transactions of the same product. This normalization approach allows for a clean interpretation of the value of an MBG policy – the percentage increase in consumer willingness-to-pay when a seller switches from no-MBG to MBG, or $\frac{V}{V_0}$. For both MBG and no-MBG transactions, we remove the highest and lowest 1% of y_i to mitigate the noise from outliers.

As we will show in Section 4.4, the appropriate econometric framework involves two dependent variables, y_i and I_i. Since y_i is our primary interest, we label it as the outcome variable and hence its regression as the outcome equation. The additional regression for I_i is used to control for a seller's endogenous choice of whether to offer an MBG. Therefore, we label I_i as the choice variable and its regression as the choice equation.

Independent Variables

Forward shipping fee. The dollar value of the shipping fee is likely to vary across products; therefore, we normalize the shipping charge using the same approach as in the previous section to create *shipping_charge*$_i$. Around 28% of the transactions in our data provide free shipping. For those with a positive shipping charge, the mean is 7.85% of the average total price for no-MBG transactions. Table 4.3 provides the detailed summary statistics.

Seller reputation. We use literature-based reputation measures – *slr_posi_log*$_i$ for positive seller reputation and *slr_nega_log*$_i$ for negative seller reputation (e.g.

Table 4.3 Summary Statistics for the Shipping Fee Variable

	Free shipping	Buyer pays shipping	
	Observations	Observations	Mean
No MBG	600	1640	7.82
MBG	222	484	7.98

Subramanian and Subramanyam (2012)). Specifically, $slr\_posi\_log_i$ ($slr\_nega\_log_i$) is the logged number of positive (neutral and negative) feedback received by the seller in the past 12 months. We use a seller's six-month review statistics as an alternative measure for $slr\_posi\_log_i$ and $slr\_nega\_log_i$ in a robustness check, as described in Section 4.5

Transaction-related controls. The previous literature suggests that the total price paid in an online auction is related to its activeness, length, and ending time (Houser and Wooders (2006), Resnick et al. (2006) and Standifird (2001)). For transaction i, $bids\_log_i$ measures the logged number of bids, $duration_i$ is the number of days the auction lasts, $weekend_i$ is whether the auction ends during the weekend (Saturday and Sunday) and $night_i$ is whether it ends at night (6 p.m. to 6 a.m.). $duration_i$ takes values of 1, 3, 5, 7, and 10 days, which are eBay's available duration choices. Since our data collection period spans three months, we create two month dummies, $february_i$ and $march_i$, to account for the rapid value depreciation of consumer electronics. Furthermore, we use $handling\_time_i$ to account for the transaction's indicated order processing time, which could be 1, 2 or 3 days. We control for the type of shipping service associated with each transaction. Our data contains economy (17.8%), standard (45.2%), and expedited (36.0%) shipping services, which are coded by two dummy variables $shipping\_standard_i$ and $shipping\_expedited_i$.

Seller-related controls. An "eBay store" is a marketing tool sold by eBay to its sellers for reaching a larger audience. Sellers operating a store may be more attractive and thus command a higher price premium from consumers. We use a dummy variable, slr_store_i, to account for this effect. Similarly, the length of existence of a seller could also confound the relationship between reputation and valuation (Subramanian and Subramanyam, 2012). We use $slr\_tenure\_log_i$ to measure the logged number of days from a seller's registration date to the date transaction i is made.

Product-related controls. We use $pdt\_avg\_price_i$ to control for the price level of a product, which is simply the average price paid for this product (in hundreds of dollars) during our data collection period. Consistent with Subramanian and Subramanyam (2012), we also account for the demand of a product. $pdt\_demand\_log_i$ measures the logged number of new units sold of a product during our data collection period. In addition, we use $pdt\_review\_log_i$, the logged number of reviews received by a product, to capture the extent of word-of-mouth information available (Dimoka et al., 2012). Lastly, given that we have 16 product categories, we create 15 dummy variables to account for any category-level fixed effects.

Table 4.4 provides the descriptive statistics for all variables discussed above, except the categorical dummies. Note that when performing a logarithmic transformation for a variable, we use $new = \log(old + 1)$ if the original variable has a zero minimum.

4.4 The Econometric Model

When a seller decides to list an item on eBay, she is presented with a choice of whether or not to offer an MBG policy. It is reasonable to expect that sellers will intentionally choose the option that is expected to make them better off. Therefore, if we conceptualize the binary variable of MBG choice I_i as the "treatment assignment" in econometric terms, we cannot consider the treatment assignment process random. Instead, endogeneity in I_i will affect our empirical strategy if some unobserved factors influencing I_i also affect the outcome variable y_i. The more obvious consequence for the correlation between I_i and the error term in y_i is that simple estimators

Table 4.4 Descriptive Statistics

Variable	Mean	SD	Min	Max
outcome DV y_i	101.40	5.98	83.89	133.36
choice DV I_i	0.24	0.43	0	1
$shipping\_charge_i$	5.66	6.36	0	49.35
$slr\_posi\_log_i$	3.78	1.82	0	11.91
$slr\_nega\_log_i$	0.57	0.95	0	7.24
slr_store_i	0.10	0.30	0	1
$slr\_tenure\_log_i$	7.34	1.34	0	8.64
$pdt\_avg\_price_i$	2.02	1.45	0.11	8.03
$pdt\_demand\_log_i$	5.87	1.16	1.39	7.24
$pdt\_review\_log_i$	4.38	1.92	0.69	8.26
$bids\_log_i$	1.78	1.29	0	4.37
$duration_i$	3.99	2.56	1	10
$shipping\_standard_i$	0.45	0.50	0	1
$shipping\_expedited_i$	0.37	0.48	0	1
$handling\_time_i$	1.96	0.76	1	3
$weekend_i$	0.31	0.46	0	1
$night_i$	0.35	0.48	0	1
$february_i$	0.33	0.47	0	1
$march_i$	0.21	0.41	0	1

Categorical dummies are excluded to conserve space.

such as OLS will produce biased estimates. What may be less obvious is that the instrumental variable (IV) approach, an often-used method for resolving endogeneity, is also inappropriate in our context. In the following, we discuss these issues in detail and construct a Full Information Maximum Likelihood (FIML) estimator to eliminate the endogeneity concern.

For transaction i, let y^*_{1i} denote the potential product valuation when an MBG is offered and y^*_{0i} when it is not offered. We use the term "potential" because the valuations are latent variables. In the actual data, we do not observe y^*_{1i} for the $I_i = 0$ transactions or y^*_{0i} for the $I_i = 1$ transactions. However, it is important for us to model the potential outcomes (y^*_{1i} and y^*_{0i}) as opposed to directly regressing the actuals (y_i) since our desired economic interpretation is embedded in the potentials. That is, we want to estimate the change in consumer product valuation if a seller switches from no-MBG to MBG (or from MBG to no-MBG). This change represents the value of an MBG policy.

Specifically, y^*_{1i} and y^*_{0i} are modeled as two linear equations in their respective explanatory variables:

$$
\begin{aligned}
y^*_{0i} &= \alpha_0 + \alpha' X_i + e_{0i} &\quad (4.1a)\\
y^*_{1i} &= \beta_0 + \beta' X_i + e_{1i} &\quad (4.1b)
\end{aligned}
$$

where α_0 and β_0 are intercepts, α' and β' are vectors of coefficients, and the error terms follow normal distributions such that $e_{0i} \sim N(0, \sigma_0)$ and $e_{1i} \sim N(0, \sigma_1)$. We use different notations for the coefficient vectors in (4.1a) and (4.1b) since our key interests reside in the difference between these two equations. Since we do not have any variable that is observed only in part of the sample (i.e. when $I_i = 0$ or when $I_i = 1$), the same set of explanatory variables are applied to both equations. The above setup can be combined into a single equation, utilizing the MBG status dummy, I_i.

$$y^*_i = \alpha_0 + \delta_0 I_i + \alpha' X_i + \delta' X_i I_i + e_{0i}(1 - I_i) + e_{1i} I_i \quad (4.2)$$

where $\delta_0 = \beta_0 - \alpha_0$ and $\delta' = \beta' - \alpha'$.

We first consider an OLS estimator for (4.2), which is obtained by directly regressing y_i on X_i and I_i:

$$y_i = \alpha_0^{ols} + \delta_0^{ols} I_i + \alpha^{'ols} X_i + \delta^{'ols} X_i I_i + \varepsilon_i \quad (4.3)$$

The problem with this approach is that I_i is likely to be correlated with e_{1i} (a component in ε_i) and hence correlated with ε_i as well. Consequently, the OLS estimates will be biased.

By construction of the OLS estimator, the potential correlation between I_i and ε_i should come from some common unobserved factors that affect both I_i and ε_i. We use a probit regression to separate the effects of observed and unobserved factors on I_i:

$$I_i = 1(\gamma_0 + \gamma' Z_i + u_i > 0) \tag{4.4}$$

where γ_0 is the intercept, γ' is the vector of coefficients, and the error term is normally distributed such that $u_i \sim N(0, 1)$. The standard deviation of u_i is normalized to 1, a common identification requirement in choice models. Z_i could contain different variables than X_i. We believe that the common unobserved factors in u_i and ε_i might relate to the textual contents of a seller's previous reviews. First, if a seller has received reviews explicitly mentioning their meritorious returns handling process in the past, consumers would be willing to pay more for their MBG policy. Factoring in the higher gains from implementing an MBG policy, such sellers are more likely to offer an MBG in the first place. Essentially, this implies a positive correlation between u_i and e_{1i}. Second, even without specific reviews on returns, consumers might still use reviews related to other types of customer service to infer a seller's friendliness in handling returns. Similarly, sellers who have received meritorious comments might be more likely to offer MBGs and consumers might be willing to pay more for the MBGs from these sellers. Again, this leads to a positive correlation between u_i and e_{1i}. Note that sellers' heterogeneous costs of processing returns should not be considered as a driver for this correlation, since cost is visible to sellers but not to consumers. That is, while the cost enters into u_i, it is not included in e_{0i} or e_{1i}. Formally, we denote the correlation between u_i and e_{1i} by ρ_1 and that between u_i and e_{0i} by ρ_0. The above discussion gives us strong *a priori* belief that ρ_1 is positive, while the directionality of ρ_0 is not *a priori* clear. Since u_i and e_{1i} are parts of I_i and ε_i, respectively, the correlation between the former will lead to correlation between the latter and thus bias the OLS results.

Given that an IV approach, such as the two-stage-least-squares (2SLS) procedure, is a commonly used method to resolve the endogeneity issue, we consider its appropriateness in our context. Suppose that q_i is an instrument for I_i. In the first stage, I_i is regressed on Z_i and q_i and X_iI_i is regressed on Z_i and X_iq_i. The linear predictions, $\widehat{I}_i$ and $\widehat{X_iI_i}$, are used subsequently in the second stage to replace I_i and X_iI_i. The following regression is then estimated through OLS.

$$y_i = \alpha_0^{2sls} + \delta_0^{2sls}\widehat{I}_i + \alpha'^{2sls}X_i + \delta'^{2sls}\widehat{X_iI_i} + \varepsilon_i \tag{4.5}$$

As Heckman (1997) and Vella and Verbeek (1999) point out, this IV approach produces consistent estimates only when the unobserved factors in u_i affects e_{0i} and e_{1i} in the same way. Technically, this requires $\rho_1 = \rho_0$ (Vella and Verbeek, 1999, p.474) as in Guajardo et al. (2012). This assumption is especially unappealing in our context, since we have many reasons to believe $\rho_1 > 0$ but the sign of ρ_0 is not clear. This subtle but important assumption of the IV approach has strong implications for the choice of an appropriate estimator.[11]

Next, we derive the FIML estimator that will be used for our analysis. As an intermediate step for constructing the likelihood function for the whole sample, we separately derive the likelihood for an MBG transaction and a no-MBG transaction. If $I_i = 0$, we have $y_i = y_{0i}^* = \alpha_0 + \alpha'X_i + e_{0i}$ and $\gamma_0 + \gamma'Z_i + u_i < 0$. Since e_{0i} and u_i follow a bivariate normal distribution, properties of the truncated normal distribution are used to derive the likelihood for a no-MBG transaction, which is given by:

$$\begin{aligned} L_i|_{I_i=0} &= \Pr(e_{0i} = y_i - \alpha_0 - \alpha'X_i, u_i < -\gamma_0 - \gamma'Z_i) \\ &= \frac{1}{\sigma_0}\phi(y_i - \alpha_0 - \alpha'X_i)\Pr(u_i < -\gamma_0 - \gamma'Z_i|e_{0i}) \\ &= \frac{1}{\sigma_0}\phi(y_i - \alpha_0 - \alpha'X_i)\Phi[\frac{-\gamma_0 - \gamma'Z_i - \frac{\rho_0}{\sigma_0}(y_i - \alpha_0 - \alpha'X_i)}{\sqrt{1-\rho_0^2}}] \end{aligned} \tag{4.6}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of a standard normal distribution. If $I_i = 1$, we have $y_i = y_{1i}^* = \beta_0 + \beta'X_i + e_{1i}$ and $\gamma_0 + \gamma'Z_i + u_i > 0$. The likelihood for

[11] To observe the severity of the bias in OLS and 2SLS estimators, we conduct a simulation study to evaluate the finite sample properties of these approaches in Appendix G. We find that, with our econometric setup: OLS is always biased; when $\rho_1 = \rho_0$ both 2SLS and FIML are unbiased but FIML is more efficient; and when $\rho_1 \neq \rho_0$ only FIML is unbiased.

an MBG transaction is given by:

$$\begin{aligned} L_i|_{I_i=1} &= \Pr(e_{1i} = y_i - \beta_0 - \beta' X_i, u_i > -\gamma_0 - \gamma' Z_i) \\ &= \frac{1}{\sigma_1}\phi(y_i - \beta_0 - \beta' X_i)\Pr(u_i > -\gamma_0 - \gamma' Z_i | e_{1i}) \\ &= \frac{1}{\sigma_1}\phi(y_i - \beta_0 - \beta' X_i)\Phi[\frac{\gamma_0 + \gamma' Z_i + \frac{\rho_1}{\sigma_1}(y_i - \beta_0 - \beta' X_i)}{\sqrt{1-\rho_1^2}}] \end{aligned} \tag{4.7}$$

Combining (4.6) and (4.7), the likelihood for any given transaction is $L_i = (1 - I_i) \times L_i|_{I_i=0} + I_i \times L_i|_{I_i=1}$. Therefore, the log likelihood for the whole sample is as follows:

$$LL = \sum_{i=1}^{n}[(1 - I_i)\ln L_i|_{I_i=0} + I_i \ln L_i|_{I_i=1}] \tag{4.8}$$

Maximizing (4.8) yields the FIML estimator of the α and β vectors, which is asymptotically consistent (Maddala, 1983). We use Stata's "ml" language to program this estimator (see Appendix G for details).

We can use our model to test for endogeneity in two ways. The more direct test examines the statistical significance of ρ_1. Alternatively, we could also implement the Hausman's model specification test to compare the systematic difference between the FIML and OLS estimates (Hausman, 1978). We perform both tests in the next section.

With our setup, the difference between $E(y^*_{1i})$ and $E(y^*_{0i})$ entails the interpretation of "the value of an MBG policy". That is,

$$NV = E(y^*_{1i}) - E(y^*_{0i}) = \beta_0 + \beta' X_i - \alpha_0 - \alpha' X_i = \delta_0 + \delta' X_i \tag{4.9}$$

We use NV since, strictly speaking, the quantity being estimated is a normalized value: $NV = \frac{V}{V_0}$. Our three hypotheses are tested through the statistical significance of the δ vector.

Table 4.5 Estimation Results

	OLS Estimates				FIML Estimates			
	Without controls		With controls		Outcome equation y_i		Choice equation I_i	
Variables	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
	δ coefficients		**δ coefficients**		**δ coefficients**			
indicator for MBG I_i	7.509***	(0.596)	7.008***	(0.337)	5.163*	(2.051)		
$shipping\_charge_i$	-0.137*	(0.063)	-0.108**	(0.041)	-0.124**	(0.041)		
$slr\_posi\_log_i$	1.579***	(0.372)	1.112***	(0.216)	1.313***	(0.214)		
$slr\_nega\_log_i$	-2.483***	(0.606)	-2.354***	(0.436)	-1.914*	(0.819)		
	α coefficients		**α coefficients**		**α coefficients**		**γ coefficients**	
constant	98.22***	(1.404)	97.23***	(1.289)	97.12***	(2.058)	0.635^{+}	(0.368)
$shipping\_charge_i$	0.0041	(0.018)	0.0042	(0.015)	0.0056	(0.019)	-0.0185^{+}	(0.009)
$slr\_posi\_log_i$	0.944***	(0.107)	0.843***	(0.084)	0.830***	(0.085)	0.0474	(0.033)
$slr\_nega\_log_i$	-2.876***	(0.292)	-3.242***	(0.215)	-3.338***	(0.560)	0.524***	(0.066)
slr_store_i			4.969***	(0.570)	5.115***	(1.008)	0.633^{+}	(0.382)
$slr\_tenure\_log_i$			0.395***	(0.067)	0.376***	(0.069)	-0.0453	(0.029)
$pdt\_avg\_price_i$			0.251**	(0.084)	0.245*	(0.109)	-0.110***	(0.026)
$pdt\_demand\_log_i$			-0.202^{+}	(0.120)	-0.183	(0.116)	-0.0526	(0.041)
$pdt\_review\_log_i$			-0.162**	(0.050)	-0.158***	(0.044)	0.00204	(0.013)
$bids\_log_i$			0.232***	(0.064)	0.252***	(0.064)	0.0414	(0.027)
$duration_i$			0.239***	(0.034)	0.229***	(0.036)	0.0107	(0.015)
$shipping\_standard_i$			0.116	(0.261)	0.128	(0.260)	0.0257	(0.136)
$shipping\_expedited_i$			0.673**	(0.258)	0.666*	(0.273)	0.0872	(0.138)
$handling\_time_i$			-0.403**	(0.145)	-0.465*	(0.212)	-0.258***	(0.052)
					error correlations[1]: $\widehat{\rho_0} = -0.076$, $\widehat{\rho_1} = 0.375^{*}$			
Log likelihood[2]	-10042.43		-9616.58		-9602.31			

$weekend_i$, $night_i$, $february_i$, and $march_i$ are included in OLS (with controls) and FIML models. *categorical dummies* is included in all models.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Cluster-robust (product level) standard errors in parentheses.

4.5 Analysis and Results

We fit both the OLS estimator as in (4.3) and the FIML estimator as in (4.8) to the eBay data[12]. The OLS estimator is fitted both with and without control variables. The endogeneity issue is addressed progressively through these three models. For the outcome equation (y_i), we estimate both α and δ coefficients for $shipping\_charge_i$, $slr\_posi\_log_i$, and $slr\_nega\_log_i$. For all other covariates (the control variables), we estimate only the α coefficients and constrain the δ coefficients to zero[13]. We also center $slr\_posi\_log_i$ and $slr\_nega\_log_i$ to give δ_0 (the constant term in NV) an interpretation of the value of an MBG when forward shipping is free and reputation indices are at the mean. Furthermore, we use cluster robust standard errors. Results are reported in Table 4.5.

As discussed earlier, the main econometric issue involved in estimating the value of an MBG is a seller's endogenous choice to offer the MBG. This endogeneity implies that the sellers who expect a higher consumer WTP from offering MBGs are also likely to be the ones who actually offer MBGs. Therefore, the consequence of not accounting for endogeneity is an exaggerated value of the MBG policy. While the OLS without controls model takes the contingency of the MBG value (V depends on forward shipping and seller reputation) into account, it ignores the fact that the decision to offer an MBG is endogenous. As a result, this model produces a misleadingly high estimate of V ($\delta_0 = 7.509$). The OLS with controls model partially addresses the endogeneity issue through the inclusion of control variables. Specifically, some variables, such as slr_store_i, exert a dual impact on both the seller's choice to offer an MBG and the consumer's product valuation. Therefore, controlling for these variables should attenuate the endogeneity problem, which is evident from

[12] 2SLS is not included in our empirical analysis because it is less appropriate than FIML in our context. Furthermore, a solid instrument for MBG choice is not present in our data.

[13] This is analogous to the conventional practice that only the hypothesized interaction effects are estimated in a linear regression model.

the decrease in δ_0 from 7.509 to 7.008. The OLS model, however, cannot account for the unobserved factors that cause the endogneity problem. Through modeling the correlation between the error terms of the outcome and choice equations (i.e., where the unobserved factors enter), the FIML model does account for the influence of these unobserved factors and estimates a lower value of an MBG ($\delta_0 = 5.163$).

Next, we conduct two formal endogeneity tests. The first test directly examines the direction and significance of $\widehat{\rho_1}$. Recall our *a priori* expectation is that some unobserved factors (e.g. content of reviews) influencing product valuation also affect the sellers' decision to offer the MBG, which gives rise to a positive ρ_1. Our estimate of $\widehat{\rho_1} = 0.375$ (significant at 0.05 level) verifies this expectation. Our second test is the well-known Hausman's model specification test (Hausman, 1978). While the null hypothesis is that both OLS (with controls) and FIML estimators are consistent and hence OLS is more efficient[14], the alternative hypothesis is that only the FIML estimator is consistent. Rejecting the null in our context would imply that the endogeneity of the MBG choice makes the OLS estimates systematically different from the FIML estimates. We implement the Hausman test on the δ coefficients ($\chi^2(4) = 8.07$, $p < 0.1$) and find that the endogeneity problem has significantly influenced the OLS estimation of our key parameters.

Having verified our choice of the estimation procedures, we now test the statistical significance of our hypotheses. The Wald tests in Table 4.5 show preliminary supportive evidence for all three hypotheses. Specifically, the coefficients on $shipping\_charge_i$, $slr\_posi\_log_i$, and $slr\_nega\_log_i$ are significant at 0.05 or lower levels. We also conduct the more formal likelihood ratio test, which we illustrate here with $shipping\_charge_i$. The log likelihood for the unconstrained model (i.e. FIML model in Table 4.5) is -9602.31. Re-estimating the model with the δ coefficient on $shipping\_charge_i$ constrained to zero reduces the log likelihood to -9609.75. Two times

[14]The efficiency of OLS comes from estimating fewer parameters than FIML.

the difference, 14.89, follows a Chi-square distribution with 1 degree of freedom, which leads to a p-value smaller than 0.001. Therefore, the effect of shipping charge on the value of an MBG policy is negative and significant, supporting Hypothesis 1. All likelihood ratio test results are shown in Table 4.6.

Table 4.6 Likelihood Ratio Tests

	$shipping\_charge_i$	$slr\_posi\_log_i$	$slr\_nega\_log_i$
Effect size δ	-0.124	1.313	-1.914
Constrained LL	-9609.75	-9628.55	-9610.95
χ^2 statistic	14.89	52.47	17.28
Test result	$p < 0.001$	$p < 0.001$	$p < 0.001$
Conclusion	H1 supported	H2a supported	H2b supported

In the following, we perform additional analysis to derive economic implications of our hypothesized relationships. We also conduct a number of robustness checks of our results. Lastly, we show how our estimation results may be used to calibrate a shipping charge optimization model.

Shipping Charge

On one hand, our analysis confirms the previous literature that the shipping charge alone does not affect the consumers' product valuations (Leng and Becerril-Arreola, 2010; Lewis, 2006). That is, in the absence of returns, a lower shipping charge does not alter the total amount of money a consumer is willing to pay for a product (i.e. the α coefficient on $shipping\_charge_i$ is insignificant). However, we extend the literature on on-line retailing by showing that a forward shipping charge *is considered an implicit restocking fee* by consumers and hence negatively affects the value of an MBG policy.

There are several managerial implications related to this result. First, we empirically validate the intuitive notion that the value of an MBG policy is highest when forward shipping is free. We quantify this value for the consumer electronics segment, finding that, for a seller with an average reputation (i.e. $slr\_posi\_log_i$ and $slr\_nega\_log_i$ are at their respective means), the value of an MBG to consumers is $\widehat{\delta_0} = 5.163\%$. In other words, if the seller switches from "returns not accepted" to "money back guaranteed" given free forward shipping, she should expect a 5.163% lift in consumers' willingness to pay. Second, when both an MBG and forward shipping charges exist, the latter erodes the value of the former considerably. The effect size of $shipping\_charge_i$, -0.124, offers some insights into this point. As an example, if the forward shipping charge increases from free to 10% of the total price, the value of an MBG drops from 5.163% to 3.923%. Consider a \$90 non-returnable digital camera with \$10 shipping charge, for a total price of \$100. Our results suggest that consumers are willing to pay \$103.92 for the same camera with an MBG policy. However, if the non-refundable \$10 was eliminated, the WTP increases \$1.24 to \$105.16. To further explore the negative effect of $shipping\_charge_i$, we draw a confidence band plot (Miller et al., 2013) in Figure 4.1, which depicts how the value of an MBG and its 95% confidence interval (in shaded area) decrease as $shipping\_charge_i$ increases from 0 to 20[15]. When shipping charge is close to 20% of total price, the value of an MBG becomes statistically insignificant from zero.

Finally, we observe in the choice equation in Table 4.5 that the γ coefficient on the shipping charge is negative and marginally significant, which means that a higher shipping charge correlates with a lower probability to offer an MBG. This might be indicative that the sellers on eBay are aware of the "incompatibility" between offering an MBG and charging a high shipping fee.

[15] We choose 20 as the maximum because the impact of $shipping\_charge_i$ on NV is identified by the transactions that offer MBG and charge a positive shipping fee. Among these transactions, the 90th percentile of $shipping\_charge_i$ is approximately 20. Thus, for $shipping\_charge_i > 20$, the effect of $shipping\_charge_i$ is driven less by data but more by extrapolation.

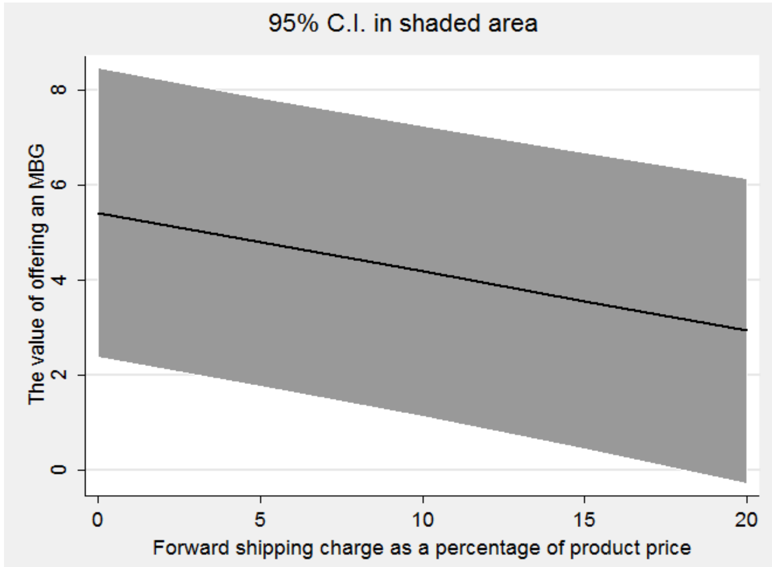

Figure 4.1 Effect of Shipping Charge on the Value of MBG

Seller Reputation

The α coefficients on $slr\_posi\_log_i$ and $slr\_nega\_log_i$ are 0.83 and -3.338, respectively (Table 4.5). These results are in line with the two conclusions drawn by the previous studies on e-reputation (e.g. Standifird, 2001): (1) positive and negative reputation have opposing effects on consumers' product valuation, and (2) the negative impact tends to be larger in size. We contribute to this literature by showing that a seller's e-reputation also affects the buyers' valuation of an MBG policy. Specifically, the δ coefficients on $slr\_posi\_log_i$ and $slr\_nega\_log_i$ show significant and opposite effects. Furthermore, the negative impact has a larger effect size, $|-1.914| > |1.313|$. But this difference is not statistically significant.

Reputation differs from shipping charge in that it is not generally considered to be a seller's decision (at least in the short term). Rather, it is a characteristic of the seller. This fact leads us to use a different approach to derive managerial implications from the reputation-related results. Specifically, we first cluster the sellers in our data according to $slr\_posi\_log_i$ and $slr\_nega\_log_i$ using the k-means algorithm[16]. It appears that three is the maximum number of clusters that would assure a reasonable degree of heterogeneity across clusters (see Appendix H for details). Then, for each cluster we derive the value of an MBG under two conditions; free shipping and a 20% shipping cost. The results are presented in Table 4.7.

The average value of an MBG is smaller than 7% of product value across all six scenarios. The main difference between clusters is the activeness of sellers. Interestingly, this does not *a priori* predict which seller cluster should enjoy the highest value of an MBG, since more active sellers have a higher number of both positive and negative reviews, which exert opposing effects. Our analysis indicates that while the occasional sellers (least active) have the lowest value of MBGs, it is not the volume sellers (most active) who enjoy the highest benefits of offering MBGs. Rather, it is the moderately active sellers (Cluster 2) who receive the best payoff from offering an MBG policy. The reason for this result is that as we move from Cluster 2 to Cluster 3, the increases in $slr\_nega\_log_i$ and $slr\_posi\_log_i$ are similar but $slr\_nega\_log_i$ has a stronger impact. Furthermore, if an occasional seller charges a relatively high shipping fee, her expected payoff from offering an MBG policy could be close to zero.

Table 4.7 The Value of MBGs by Seller Cluster

	Cluster 1 Occasional Sellers (2.10,0.10)[1]	Cluster 2 Moderate Sellers (4.28,0.42)	Cluster 3 Volume Sellers (6.92,2.45)
$shipping\_charge_i = 0$	4.13	6.38	5.96
$shipping\_charge_i = 20$	1.64	3.89	3.47

[1](x,y) are the means of slr_posi_log$_i$ and slr_nega_log$_i$ (no centering) for each cluster.

Finally, we turn to the choice regression. While the γ coefficient on $slr\_posi\_log_i$

[16]Alternatively, we could do a median-split for $slr\_posi\_log_i$ and $slr\_nega\_log_i$ to create four seller profiles. However, this approach deviates from reality since sellers with few positive reviews but a lot of negative reviews are extremely rare.

is insignificant, that on $slr\_nega\_log_i$ is positive and significant. This suggests that sellers are acting in a way detrimental to themselves. That is, although sellers with more negative reviews realize less value from an MBG, they are more likely to offer MBGs. One explanation is that sellers struggling with negative feedback might be using the MBG as a longer-term strategy to shift their reviews more positively. Such strategic behavior is not captured by our analysis but is an interesting area for future research.

Robustness Checks

We provide a number of robustness checks, which are summarized in Table 4.8. For each robustness check model, we run our FIML estimator. To conserve space, we present only the δ parameter estimates in Table 4.8, but the full set of results is available from the authors upon request. First, we examine whether our results are sensitive to the exclusion of some "extreme products" for which the MBG might have a very high or very low importance to consumers. To this end, we rank the 86 products in our sample according to $\overline{y_i}|_{I=1} - \overline{y_i}|_{I=0}$, the normalized price differential between MBG and no-MBG transactions. Model 1 contains the estimation results from excluding the top 3 and bottom 3 products in this rank. Second, as discussed in Section 4.3, a small fraction of our MBG transactions have a return window of 7 or 30 days, while the majority have a 14-day time window. We reran our analysis in Model 2 by excluding this small fraction of transactions. Third, we have used the number of reviews a seller received in the past twelve months to construct measures for positive and negative reputation. In Model 3, we use the six-month statistics[17] to create alternative measures for $slr\_posi\_log_i$ and $slr\_nega\_log_i$. Fourth, we have allowed both ρ_0 and ρ_1 to be freely estimated. While the rationale behind a positive

[17] Although eBay also publishes three-month review statistics of each seller, we choose not to use them because the majority (84.1%) of the sellers in our data received zero negative reviews in the past three months.

ρ_1 is strong, our *ex ante* expectation for ρ_0 is not very clear. Therefore, the correct error structure might have $\rho_0 = 0$. In Model 4, we explore whether our estimation results are sensitive to this constraint. Lastly, while our empirical model is identified with exactly the same set of variables in X_i and Z_i, it is sometimes advised to include additional variables in Z_i to enhance identification (Heckman, 1997; Maddala, 1983; Vella and Verbeek, 1999). To this end, we create $MBG\_popularity\_log_i$ to capture the popularity of offering an MBG when transaction i was listed by its seller. Specifically, $MBG\_popularity\_log_i$ is the logged number of with-MBG listings of the same product at the time point when the seller posted i on eBay. This variable is expected to affect a seller's choice of offering an MBG, but not directly affect a buyer's product valuation since it is not visible when an auction completes. In Model 5, we insert $MBG\_popularity\_log_i$ into Z_i in addition to all the variables that appear in X_i.

As shown in Table 4.8, the results from our robustness checks are generally consistent with those in Table 4.5. One exception is Model 3, where weaker reputation effects are observed. The exposition of seller profile web-pages might explain this observation. Specifically, the homepage of an eBay seller displays only the twelve-month review statistics, whereas the six-month statistics are displayed on the feedback page – one more click from the home page. Consequently, the twelve-month statistics are more likely to draw attention from consumers and hence better capture the effects of e-reputation.

Example Application: Managing the Forward Shipping Fee

Our results suggest that the forward shipping fee has implications for both the costs and benefits of offering an MBG. Specifically, our analysis shows that the benefit of a lower shipping charge relates to a higher value for an MBG. The costs, on the other hand, include those associated with a higher probability of returns from consumers. In the following, we use our estimated parameters as inputs to construct a simple

Table 4.8 Robustness Checks

	Model 1	Model 2	Model 3	Model 4	Model 5
Variables	δ coefficients (cluster robust standard errors)				
indicator for MBG I_i	4.581**	4.906**	7.140***	6.058*	4.836**
	(1.695)	(1.869)	(1.939)	(2.694)	(1.483)
$shipping\_charge_i$	-0.132**	-0.125**	-0.137***	-0.126**	-0.124**
	(0.042)	(0.041)	(0.041)	(0.042)	(0.041)
$slr\_posi\_log_i$	1.283***	1.314***	0.978***	1.326***	1.312***
	(0.219)	(0.218)	(0.243)	(0.231)	(0.213)
$slr\_nega\_log_i$	-1.939**	-1.863*	-0.803	-1.698$^+$	-1.975***
	(0.664)	(0.757)	(0.628)	(0.925)	(0.562)
N	2894	2933	2946	2946	2946

$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

model that incorporates the above trade-off. We consider the case where the retailer chooses the forward shipping charge (κ) that will maximize its expected profit (π) from each sale.

Let π_{keep} (π_{return}) denote the unit profit of a sold item that is kept (returned) by a consumer. Unit expected profit can be expressed as $\pi = [1 - \Pr(return)] \times \pi_{keep} + \Pr(return) \times \pi_{return}$, where π_{keep} is simply the difference between unit price (p) and unit cost of producing and shipping the product (c), i.e. $\pi_{keep} = p - c$. In order to focus on our variable of interest, the shipping fee, we assume that $\underline{p}$, the price without an MBG policy, is exogenous to the model. With an MBG in place, the retailer adjust price upward according to the estimated value of an MBG such that $p = \underline{p}(1 + \frac{NV}{100})$. Assuming the retailer has average reputation indices, we have $NV = \delta_0 + \delta_\kappa \frac{100\kappa}{\underline{p}}$, where δ_κ and δ_0 are the coefficients on $shipping\_charge_i$ and I_i, and $\frac{100\kappa}{\underline{p}}$ converts κ into the percentage measure of the shipping charge in our empirical analysis. Thus, $\pi_{keep} = \underline{p} + \frac{\underline{p}\delta_0}{100} + \kappa\delta_\kappa - c$. If a consumer decides to return the purchased item, the retailer keeps the forward shipping charge, salvages the returned item, and incurs the unit cost. That is, $\pi_{return} = \kappa + s - c$, where s

denotes the salvage value. Let λ denote the return probability under free shipping, i.e. $\Pr(return|\kappa = 0) = \lambda$. Since $\kappa = p$ essentially means no return is allowed, we have $\Pr(return|\kappa = p) = 0$. For simplicity, we assume the return probability moves linearly in κ such that $\Pr(return) = \lambda(1 - \frac{\kappa}{p}) = \lambda - \frac{\lambda\kappa}{\underline{p} + \frac{p\delta_0}{100} + \kappa\delta_\kappa}$. Making the relevant substitutions, we have the following retailer profit function:

$$\max_{0 \leq \kappa \leq 0.2\underline{p}} \pi = [1 - \lambda + \frac{\lambda\kappa}{\underline{p} + \frac{p\delta_0}{100} + \kappa\delta_\kappa}](\underline{p} + \frac{p\delta_0}{100} + \kappa\delta_\kappa) + [\lambda - \frac{\lambda\kappa}{\underline{p} + \frac{p\delta_0}{100} + \kappa\delta_\kappa}] \times (\kappa + s) - c$$

We let the feasible range of shipping charge be between 0 and $0.2\underline{p}$ because our estimates of δ_0 and δ_κ are most reliable within this range – the same reason that Figure 4.1 has its x-axis between 0 and 20. The unit cost, c, can be normalized to zero since it does not affect the choice of the forward shipping fee.

We use the Nikon Coolpix L810 digital camera ("camera" hereafter) to illustrate the implementation of the above model. While δ_0 and δ_κ are available from Table 4.5, additional data from eBay is collected to calibrate $\underline{p}$ and s. Specifically, we collect the fixed-price transactions of the camera in new, open-box, and used conditions. The average price of the transactions in a new condition and without an MBG policy is used as an approximation for $\underline{p}$. The average prices of open-box and used transactions are used as the upper (s_H) and lower (s_L) bounds of s, giving $\underline{p} = \$190$, $s_H = \$180$ and $s_L = \$100$ (rounded to tens). The last parameter to calibrate is λ. Douthit et al. (2011) and Tuttle (2011) report that the return rate for consumer electronics is between 10% to 20%. Similar to s, we use these two numbers as the upper (λ_H) and lower (λ_L) bounds of λ. In Table 4.9, we show how the optimal forward shipping fee, κ^*, varies as a function of the salvage value and the return rate under free shipping. Note that the feasible range for κ is between \$0 and \$38, which is given by $0 \leq \kappa \leq 0.2\underline{p}$. If the optimized unit profit is smaller than the profit without an MBG (\$190), we replace the shipping fee with "No MBG" in the table.

Several observations can be made from Table 4.9. When the return rate is high, the retailer should either consider charging a high shipping fee or not offering an MBG. In

Table 4.9 Optimal Forward Shipping Fee

$\underline{p}$ = \$190		Return rate given free shipping					
		λ_L					λ_H
Salvage value		10%	12%	14%	16%	18%	20%
s_L	\$100	\$35	No MBG	No MBG	No MBG	No MBG	No MBG
	\$120	\$26	\$38	\$38	\$38	No MBG	No MBG
	\$140	\$17	\$35	\$38	\$38	\$38	\$38[1]
	\$160	\$8	\$26	\$38	\$38	\$38	\$38
s_H	\$180	\$0	\$17	\$29	\$38	\$38	\$38

[1] The upper bound of foreward shipping fee is \$38.

this case, consumers' increased WTP resulting from offering an MBG policy does not cover the loss incurred from handling returns. When the return rate is relatively low, how the retailer should set its forward shipping fee largely depends on the product's salvage value. Therefore, once a retailer commits to offering an MBG, the forward shipping fee should, if possible, be treated as a strategic decision variable. Since the optimal shipping fee is likely to vary across products, there is incentive for the retailer to customize its shipping charges this way[18].

4.6 Conclusion

This study represents the first in-depth empirical examination of the value of an MBG policy in online retailing, and we provide new insights into the influence of shipping charges and seller reputation on this value. As revealed by the empirical analysis, the value of an MBG for consumer electronics sold online is, on average, smaller than 7% of the total product value (Table 4.7). This is substantially lower than the 20% to 30% for catalog apparel products estimated by Anderson et al. (2009). This difference may be explained by the varied degree of word-of-mouth information available for the

[18] As a cautionary note to the above discussion, we clarify that the extent to which a retailer can adjust the partition between base product price and shipping charge may be constrained by the OEM. It appears that some OEMs allow a minimal degree of freedom for their retailers to change the product price. For example, most Apple products and Bose headphones are nearly identically priced across major retailers. In these cases, the strategic optimization of the shipping fee discussed in this section may not be a realistic option.

two product categories. For consumer electronics, there is abundant product review information on the web (each product included in our sample has its own review page on eBay). Consumers could also augment this information by searching on other websites such as Amazon and CNET. As a result, much of the uncertainty regarding product fit may potentially be resolved before purchase, which limits the value of an MBG policy. In comparison, word-of-mouth information for apparel products is much less available, given their wide variety. Furthermore, it is likely to be more difficult for consumers to use word-of-mouth information to determine the fit of an apparel product.

Specific to on-line retailing, we find that consumers treat the forward shipping fee as an implicit restocking fee. As a result, it erodes the value of an MBG, and the magnitude of this negative effect is large. To illustrate, if 20% of a product's total price is attributed to shipping, the value of an MBG could be statistically indifferent from zero for some sellers. Given that a non-refundable shipping fee also affects the probability and the cost of returns, retailers may consider using it strategically to optimize profit as demonstrated in Section 4.5. We also contribute to the e-reputation literature by showing how consumers may infer the credibility of a seller's MBG promise from its reputation indices. This result leads to the finding that the most active sellers, who accumulate more positive as well as negative reviews, are not necessarily the ones who enjoy the highest payoff from offering an MBG.

Our work also helps illustrate the problem of using overly simplistic methods to estimate the value of MBGs. Table 4.10 provides a by-category comparison of the MBG values estimated through different methods. The naïve approach first takes a simple mean difference of MBG and no-MBG transactions for each product and then summarizes these mean differences by category (also shown in Table 4.2). Although it is very straightforward to implement, this approach does not account for differences in shipping fees nor the sellers' endogenous choice to offer an MBG policy. Furthermore, there is no clear pattern that the naïve approach over or under estimates the value of an MBG. It may have particularly poor performance when MGB and non-MBG transactions exhibit considerable differences in other dimensions. For example, in the scanner category, sellers with an MBG policy happen to be the ones who charge higher forward shipping fees and have lower reputation. As a result, the naïve approach significantly under-estimates the value of an MBG for the scanner category. As compared with the naïve approach, the OLS approach takes the contingency of the MBG value into account, but it generally over-estimates the value of an MBG policy. The extent of exaggeration for the OLS approach can be quite large, which is evident from comparing the OLS and FIML estimates. In summary, while return managers could use simple methods (such as the naïve and OLS approaches) to produce a quick preview on the value of an MBG, a more sophisticated analysis is required to model confounding factors and account for endogeneity.

Table 4.10 Estimates of MBG Value by Category

	Naïve	OLS (with controls)		FIML	
Product ($) category	Est. ($)	Est. ($)[1]	Pct. Diff.[2]	Est. ($)[1]	Pct. Diff.[2]
Audio System	4.08	8.26	103%	5.67	39%
Calculator	8.18	6.56	-20%	4.72	-42%
Camcorder	12.07	23.97	99%	17.43	44%
Camera	10.84	18.38	70%	12.83	18%
Camera Lens	10.88	17.70	63%	12.06	11%
Vacuum Cleaner	16.2	20.95	29%	15.08	-7%
DVD Player	2.48	3.79	53%	2.72	10%
GPS	10.41	11.46	10%	8.15	-22%
Headphone	6.36	4.00	-37%	2.61	-59%
Keyboard	4.65	3.91	-16%	2.46	-47%
Modem	2.89	3.33	15%	2.12	-27%
MP3 Player	6.99	9.66	38%	6.21	-11%
Printer	5.04	3.73	-26%	2.25	-55%
Scanner	5.58	22.01	294%	14.95	168%
Tablet	21.31	15.38	-28%	10.50	-51%
Webcam	2.83	5.26	86%	3.68	30%

[1] The OLS and FIML estimates are obtained by computing η_i for each observation, converting it to the dollar measure, and summarizing the dollar measures by category.
[2] The percentage differences compare OLS and FIML estimates with the naïve estimates.

Chapter 5

Future Research

While our consumer opportunism model in Chapter 2 captures the essential elements of the consumer returns context with opportunism, there are other aspects that could be considered in future research. In an effort to discourage policy abusers, some retailers limit the maximum dollar value returnable by a single consumer within a period of time[1]. Evaluating the effectiveness of this practice requires a multi-period framework, which is an interesting area of future inquiry. In addition, a competitive model might provide a nice complement to our insights, although to remain tractable it would require a simplified analytical framework that would sacrifice some of the richness of our single-retailer analysis.

Given the evidence in Chapter 3 that using transactions data significantly improves returns forecasting, an interesting extension to this stream of work is to empirically estimate the cost savings that an improved product return forecast provides. Through exploring the drivers of return probability, we have provided guidance on how to better target buyer assistance. Future studies using more complete data sets can extend this analysis by incorporating a retailer's current allocation of buyer assistance efforts as an endogenous decision by the retailer when estimating its impact on product return rates.

Chapter 4 has shown how retailers could better management their returns with a clearer picture of the trade-off involved with shipping charges. Similarly, a longer

[1]For example, Best Buy explicitly states on their website that they reserve the right to refuse further returns if the consumer's return volume is considered too “excessive”.

return time window also has both cost and benefit implications. Our eBay data has limited variation in return time window, which restricts us from quantifying the value of time for consumers to evaluate a product. Future research could experimentally examine how consumers react to different lengths of return time windows. This information could then facilitate analytical models of optimal time window policies.

Bibliography

Akçay, Y., T. Boyacı, and D. Zhang (2013). Selling with money-back guarantees: The impact on prices, quantities, and retail profitability. *Production and Operations Management 22*(4), 777–791.

Anderson, E., K. Hansen, and D. Simester (2009). The option value of returns: Theory and empirical evidence. *Marketing Science 28*(3), 405–423.

Aviv, Y. (2007). On the benefits of collaborative forecasting partnerships between retailers and manufacturers. *Management Science 53*(5), 777–794.

Bajari, P. and A. Hortaçsu (2004). Economic insights from internet auctions. *The Journal of Economic Literature 42*(2), 457–486.

Blanchard, D. (2007). Supply chains also work in reverse. *Industry Week 256*(5), 48.

Bodea, T., M. Ferguson, and L. Garrow (2009). Data set – Choice-based revenue management: Data from a major hotel chain. *Manufacturing & Service Operations Management 11*(2), 356–361.

Boehmke, F. J., D. S. Morey, and M. Shannon (2006). Selection bias and continuous-time duration models: Consequences and a proposed solution. *American Journal of Political Science 50*(1), 192–207.

Chircu, A. and V. Mahajan (2006). Managing electronic commerce retail transaction costs for customer value. *Decision Support Systems 42*(2), 898–914.

Chu, J. and P. K. Chintagunta (2009). Quantifying the economic value of warranties in the us server market. *Marketing Science 28*(1), 99–121.

Chu, W., E. Gerstner, and J. D. Hess (1998). Managing dissatisfaction. *Journal of Service Research 1*(2), 140–155.

Clottey, T., W. Benton, and R. Srivastava (2012). Forecasting product returns for remanufacturing operations. *Decision Sciences 43*(4), 589–614.

De Brito, M. P. and R. Dekker (2003). Modelling product returns in inventory control: Exploring the validity of general assumptions. *International Journal of Production Economics 81-82*, 225–241.

Dimoka, A., Y. Hong, and P. Pavlou (2012). On product uncertainty in online markets: theory and evidence. *MIS Quarterly 36*(2), 395–426.

Douthit, D., M. Flach, and V. Agarwal (2011). Reducing the quantity and cost of product returns in consumer electronics. *Accenture*.

Ferguson, M., V. D. R. Guide, and G. C. Souza (2006). Supply chain coordination for false failure returns. *Manufacturing & Service Operations Management 8*(4), 376–393.

Ghosh, B. and M. Galbreth (2013). The impact of consumer attentiveness and search costs on firm quality disclosure: A competitive analysis. *Management Science Forthcoming*.

Griffis, S., S. Rao, T. Goldsby, and T. Niranjan (2012). The customer consequences of returns in online retailing: An empirical analysis. *Journal of Operations Management 30*(4), 282–294.

Grind, K. (2013). Retailer REI ends era of many happy returns. *Wall Street Journal* (September 15, 2013).

Gu, B., J. Park, and P. Konana (2012). Research note – The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research 23*(1), 182–196.

Guajardo, J. A., M. A. Cohen, S.-H. Kim, and S. Netessine (2012). Impact of performance-based contracting on product reliability: an empirical analysis. *Management Science 58*(5), 961–979.

Guide, V., G. Souza, L. Van Wassenhove, and J. Blackburn (2006). Time value of commercial product returns. *Management Science 52*(8), 1200–1214.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society 46*(6), 1251–1271.

Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources 32*(3), 441–462.

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

Hess, J. D., W. Chu, and E. Gerstner (1996). Controlling product returns in direct marketing. *Marketing Letters 7*(4), 307–317.

Hess, J. D. and G. E. Mayhew (1997). Modeling merchandise returns in direct marketing. *Journal of Direct Marketing 11*(2), 20–35.

Hjort, K. and B. Lantz (2012). (r) e-tail borrowing of party dresses: an experimental study. *International Journal of Retail & Distribution Management 40*(12), 997–1012.

Houser, D. and J. Wooders (2006). Reputation in auctions: Theory, and evidence from ebay. *Journal of Economics & Management Strategy 15*(2), 353–369.

Hsiao, L. and Y. Chen (2012). Returns policy and quality risk in e-business. *Production and Operations Management 21*(3), 489–503.

Kahn, B. E. (1998). Dynamic relationships with customers: high-variety strategies. *Journal of the Academy of Marketing Science 26*(1), 45–53.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The statistical analysis of failure time data* (2 ed.). New York: John Wiley & Sons.

Kalish, S. (1985). A new product adoption model with price, advertising, and uncertainty. *Management science 31*(12), 1569–1585.

Ketzenberg, M. E. and R. A. Zuidwijk (2009). Optimal pricing, ordering, and return policies for consumer goods. *Production and Operations Management 18*(3), 344–360.

Kotz, S., N. Balakrishnan, and N. L. Johnson (2004). *Continuous multivariate distributions, models and applications*, Volume 334. Wiley-Interscience.

Kumar, N., V. D. R. Guide Jr, and L. Van Wassenhove (2002). Managing product returns at Hewlett Packard, INSEAD teaching case 05/2002-4940.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society 51*(2), 507–12.

Lehmann, D. R. (1998). Customer reactions to variety: Too much of a good thing? *Journal of the Academy of Marketing Science 26*(1), 62–65.

Leng, M. and R. Becerril-Arreola (2010). Joint pricing and contingent free-shipping decisions in B2C transactions. *Production and Operations Management 19*(4), 390–405.

Lewis, M. (2006). The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. *Journal of Retailing 82*(1), 13–23.

Li, K. J., D. K. Fong, and S. H. Xu (2011). Managing trade-in programs based on product characteristics and customer heterogeneity in business-to-business markets. *Manufacturing & Service Operations Management 13*(1), 108–123.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Number 3. Cambridge University Press.

Mazar, N., O. Amir, and D. Ariely (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research 45*(6), 633–644.

Mazar, N. and D. Ariely (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing 25*(1), 117–126.

McWilliams, B. (2012). Money-back guarantees: Helping the low-quality retailer. *Management Science 58*(8), 1521–1524.

McWilliams, G. (2004). Analyzing customers, best buy decides not all are welcome. *The Wall Street Journal Online* (November 8), A1.

Miller, J. W., W. R. Stromeyer, and M. A. Schwieterman (2013). Extensions of the johnson-neyman technique to linear models with curvilinear effects: Derivations and analytical tools. *Multivariate Behavioral Research 48*(2), 267–300.

National Retail Federation (2011a, December). Holiday consumer intentions and actions survey. http://www.nrf.com.

National Retail Federation (2011b, October). Return fraud survey. http://www.nrf.com.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy 78*(2), 311–329.

Ni, J., S. A. Neslin, and B. Sun (2012). Database submission - the ISMS durable goods data sets. *Marketing Science 31*(6), 1008–1013.

Obloj, T. and L. Capron (2011). Role of resource gap and value appropriation: Effect of reputation gap on price premium in online auctions. *Strategic Management Journal 32*(4), 447–456.

Ofek, E., Z. Katona, and M. Sarvary (2011). Bricks and clicks: The impact of product returns on the strategies of multichannel retailers. *Marketing Science 30*(1), 42–60.

Petersen, J. and V. Kumar (2009). Are product returns a necessary evil? Antecedents and consequences. *Journal of Marketing 73*(3), 35–51.

Pince, C., M. Ferguson, and B. Toktay (2013). Extracting maximum value from consumer returns: Allocating between selling refurbished product and meeting warranty demand. *Georgia Institute of Technology Working Paper*.

Prieger, J. E. (2002). A flexible parametric selection model for non-normal data with application to health care usage. *Journal of Applied Econometrics 17*(4), 367.

Puhani, P. (2000). The heckman correction for sample selection and its critique. *Journal of economic surveys 14*(1), 53–68.

Randall, T., K. Ulrich, and D. Reibstein (1998). Brand equity and vertical product line extent. *Marketing Science 17*(4), 356–379.

Renneboog, L. and C. Spaenjers (2013). Buying beauty: On prices and returns in the art market. *Management Science 59*(1), 36–53.

Resnick, P., R. Zeckhauser, J. Swanson, and K. Lockwood (2006). The value of reputation on ebay: A controlled experiment. *Experimental Economics 9*(2), 79–101.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy 82*(1), 34–55.

Rosenbaum, M. S. and R. Kuntze (2005). Looking good at the retailer's expense: investigating unethical retail disposition behavior among compulsive buyers. *Journal of Retailing and Consumer Services 12*(3), 217–225.

Shulman, J., A. Coughlan, and R. Savaskan (2009). Optimal restocking fees and information provision in an integrated demand-supply model of product returns. *Manufacturing & Service Operations Management 11*(4), 577–594.

Shulman, J. D., A. T. Coughlan, and R. C. Savaskan (2010). Optimal reverse channel structure for consumer product returns. *Marketing Science 29*(6), 1071–1085.

Shulman, J. D., A. T. Coughlan, and R. C. Savaskan (2011). Managing consumer returns in a competitive environment. *Management Science 57*(2), 347–362.

Smith, M. D. (2003). Modelling sample selection using archimedean copulas. *The Econometrics Journal 6*(1), 99–123.

Speights, D. and M. Hilinski (2005). Return fraud and abuse: How to protect profit. *Retail Issues Letter 17*(1), 1–6.

Standifird, S. (2001). Reputation and e-commerce: ebay auctions and the asymmetrical impact of positive and negative ratings. *Journal of Management 27*(3), 279–295.

Su, X. (2009). Consumer returns policies and supply chain performance. *Manufacturing & Service Operations Management 11*(4), 595–612.

Subramanian, R. and R. Subramanyam (2012). Key factors in the market for remanufactured products. *Manufacturing & Service Operations Management 14*(2), 315–326.

Toktay, B., E. A. van der Laan, and M. P. Brito (2003). Managing product returns: The role of forecasting. Technical report, Erasmus Research Institute of Management (ERIM).

Toktay, L., L. Wein, and S. Zenios (2000). Inventory management of remanufacturable products. *Management Science 46*(11), 1412–1426.

Tuttle, B. (2011). Consumers to tech companies: Please, stop the flood of new gadgets. *Time* (December 14, 2011).

Vella, F. (1998). Estimating models with sample selection bias: a survey. *Journal of Human Resources 33*(1), 127–169.

Vella, F. and M. Verbeek (1999). Estimating and interpreting models with endogenous treatment effects. *Journal of Business & Economic Statistics 17*(4), 473–478.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society 48*(4), 817–838.

Willems, S. P. (2008). Data set: Real-world multiechelon supply chains used for inventory optimization. *Manufacturing & service operations management 10*(1), 19–23.

Wood, S. L. (2001). Remote purchase environments: The influence of return policy leniency on two-stage decision processes. *Journal of Marketing Research 38*(2), 157–169.

Appendix A

Proof of Lemma 2.1

We show in Equation (2.3) the retailer's profit function given that both segments are willing to buy (π_1 and π_2 in Table A). To ensure that serving both segments is indeed the retailer's optimal choice, we also need to calculate its profit when serving only one segment and show that such profit is strictly dominated by serving both segments. The scenario where the retailer sells only to ordinary consumers can be eliminated as follows. Observe that $U_{op} - U_{or}$ is equal to $\frac{\beta(p-f)^2}{2(1-\beta)}$ for $\frac{p-f}{1-\beta} < 1$ and to $\frac{\beta-(1+f-p)^2}{2}$ for $\frac{p-f}{1-\beta} > 1$. We can conclude $U_{op} > U_{or}$ from both cases. Thus, if opportunistic consumers are unwilling to buy ($U_{op} < 0$), ordinary consumers are also unwilling to buy ($U_{or} < U_{op} < 0$). Therefore, if the retailer considers serving only one segment, it must be the segment of opportunistic consumers (π_3 and π_4 in Table A)[1]. Intuitively, this might happen when there are many opportunistic consumers – γ is close to its upper limit, $\frac{1}{2}$.

Table A.1 Retailer Profit Function (Main Model)

	$p - f \leq 1 - \beta$	$p - f \geq 1 - \beta$
$U_{or} \geq 0, U_{op} \geq 0$	$\pi_1 = (1-\gamma)\pi_{or} + \gamma\pi_{op1}$	$\pi_2 = (1-\gamma)\pi_{or} + \gamma\pi_{op2}$
$U_{or} < 0, U_{op} \geq 0$	$\pi_3 = \gamma\pi_{op1}$	$\pi_4 = \gamma\pi_{op2}$

[1] If both U_{op} and U_{or} are negative, no consumer will buy and the retailer's profit is zero. This case is strictly dominated by serving both segments since the optimal profits in Table A.2 are positive.

Table A.2 Optimal Return Policies (Main Model)

	Case I	Case II	Case III
	low β	high β, low γ	high β, high γ
p^*	$\frac{1+\beta^2(1-2\gamma)^2+s_1^2(1-\beta)^2(1-\gamma)^2+2s_1s_2(1-\beta)(1-\gamma)\gamma+s_2^2\gamma^2-\beta(2-4\gamma)}{2(1-\beta+2\beta\gamma)^2}$	$\frac{1+s_1^2-2\left(2+s_1+s_1^2\right)\gamma+(5+s_1(2+s_1))\gamma^2}{2(1-2\gamma)^2}$	$\frac{2(1-\beta)+\beta^2}{2}$
f^*	$\frac{(-1+\beta+s_1(1-\beta)(1-\gamma)+s_1\gamma-2\beta\gamma)^2}{2(1-\beta+2\beta\gamma)^2}$	$\frac{(1-s_1)^2(1-\gamma)^2}{2(1-2\gamma)^2}$	$\frac{\beta^2}{2}$
π^*	$\frac{(1-2c+s_1^2)(1-\beta)^2+2(1-\beta)(s_1(s_2-s_1(1-\beta))+\beta-2c\beta)\gamma+(s_2-s_1(1-\beta))^2\gamma^2}{2(1-\beta)(1-\beta(1-2\gamma))}$	$\frac{1+s_1^2-(21+s_1+s_1^2-s_2)\gamma+((1+s_1)^2-4s_2)\gamma^2-c(2-4\gamma)}{2-4\gamma}$	$\frac{2(s_1+\beta)(1-\beta)(1-\gamma)+2s_2\gamma+\beta^2-2c}{2}$

Case I conditions: $\beta \in [0, 1-s_1] \cup \gamma \in [0, \frac{1}{2}]$, or $\beta \in [1-s_1, 1-s_2] \cup \gamma \in [\widehat{\gamma}, \frac{1}{2}]$, or $\beta \in [1-s_2, \frac{2-s_1-s_2}{2-s_1}] \cup \gamma \in [\frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}, \frac{1}{2}]$

Case II conditions: $\beta \in [1-s_1, 1-s_2] \cup \gamma \in [0, \widehat{\gamma}]$, or $\beta \in [1-s_2, 1] \cup \gamma \in [0, \frac{\beta+s_1-1}{2\beta+s_1-1}]$

Case III conditions: $\beta \in [1-s_2, \frac{2-s_1-s_2}{2-s_1}] \cup \gamma \in [\frac{\beta+s_1-1}{2\beta+s_1-1}, \frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}]$, or $\beta \in [\frac{2-s_1-s_2}{2-s_1}, 1] \cup \gamma \in [\frac{\beta+s_1-1}{2\beta+s_1-1}, \frac{1}{2}]$

where $\widehat{\gamma}=\frac{s_2^2+4s_1^2(1-\beta)-2s_1(1+3s_2-3\beta)(1-\beta)-(-1+\beta)^2+4s_2(1-\beta)(1-2\beta)-\sqrt{(s_2+\beta-1)^2(s_2^2-6s_2(1-\beta)+(1-\beta)^2+4s_1(1-\beta)(1+s_2+\beta)+4s_1^2(1-\beta^2))}}{4(s_1-s_2)^2+\beta+(2-s_1)(s_1-2s_2)\beta-(1+2s_1-4s_2)\beta^2}$

In the following, we use a two-step procedure to prove Lemma 2.1. First, we assume the retailer serves both segments and derive p^*, f^*, and π^*. Second, we show that within the parameter space of interest, $\gamma \in [0, \frac{1}{2}]$ and $\beta \in [0, 1]$, serving only the opportunistic segment is strictly dominated given $c < Min[\frac{1-s_1+2s_1^2}{2}, \frac{2s_2+s_1^2(-1+s_2(2+s_2))}{4s_2}]$.

$U_{or} \geq 0$ and $p - f \leq 1 - \beta$. The Hessian of π_1 is given by $\begin{pmatrix} -\frac{2-2\beta+2\beta\gamma}{1-\beta} & \frac{2-2\beta+2\beta\gamma}{1-\beta} \\ \frac{2-2\beta+2\beta\gamma}{1-\beta} & -\frac{2-2\beta+2\beta\gamma}{1-\beta} \end{pmatrix}$, which is negative definite given $\gamma \in [0, \frac{1}{2}]$ and $\beta \in [0, 1]$; i.e. the profit function is jointly strictly concave in p and f. Rewriting π_1 into the Lagrangian format, we have $L_1 = \pi_1 + \mu_1 U_{or} + \mu_2(1 - \beta - p + f)$. It is straightforward to verify that π_1^* is obtained either with only $U_{or} = 0$ or both constraints binding. p_1^*, f_1^*, and π_1^* in the former scenario is given in Case I in Table A.2; the corresponding parameter conditions are $\beta \in [0, 1 - s_1] \cup \gamma \in [0, \frac{1}{2}]$ or $\beta \in [1 - s_1, \frac{2-s_1-s_2}{2-s_1}] \cup \gamma \in [\frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}, \frac{1}{2}]$. p_1^*, f_1^*, and π_1^* in the latter scenario is given in Case III in Table A.2; the corresponding parameter conditions are $\beta \in [\frac{2-s_1-s_2}{2-s_1}, 1] \cup \gamma \in [0, \frac{1}{2}]$ or $\beta \in [1 - s_1, \frac{2-s_1-s_2}{2-s_1}] \cup \gamma \in [0, \frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}]$. Since f_1^* in both cases is strictly positive and $p_1^* > f_1^*$ must be true given the second constraint, we also have $p_1^* > f_1^* > 0$.

$U_{or} \geq 0$ and $p - f \geq 1 - \beta$. The Hessian of π_2 is given by $\begin{pmatrix} -2(1-\gamma) & 2(1-\gamma) \\ 2(1-\gamma) & -2(1-\gamma) \end{pmatrix}$, which is negative definite given $\gamma \in [0, \frac{1}{2}]$; i.e. the profit function is jointly strictly concave in p and f. Similarly, π_2^* is also obtained either with with only $U_{or} = 0$ or both constraints binding. p_2^*, f_2^*, and π_2^* in the former scenario is given in Case II in Table A.2; the corresponding parameter conditions are $\beta \in [1 - s_1, 1] \cup \gamma \in [0, \frac{\beta+s_1-1}{2\beta+s_1-1}]$. p_2^*, f_2^*, and π_2^* in the latter scenario is given in Case III in Table A.2; the corresponding parameter conditions are $\beta \in [1-s_1, 1] \cup \gamma \in [\frac{\beta+s_1-1}{2\beta+s_1-1}, \frac{1}{2}]$ or $\beta \in [0, 1-s_1] \cup \gamma \in [0, \frac{1}{2}]$. Again, $p_1^* > f_1^* > 0$ is verified.

Compare π_1^* with π_2^*. Observe that $\pi_1^* = \pi_2^*$ when $p - f = 1 - \beta$. Therefore, if π_1^* is obtained with $p - f = 1 - \beta$, it must be true that $\pi_1^* \leq \pi_2^*$, *vice versa.* As a result, we can immediately show that (1) $\pi_1^* > \pi_2^*$ and Case I is optimal when $\beta \in [0, 1 - s_1]$ $\cup$ $\gamma \in [0, \frac{1}{2}]$ or $\beta \in [1 - s_1, 1 - s_2] \cup \gamma \in [\frac{\beta+s_1-1}{2\beta+s_1-1}, \frac{1}{2}]$ or $\beta \in [1 - s_2, \frac{2-s_1-s_2}{2-s_1}] \cup$

$\gamma \in [\frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}, \frac{1}{2}]$, (2) $\pi_1^* < \pi_2^*$ and Case II is optimal when $\beta \in [1-s_1, 1-s_2]$ $\cup$ $\gamma \in [0, \frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}]$ or $\beta \in [1-s_2, 1] \cup \gamma \in [0, \frac{\beta+s_1-1}{2\beta+s_1-1}]$, and (3) $\pi_1^* = \pi_2^*$ and Case III is optimal when $\beta \in [1-s_2, \frac{2-s_1-s_2}{2-s_1}] \cup \gamma \in [\frac{\beta+s_1-1}{2\beta+s_1-1}, \frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}]$ or $\beta \in [\frac{2-s_1-s_2}{2-s_1}, 1] \cup \gamma \in [\frac{\beta+s_1-1}{2\beta+s_1-1}, \frac{1}{2}]$. The only parameter space left is $\beta \in [1-s_1, 1-s_2] \cup$ $\gamma \in [\frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}, \frac{\beta+s_1-1}{2\beta+s_1-1}]$, where both π_1^* and π_2^* are obtained without $p-f \geq 1-\beta$ binding. By explicitly examining the sign of $\pi_1^* - \pi_2^*$, we can show that it is positive when $\widehat{\gamma} < \gamma$ and negative otherwise. The above gives us all the β and γ conditions in Table A.2.

Compare $Max[\pi_1^*, \pi_2^*]$ with $Max[\pi_3^*, \pi_4^*]$. Next, we rule out the possibility that the retailer wants to serve only the opportunistic segment; i.e. $Max[\pi_1^*, \pi_2^*] < Max[\pi_3^*, \pi_4^*]$. The respective optimization problems for π_3 and π_4 are rather straightforward. Specifically, $\pi_3^* = \frac{(1+s_2^2-2c(1-\beta)-\beta)\gamma}{2(1-\beta)}$ when $\beta \in [1, 1-s_2]$ and $\pi_3^* = \frac{(\beta+2s_2-2c)\gamma}{2}$ when $\beta \in [1, 1-s_2]$, while π_4^* is always obtained with $p-f \geq 1-\beta$ binding. Since $\pi_3^* = \pi_4^*$ given $p-f = 1-\beta$, we conclude that π_3^* dominates π_4^*. Now, we need to verify that $Max[\pi_1^*, \pi_2^*] > \pi_3^*$. We separate this comparison into the sub-scenarios where $\beta < 1-s_1$, $1-s_1 < \beta < 1-s_2$ and $\beta > 1-s_2$. When $\beta < 1-s_1$, $Max[\pi_1^*, \pi_2^*] \geq \pi_1^* \geq \pi_{CaseI}^*$ (the optimal profit in Case I). It is easy to verify that $\pi_{CaseI}^* - \pi_3^*$ is inverted U-shaped in s_2, which implies that the minimum of $\pi_{case1}^* - \pi_3^*$ is either at $s_2 = 0$ or $s_2 = s_1$. Given $c < \frac{1-s_1+2s1^2}{2}$, both $\pi_{CaseI}^* - \pi_3^*|_{s_2 \to 0} > 0$ and $\pi_{CaseI}^* - \pi_3^*|_{s_2 \to s_1} > 0$, which ensures $\pi_{CaseI}^* > \pi_3^*$. Similarly, when $1-s_1 < \beta < 1-s_2$, $Max[\pi_1^*, \pi_2^*] \geq \pi_1^* \geq \pi_{CaseI}^*$ and $\pi_{CaseI}^* - \pi_3^*$ is inverted U-shaped in s_2. Given $c < \frac{2s_2+s_1^2(-1+s_2(2+s_2))}{4s_2}$, again both $\pi_{CaseI}^* - \pi_3^*|_{s_2 \to 0} > 0$ and $\pi_{CaseI}^* - \pi_3^*|_{s_2 \to s_1} > 0$. When $\beta > 1-s_2$, $Max[\pi_1^*, \pi_2^*] \geq \pi_1^* \geq \pi_{CaseIII}^*$ (the optimal profit in Case 3). $\pi_{CaseIII}^*$ is strictly larger than π_3^* given the range of β. To summarize, as long as $c < Min[\frac{1-s_1+2s1^2}{2}, \frac{2s_2+s_1^2(-1+s_2(2+s_2))}{4s_2}]$, retailer is strictly better off with serving both segments and the specific optimal strategies are listed in Table A.2.

Appendix B

Proof of Proposition 2.1

In Case I, $\frac{\partial p^*}{\partial \gamma} = -\frac{(1-\beta)(s_1-s_2+s_1\beta)(s_1(1-\beta)(1-\gamma)+s_2\gamma)}{(1+\beta(-1+2\gamma))^3} < 0$, $\frac{\partial f^*}{\partial \gamma} = \frac{(1-\beta)(s_1-s_2+s_1\beta)(1-\beta-s_2\gamma+2\beta\gamma)}{(1+\beta(-1+2\gamma))^3} - \frac{s_1(1-\beta)(1-\gamma)}{1+\beta(-1+2\gamma))^3} > 0$, and $\left|\frac{\partial p^*}{\partial \gamma}\right| - \left|\frac{\partial f^*}{\partial \gamma}\right| = -\frac{(1-\beta)(s_1-s_2+s_1\beta)(1-\beta-2s_1(1-\beta)(1-\gamma)+2s_2\gamma-2\beta\gamma)}{(1+\beta(-1+2\gamma))^3} < 0$. In Case II, $\frac{\partial p^*}{\partial \gamma} = -\frac{(1-s_1)(s_1(1-\gamma)-\gamma)}{(1-2\gamma)^3} < 0$, $\frac{\partial f^*}{\partial \gamma} = \frac{(1-s_1)^2(1-\gamma)}{(1-2\gamma)^3} > 0$, and $\left|\frac{\partial p^*}{\partial \gamma}\right| - \left|\frac{\partial f^*}{\partial \gamma}\right| = -\frac{(1-s_1)(1+2s_1(-1+\gamma))}{(1-2\gamma)^3} < 0$. In Case III, γ has no impact on p^* and f^*. Thus, $\frac{\partial p^*}{\partial \gamma} \leq 0$, $\frac{\partial f^*}{\partial \gamma} \geq 0$, and $\left|\frac{\partial p^*}{\partial \gamma}\right| \leq \left|\frac{\partial f^*}{\partial \gamma}\right|$.

Appendix C

Proof of Proposition 2.2 and Corollary 2.1

In Case I, $\frac{\partial \pi^*}{\partial \gamma} = \frac{s_1(s_1-s_2)(1-\beta)^2-(s_2-s_1(1-\beta))^2((1-\beta)\gamma+\beta\gamma^2)}{(-1+\beta)(-1+\beta-2\beta\gamma)^2}$. Given $\beta < \widehat{\beta}$, $\frac{\partial \pi^*}{\partial \gamma}$ is strictly negative. For $\beta > \widehat{\beta}$, $\frac{\partial \pi^*}{\partial \gamma} < 0$ when $\gamma < \frac{s_2}{s_2-s_1(1-\beta)} - \frac{1}{\beta}$ and $\frac{\partial \pi^*}{\partial \gamma} > 0$ otherwise. In Case II, $\frac{\partial \pi^*}{\partial \gamma} = -\frac{(1+2s_1+s_1^2-4s_2)(1-2\gamma)^2+(1-s_1)^2}{4(1-2\gamma)^2}$, which is positive only when $\beta > \frac{1-s_1+\sqrt{(1+s_1)^2-4s_2}}{2}$ and $\gamma > \frac{1-s_1-\sqrt{(1+s_1)^2-4s_2}}{2\sqrt{(1+s_1)^2-4s_2}}$. In Case III, $\frac{\partial \pi^*}{\partial \gamma} = s_2 - (s_1+\beta)(1-\beta)$, which is positive when $\beta > \frac{1-s_1+\sqrt{(1+s_1)^2-4s_2}}{2}$. Taking the above together, we have shown that (1) when $\beta < \widehat{\beta}$, $\frac{\partial \pi^*}{\partial \gamma} < 0$, (2) when $\widehat{\beta} < \beta < \frac{1-s_1+\sqrt{(1+s_1)^2-4s_2}}{2}$, π^* is U-shaped in γ with $\frac{\partial \pi^*}{\partial \gamma}|_{\gamma \to \frac{s_2}{s_2-s_1(1-\beta)}-\frac{1}{\beta}} = 0$, and (3) when $\beta > \frac{1-s_1+\sqrt{(1+s_1)^2-4s_2}}{2}$, π^* is U-shaped in γ with $\frac{\partial \pi^*}{\partial \gamma}|_{\gamma \to \frac{1-s_1-\sqrt{(1+s_1)^2-4s_2}}{2\sqrt{(1+s_1)^2-4s_2}}} = 0$. In addition, it is easy to verify that $\frac{\partial^2 \pi^*}{\partial^2 \gamma} > 0$ holds globally.

Appendix D

Proof of Proposition 2.3

In Case I, $\frac{\partial \pi^*}{\partial \beta} = -\frac{\gamma(s_1(1-\beta)(1-\gamma)+s_2\gamma)(s_1(1-\beta)(1-\gamma)-s_2(1-\beta-\gamma+2\beta\gamma))}{(1-\beta)^2(1-\beta+2\beta\gamma)^2}$, which is positive for either $\beta \in [\frac{s_1-s_2}{s_1}, 1-s_2] \cup \gamma \in [\frac{(s_1-s_2)(-1+\beta)}{s_2-s_1(1-\beta)-2s_2\beta}, \frac{1}{2}]$ or $\beta \in [1-s_2, \frac{2-s_1-s_2}{2-s_1}] \cup \gamma \in [\frac{(1-\beta)(\beta+s_1-1)}{(s_1+2\beta)(1-\beta)-s_2}, \frac{1}{2}]$ and negative otherwise. In Case II, $\frac{\partial \pi^*}{\partial \beta} = 0$. In Case III, $\frac{\partial \pi^*}{\partial \beta} = (1-\gamma)(1-s_1-\beta)+\beta\gamma$, which is strictly positive given the boundary conditions of Case 3. In summary, we have $\frac{\partial \pi^*}{\partial \beta} > 0$ for $\beta \in [\frac{s_1-s_2}{s_1}, 1]$ and $\gamma \in [Max[\frac{(s_1-s_2)(-1+\beta)}{s_2-s_1(1-\beta)-2s_2\beta}, \frac{\beta+s_1-1}{2\beta+s_1-1}], \frac{1}{2}]$, $\frac{\partial \pi^*}{\partial \beta} = 0$ for $\beta \in [1-s_1, 1]$ and $\gamma \in [0, Min[\widehat{\gamma}, \frac{\beta+s_1-1}{2\beta+s_1-1}]]$, and $\frac{\partial \pi^*}{\partial \beta} < 0$ otherwise.

Appendix E

Proof of Proposition 2.4

Solving the retailer's optimization problem in the extended model yields the following lemma:

Lemma E.1. *The retailer's optimal price (p^*) and restocking fee (f^*) depend on the extent (γ) and the benefit (β) of opportunism, which together define three regions of the solution space. Table E summarizes p^*, f^*, and optimal profit (π^*), as well as the boundary conditions for each region.*

Analysis for Lemma E.1 follows exactly the same steps as for Lemma 2.1. Again, a mild condition on the unit production cost, $c < \frac{1-s+s^2}{2}$, ensures that it is optimal for the retailer to serve both segments.

In Case I, $\frac{\partial \pi^*}{\partial \beta} = \frac{\gamma(s(1+\beta(-1+\gamma))-\beta\gamma\delta)(s\beta\gamma+(-1+\beta-\beta\gamma)\delta)}{(1-\beta)^2(1-\beta+2\beta\gamma)^2}$, which is positive for $\beta \in [\frac{2\delta}{s+\delta}, 1]$ and $\gamma \in [\frac{\delta(1-\beta)}{\beta(s-\delta)}, \frac{1}{2}]$ and negative otherwise. In Case II, $\frac{\partial \pi^*}{\partial \beta} = -\gamma\delta$. In Case III, $\frac{\partial \pi^*}{\partial \beta} = 1-s(1-\gamma)-\beta(1-2\gamma)-\gamma(1+\delta)$, which is positive for either $\beta \in [1-s+\delta, \frac{2(1-s)}{2-s-\delta}]$ and $\gamma \in [\frac{\delta(1-\beta)}{\beta(s-\delta)}, \frac{(1-\beta)(s+\beta-1)}{\beta(2-s-2\beta+\delta)}]$ or $\beta \in [\frac{2(1-s)}{2-s-\delta}, 1]$ and $\gamma \in [\frac{\delta(1-\beta)}{\beta(s-\delta)}, \frac{1}{2}]$, and negative otherwise. To summarize, we have $\frac{\partial \pi^*}{\partial \beta} > 0$ for $\beta \in [\frac{2\delta}{s+\delta}, 1]$ and $\gamma \in [\frac{\delta(1-\beta)}{\beta(s-\delta)}, \frac{1}{2}]$ and $\frac{\partial \pi^*}{\partial \beta} < 0$ otherwise.

Table E.1 Optimal Return Policies When Salvage Value Linked to Benefit of Opportunism

	Case I	Case II	Case III
	low β	high β, low γ	high β, high γ
p^*	$\frac{s^2(1-\beta(1-\gamma))^2+(1-\beta(1-2\gamma))^2-2s\beta\gamma(1-\beta+\beta\gamma)\delta+\beta^2\gamma^2\delta^2}{2(1-\beta+2\beta\gamma)^2}$	$\frac{1+s^2-2\left(2+s+s^2\right)\gamma+(5+s(2+s))\gamma^2}{2(1-2\gamma)^2}$	$\frac{2(1-\beta)+\beta^2}{2}$
f^*	$\frac{(-1+\beta+s(1+\beta(-1+\gamma))-\beta\gamma(2+\delta))^2}{2(1-\beta+2\beta\gamma)^2}$	$\frac{(1-s)^2(1-\gamma)^2}{2(1-2\gamma)^2}$	$\frac{\beta^2}{2}$
π^*	$\frac{1+s^2(1+\beta(-1+\gamma))^2+2s\beta\gamma(-1+\beta-\beta\gamma)\delta-\beta(2-\beta-2\gamma+2\beta\gamma-\beta\gamma^2\delta^2)}{2(1-\beta)(1-\beta(1-2\gamma))} - c$	$\frac{(1-\gamma)^2+s^2(1-\gamma)^2-2s\gamma^2+2\beta\gamma(1-2\gamma)\delta-c(2-4\gamma)}{2-4\gamma}$	$\frac{2(s+\beta)(1-\beta)(1-\gamma)+2(s-\beta\delta)\gamma+\beta^2-2c}{2}$

Case 1 conditions: $\beta \in [0, 1-s] \cup \gamma \in [0, \frac{1}{2}]$, or $\beta \in [1-s, \frac{1-s}{1-\delta}] \cup \gamma \in [\hat{\hat{\gamma}}, \frac{1}{2}]$, or $\beta \in [\frac{1-s}{1-\delta}, \frac{2(1-s)}{2-s-\delta}] \cup \gamma \in [\frac{(1-\beta)(s+\beta-1)}{\beta(2-s-2\beta+\delta)}, \frac{1}{2}]$

Case 2 conditions: $\beta \in [1-s, \frac{1-s}{1-\delta}] \cup \gamma \in [0, \hat{\hat{\gamma}}]$, or $\beta \in [\frac{1-s}{1-\delta}, 1] \cup \gamma \in [0, \frac{s+\beta-1}{s+2\beta-1}]$

Case 3 conditions: $\beta \in [\frac{1-s}{1-\delta}, \frac{2(1-s)}{2-s-\delta}] \cup \gamma \in [\frac{1-s-\beta}{1-s-2\beta}, \frac{(1-\beta)(s+\beta-1)}{\beta(2-s-2\beta+\delta)}]$, or $\beta \in [\frac{2(1-s)}{2-s-\delta}, 1] \cup \gamma \in [\frac{s+\beta-1}{s+2\beta-1}, \frac{1}{2}]$

where $\hat{\hat{\gamma}}=\frac{(2\beta-1)(-1+\beta-s(-2+s+2\beta)+2\beta(-2+s+2\beta)\delta-\beta\delta^2)+\beta(-1+\beta)(1+\delta)(-1+2s+\delta)-\sqrt{(-1+s+\beta-\beta\delta)^2((1-\beta+s(-1+2\beta))^2+2\beta(3-3\beta+s(-3+2\beta))\delta+\beta^2\delta^2)}}{4(-1+\beta-s(-2+s+2\beta)+2\beta(-2+s+2\beta)\delta-\beta\delta^2)\beta}$

Appendix F

More on Forecasting Model Specification

Specifications of $\widehat{R_t}$ in the two-week and one-month scenarios are presented in Table F.1. When length per period is one month, we incorporate seasonality in the benchmark forecasts of R_t and $R_{c,t}$. Further, when including the *full* set of covariates, the maximum likelihood estimation of our econometric model sometimes does not converge. We constrain the correlation parameter, θ, to zero to ensure convergence in these cases.

Table F.1 Model Specification for Two-week and One-month Cases

	Two Weeks Per Period	One Month Per Period
$\widehat{R_t^{Baseline}}$	same as Equation(3.8)	$[\frac{1}{int(\frac{t}{12})}\sum_{j=1}^{int(\frac{t}{12})}\frac{R_{t-12j}}{Q_{t-12j}}] \times Q_t$
		$int(\frac{t}{12})$ is the integer part of $\frac{t}{12}$
$\widehat{R_t^{Smoothing}}$	same as Equation(3.9)	same as Equation(3.9)
$\widehat{R_t^{ARIMA}}$	$a_0 + \alpha_1 R_{t-1} + \alpha_2 e_{t-1} + e_t$	$\alpha_0 + \alpha_1 R_{t-1} + \alpha_2 e_{t-12} + e_t$
$\widehat{R_t^{LagSales}}$	$\alpha_0 + \sum\limits_{j=1}^{9} \alpha_j Q_{t-j} + e_t$	$\alpha_0 + \sum\limits_{j=1}^{4} \alpha_j Q_{t-j} + holiday_t + e_t$
$\widehat{R_t^{ARIMAX}}$	$\alpha_0 + \sum\limits_{j=1}^{9} \alpha_j Q_{t-j} + \kappa_t$	$\alpha_0 + \sum\limits_{j=1}^{4} \alpha_j Q_{t-j} + holiday_t + \kappa_t$
	$\kappa_t = \alpha_{10}\kappa_{t-1} + \alpha_{11} e_{t-1} + e_t$	$\kappa_t = \alpha_{10}\kappa_{t-1} + e_t$
$\widehat{R_{p,t}}$	same as Equation(3.13)	same as Equation(3.13)

[1]ARIMA model in the one month case is essentially a SARIMA(1,0,0)×(0,0,1)$_{12}$ model.

[2]Triple exponential smoothing cannot converge in the one month case because there are only four cycles to start with. Therefore, we continued the use of a simple exponential smoothing model.

[3]Using a full set of month dummies in the lagged sales and ARIMAX models overfits data and produces poor forecast. A holiday dummy differentiates November and December.

Appendix G

Finite Sample Performance of the Estimators

We conduct a simulation study to evaluate the finite sample properties of the OLS, 2SLS, and FIML estimators, when the underlying econometric model is specified by the three-equation system as in (4.1a), (4.1b), and (4.4). The following Data Generating Process (DGP) is used:

$$I_i = 1(-0.6 + w_{1i} + w_{2i} + w_{3i} + u_i > 0) \tag{G.1}$$

$$y^*_{0i} = 1 + w_{1i} + w_{2i} + e_{0i} \tag{G.2}$$

$$y^*_{1i} = 2 + 2w_{1i} + w_{2i} + e_{1i} \tag{G.3}$$

where $\sigma_o = \sigma_1 = 1$, $\rho_0 = 0.2$ or 0.5, and $\rho_1 = 0.5$. The two versions of ρ_0 creates one case where $\rho_0 \neq \rho_1$ and the other case where $\rho_0 = \rho_1$. This would help us illustrate the point that 2SLS is appropriate only when $\rho_0 = \rho_1$. y_i is created by the rule that $y_i = y^*_{0i}$ if $I_i = 0$ and $y_i = y^*_{1i}$ if $I_i = 1$. Given the DGP, we know the true parameter values are $\alpha_0 = 1$, $\alpha' = (1, 1)$, $\delta_0 = 1$, and $\delta' = (1, 1)$[1].

We generate 500 simulated samples for each of the $\rho_0 = \rho_1$ and $\rho_0 \neq \rho_1$ cases. Each sample has 10000 observations. Next, we use the three estimators to recover the α and δ parameters. w_{3i} is used as an instrument for I_i when implementing the 2SLS estimator. It is, by the construction of DGP, a solid instrument – correlated with I_i but not y_i. Therefore, bias in the 2SLS estimator could only be attributed to the problem of the estimator itself but not to the validity of the instrument. OLS and 2SLS estimators are implemented with Stata's "reg" and "ivreg" commands.

[1] Recall that the δ parameters are the difference between the α and β parameters.

The FIML estimator is programed with Stata's general purpose maximum likelihood estimation language "ml" as follows.

```
program define split\_heck
args lnf xb0 xb1 xb zg lnsig0 lnsig1 athrho0 athrho1
tempvar sigma0 sigma1 rho0 rho1 u0 u1
qui gen double ‘sigma0’=exp(‘lnsig0’)
qui gen double ‘sigma1’=exp(‘lnsig1’)
qui gen double ‘rho0’=tanh(‘athrho0’)
qui gen double ‘rho1’=tanh(‘athrho1’)
qui gen double ‘u0’=\$ML\_y1-‘xb0’-‘xb’
qui gen double ‘u1’=\$ML\_y1-‘xb1’-‘xb’
qui replace ‘lnf’=ln(normalden(‘u0’,0,‘sigma0’)*normal((-‘zg’-(‘rho0’/
 ‘sigma0’)*(‘u0’))/sqrt(1-(‘rho0’)^2))) if $ML_y2==0
qui replace ‘lnf’=ln(normalden(‘u1’,0,‘sigma1’)*normal((‘zg’+(‘rho1’/
 ‘sigma1’)*(‘u1’))/sqrt(1-(‘rho1’)^2))) if $ML_y2==1
end
ml model lf split\_heck (xb0: y I = w1) (xb1: w1) (xb: w2, nocons)
(zg: w1 w2 w3) (lnsig0: ) (lnsig1: ) (athrho0: ) (athrho1: )
ml max, diff
```

Results are summarized in Table G.1. We can make three observations from this simulation study. First, OLS is always biased. Second, when $\rho_0 = \rho_1$, both 2SLS and FIML are unbiased. However, FIML is more efficient, which is evident in the smaller standard errors. Third, when $\rho_0 \neq \rho_1$, only FIML is unbiased.

Table G.1 Finite Sample Performance of OLS, 2SLS, and FIML Estimators

	OLS			2SLS			FIML		
	Mean[1]	Bias[2]	S.E.[3]	Mean	Bias	S.E.	Mean	Bias	S.E.
					$\rho_0 = \rho_1$				
$\widehat{\alpha}_0$	0.7801	0.2199	0.0130	1.0004	0.0004	0.0239	0.9999	0.0001	0.0193
$\widehat{\alpha}_1$	0.8985	0.1015	0.0133	1.0008	0.0008	0.0294	0.9995	0.0005	0.0154
$\widehat{\alpha}_2$	0.8885	0.1115	0.0110	1.0002	0.0002	0.0149	1.0003	0.0003	0.0132
$\widehat{\delta}_0$	1.5889	0.5889	0.0248	0.9999	0.0001	0.0574	0.9989	0.0011	0.0414
$\widehat{\delta}_1$	0.9729	0.0271	0.0207	0.9979	0.0021	0.0683	1.0013	0.0013	0.0221
					$\rho_0 \neq \rho_1$				
$\widehat{\alpha}_0$	0.9014	0.0986	0.0133	1.0543	0.0543	0.0240	1.0000	0.0000	0.0215
$\widehat{\alpha}_1$	0.9543	0.0457	0.0134	1.0396	0.0396	0.0303	0.9997	0.0003	0.0160
$\widehat{\alpha}_2$	0.9261	0.0739	0.0116	1.0007	0.0007	0.0151	1.0005	0.0005	0.0137
$\widehat{\delta}_0$	1.4449	0.4449	0.0251	1.0588	0.0588	0.0569	0.9986	0.0014	0.0426
$\widehat{\delta}_1$	0.9251	0.0749	0.0207	0.8970	0.1030	0.0701	1.0012	0.0012	0.0227

[1]Mean is the average across 500 samples.

[2]Bias is the difference between the true parameter value and the mean of the estimated values.

[3]S.E. is the standard error of the estimated values across 500 samples.

Appendix H

Clustering the Reputation Indices

We apply the k-means algorithm to $slr_posi_log_i$and $slr\_nega\_log_i$ with k (number of clusters) equal to 2, 3, and 4. Table H.1 summarizes the results. μ_p and μ_n are used to denote the means of $slr_posi_log_i$and $slr\_nega\_log_i$ within each cluster. As we move from two clusters to three clusters, the additional cluster seems to add enough heterogeneity. This is evident from the difference of μ_p across the three clusters. The between-cluster difference of μ_n shows similar evidence. However, as we move into the four-cluster scenario, Clusters 1-3 are quite homogeneous in their μ_n. This pattern persists if more clusters are introduced. Therefore, we proceed with the three-cluster result. The pattern of μ_p and μ_n informs us that the difference of sellers across clusters is mainly their level of activeness on eBay. That is, the more active sellers receive more positive as well as negative reviews. In fact, this observation strengthens the validity of using cluster analysis to profile sellers instead of a high-low split, since the chance of having a seller with limited positive reviews but many negative reviews is very rare.

Table H.1 Cluster Analysis of Seller Reputation

	Two Clusters			Three Clusters			Four Clusters		
	μ_p	μ_n	N	μ_p	μ_n	N	μ_p	μ_n	N
Cluster 1	2.81	0.15	2003	2.10	0.10	1172	1.67	0.08	765
Cluster 2	5.82	1.45	943	4.28	0.42	1369	3.49	0.20	1191
Cluster 3				6.92	2.45	405	5.10	0.82	704
Cluster 4							7.38	2.81	286

μ_p and μ_n are the means of slr_posi_log$_i$ and slr_nega_log$_i$ (no centering) for each cluster. N is the number of observations in each cluster.