Understanding "Cherry Pickers:" How Retail Customers Split Their Shopping Baskets

Edward J. Fox[†]

John Semple[†]

August 21, 2002

[†] Edwin L. Cox School of Business, Southern Methodist University, PO Box 750333, Dallas, TX 75275-0333.

Send correspondence to Edward J. Fox; email: efox@mail.cox.smu.edu, phone: (214)768-3943, fax: : (214)768-4099.

The authors gratefully acknowledge Information Resources, Inc. for providing the data used in this study.

Understanding “Cherry Pickers:” How Retail Customers Split Their Shopping Baskets

Abstract. Recent studies of store choice assume that shoppers develop shopping lists *a priori*, then visit the store that allows them to buy needed items at the lowest expected cost. These studies implicitly assume that shoppers purchase their entire list at a single store. However, many households shop at multiple stores without any intervening consumption—a behavior that retail managers commonly term "cherry picking." To explain this behavior, we develop a mathematical model that extends the standard cost minimization framework in two important ways: First, we allow shoppers to split their basket purchases across multiple stores; and second, we introduce price uncertainty into their decision-making process. The latter is necessary since one cannot assume all prices are known when the purchase decision is made. Because deferring purchases to a subsequent store visit is essentially a gamble constructed by the shopper, the buy or defer purchase decision will depend on her risk attitude. Despite the shopper-specific nature of this attitude, we discover a rather general result: Unless the shopper is perfectly risk neutral, the probability of an item being “cherry-picked” depends on its expected price difference divided by the variance of that difference. This leads to interesting implications for retailers regarding the timing and depth of retail promotions. Our model also predicts that the amount a cherry-picking shopper purchases at a given retailer depends on the order in which the retailers are visited. We develop a discrete mixture model that incorporates various aspects of household-level heterogeneity, including differences in loyalty, risk attitudes, and price knowledge. We apply this model in the analysis of a panel dataset of consumer purchases made by cherry-picking shoppers and find uniform support for our model’s predictions. In addition, we find empirical evidence that most cherry pickers are risk-averse in their purchase decisions, and that they study retailer ads.

Key Words: Retailing, Cherry Picking, Store Choice, Shopping Behavior, Structural Mixture, Constrained Maximum Likelihood, Risk Aversion.

Understanding "Cherry Pickers:" How Retail Customers Split Their Shopping Baskets

1. Introduction

The lack of customer loyalty to a single retailer can be explained by different underlying patterns of shopping behavior. One such behavior is known as "cherry picking." As American Demographics, Inc.'s Marketing Tools explains:

> For many Americans, the … hunt for value has replaced loyalty as a deciding factor in planning which stores to patronize – and how many. Instead of going to the same outlet each week, every week, to complete their grocery shopping, price-conscious consumers often visit more than one store in search of special prices – a bargain-hunting practice known in the industry as "cherry picking." (Mogelonsky 1994, p. 10)

Simply put, cherry pickers shop around for bargains (Mogelonsky 1995, p. 27). *Consumer Reports* actually recommends that its readers "scrutinize the food-day ads and 'cherry pick' the specials," noting that 20% of its readers show little loyalty among supermarkets (Consumer Reports 1988, p. 158). Thus, the cherry picker visits multiple retailers when shopping for a basket of goods, taking advantage of the lowest item prices *across those retailers*. Cherry pickers are not the only price-conscious shoppers; others are loyal to the single retailer that they believe offers the lowest prices (Mogelonsky 1995 p. 27). Note that cherry picking is fundamentally different from *store switching*, where the shopper chooses the particular retailer offering the lowest expected total cost to acquire her market basket on a particular shopping trip (Barnard and Hensher 1992, Bell, Ho, and Tang 1998, Fox, Metters, and Semple 2002). While the store switcher may patronize different retailers like the cherry picker, she does not split her basket among multiple retailers on the same trip.

Surprisingly little research has been conducted on consumer cherry picking, though previous research has *assumed* its existence. Lal and Rao (1997) develop a theoretical model that segments shoppers into those who are time constrained and those who cherry pick. Dreze (1999) assumes a segment of shoppers that is price sensitive with low travel costs that can be induced to cherry pick by retailer price deals. Both models suggest that cherry pickers are willing to travel to multiple retailers to take advantage of price deals because of the low opportunity cost of their time.

1.1. Literature Review. Our characterization of cherry-picking behavior builds on recent store choice research (Barnard and Hensher 1992, Bell, Ho, and Tang 1998, Fox, Metters, and Semple 2002). These studies model the shopper's choice of store as one that minimizes the sum of (1) the fixed cost of shopping (e.g., traveling to and from the store) and (2) the variable cost of shopping (product quantities and product prices for the basket). Although this framework places the desired emphasis on costs, it does not apply directly to cherry pickers because the shopper is required to buy her entire basket at a single retailer. We relax this requirement, allowing the shopper to partition her basket and purchase the resulting "sublists" at different retailers (for clarity and tractability, we consider only two retailers). More precisely, the shopper visits one retailer, and, conditioned on the observed prices, may defer some of her purchases to the second retailer—if the savings she expects to realize exceed the cost of the incremental visit. This partitioning of the basket is based on both external and internal price information (Mayhew and Winer 1992), as the cherry picker compares *observed* prices at the first retailer with generally *unobserved* prices at the second retailer. In some cases, item prices at the second retailer are known because of information acquired prior to the shopping trip (e.g., feature advertising). More often, however, the cherry picker must use internal information (e.g., recently observed prices, probability of discounts) to form price expectations for the second retailer, thereby introducing uncertainty into her decisions about where to purchase items in the basket. Thus, our models also extend the store choice literature by explicitly incorporating price uncertainty. As a result, cherry-picking decisions depend on the shopper's response to this uncertainty, or risk attitude (Kahneman and Tversky 1979).

A broader literature review reveals two studies about consumer price search and multi-store shopping among grocery retailers suggesting that the cherry-picking segment is substantial (Urbany, Dickson, and Key 1991, Urbany, Dickson, and Sawyer 2000). Among surveyed consumers:

- 22%-24% of consumers regularly shop at multiple stores (UDK, UDS)
- 25% shop other than their principle store to get advertised specials (UDK)
- 42% of shoppers compare prices across stores at least once per month (UDS)
- 19% regularly shop specials at multiple stores (UDS)
- 57%-80% of shoppers read fliers to compare prices across stores (UDK, UDS)

Other related research takes the retailer’s viewpoint, investigating the effects of retailer promotions on store traffic (Walters and Rinne 1986, and Bucklin and Lattin 1992) and the shifting of sales between stores (Walters and MacKenzie 1988, Kumar and Leone 1988, Walters 1991, Grover and Srinivasan 1992). These papers find either weak or nonexistent traffic effects and limited displacement of sales from one store to another as a consequence of advertised price promotions. In sum, these studies suggest that the linkages between retailer promotions and multi-store shopping are not fully understood.

The conventional wisdom is that cherry pickers are the least profitable retail customers because they buy fewer items at each store, and only those items that are discounted (see Dreze 1999 for a counter-argument). In fact, Wal-Mart’s recently declining margins and the associated failure of earnings to match revenue growth have been attributed in part to cherry picking consumers (MMR 2002).

1.2. Overview and Contribution. In this paper, we propose an analytical and empirical framework that relates retailer prices to individual cherry picking behavior. Essential to our proposition is that the cherry picker effectively constructs a gamble by deciding which items to buy at the first retailer and which to defer to the second. An interesting and important consequence of this design is that cherry-pickers are driven by the distribution of price *differences* (i.e., savings) for items across retailers instead of the distribution of prices for items within a retailer. We prove that a risk-averse cherry picker's propensity to defer purchase of an item to the second retailer increases according to the ratio $E(d_i)/(q_i \cdot Var(d_i))$, where $E(d_i)$ is the expected price difference for item *i* between retailers, $Var(d_i)$ is the variance of the price difference, and q_i is the quantity to be purchased[1]. Our empirical analysis supports this finding, and further suggests that most cherry pickers exhibit risk aversion. The analytical and empirical findings together imply that retailers could increase sales among cherry pickers by: (1) matching the timing and depth of competitor promotions for items which they generally price below competition (i.e., reducing $Var(d_i)$), but (2) timing promotions of items priced above competition so as to avoid competitor promotions (i.e., increasing $Var(d_i)$).

Our paper makes several contributions. We are the first to propose an analytical model of cherry picking that distinguishes this behavior from store switching by focusing on the statistical properties of item-level price differences between stores. In doing so, we extend the typology of non-loyal shopping behaviors and provide an appropriate analytical framework for investigating a widespread consumer shopping strategy. With minimal model assumptions, we show that simple price difference ratios at the item level are theoretically related to the intensity of cherry picking. Building on the importance of uncertainty implied by these ratios, we develop a finite

[1] More precisely, it increases according to a piecewise linear sigmoidal function of this ratio.

mixture model that allows shoppers to internalize this uncertainty in different ways. Included are factors that could affect savings uncertainty, such as retailer advertising and price promotions, order of retailer visits, and previous price history. This analysis allows us to offer a new perspective on the connection between retailer advertising and multi-store shopping behavior.

The remainder of the paper is organized in four sections. The next section introduces and develops a mathematical framework for analyzing cherry-picking behavior, resulting in three testable propositions. Those propositions are tested in the following section, in which we specify and estimate an econometric model of consumer cherry picking. The subsequent section presents a discussion of our findings and their implications for retail managers. We conclude with a final section that discusses the model's limitations and suggests topics for future research.

2. Expected Savings and the Probability of a Better Deal

Before embarking on trip *t*, the customer compiles a *shopping list* $L = \{1,2,....,I\}$ of items i = 1,2,….,*I*, along with purchase quantities $q_{1t}, q_{2t},...., q_{It}$ for each item. The list may be written down in advance of the shopping trip (see Kahn and McAlister 1997, pp.118-9, for a review) or may be kept mentally. In our model, the quantities are treated as exogenous and fixed. There are two retailers from whom the items can be acquired, "Retailer A" and "Retailer B." This framework is similar to that of Bell, Ho, and Tang (1998), except that we do not require that the whole list be acquired at one retailer. Moreover, the order that Retailers A and B are visited is important. On trip *t*, the first retailer visited is henceforth denoted as Retailer 1 and has a vector of stochastic prices $p_t^{(1)} = (p_{1t}^{(1)}, p_{2t}^{(1)},..., p_{It}^{(1)})^T$. The second retailer visited, denoted by Retailer 2, has stochastic prices $p_t^{(2)} = (p_{1t}^{(2)}, p_{2t}^{(2)},..., p_{It}^{(2)})^T$. The shopper's objective is to acquire her list at low cost in the presence of price uncertainty.

The decision regarding where to buy the items on the list is based on the buyer's knowledge of price differences for items at the two retailers.[2] In the event that the shopper defers some—but not all—of her purchases to the second retailer, we say that she is *cherry-picking*, and a fixed cost of k is incurred for the additional visit. For clarity of exposition, we say that this additional visit takes place during the same shopping trip as the first visit, though the cost k could as easily apply to a separate trip. Since all prices at Retailer 2 are not known with certainty, the buyer's gamble may result in a gain (net savings) or loss once prices are realized and the set-up cost is included. Rationalizing the visit to Retailer 2 in probabilistic terms is an essential part of our decision model.

To describe the shopper's situation in more detail, we introduce a decision vector $\lambda_t^T = (\lambda_{1t}, \lambda_{2t}, ..., \lambda_{It})$, called the *sublist*, whose component λ_{it} is the proportion of item i purchased at Retailer 2 on trip t. Observe that we do not require the vector λ_t to be binary, a relaxation used to facilitate our analytic arguments and to permit hedging (e.g., one might buy one six-pack of beer at Retailer 1, another at Retailer 2). One could also interpret the value λ_{it} as measuring the cherry picker's preference for purchasing item i on trip t at Retailer 2. The values of the sublist and their interpretations will be discussed in greater detail once our model is formally proposed. The shopper considers two criteria when constructing her sublist: (1) the expected savings from purchasing the sublist at Retailer 2, and (2) the probability of a net gain, i.e., the probability that the trip to Retailer 2 produces a gain in excess of the setup cost k. The first criterion is needed to measure the magnitude of potential gains from purchasing a sublist at Retailer 2. However, by itself it fails to allow for known risk-averse behavior, i.e., the over-

[2] Strictly speaking, we assume the buyer knows either the distribution of price differences for each item, or the conditional distribution of price differences given the observed price at Retailer 1.

weighting of high probability outcomes in the domain of gains (see Kahneman and Tversky, 1979). This motivates the inclusion of the second criterion into the shopper's objective function—the *probability* of a net gain when purchasing λ_t at Retailer 2. The cherry picker thus constructs a gamble reflecting her personal tradeoff between risk and return. The difference between placing (i) a positive weight on the probability of a net gain when deciding where to buy items on the shopping list and (ii) zero weight leads to fundamentally different types of cherry pickers. The following table summarizes our notation.

Summary of Notation

Symbol	*Meaning*
$p_t^{(1)} = (p_{1t}^{(1)}, p_{2t}^{(1)}, ..., p_{It}^{(1)})^T$	Vector of stochastic prices for Retailer 1 on trip *t*
$p_t^{(2)} = (p_{1t}^{(2)}, p_{2t}^{(2)}, ..., p_{It}^{(2)})^T$	Vector of stochastic prices for Retailer 2 on trip *t*
$\lambda_t^T = (\lambda_{1t}, \lambda_{2t}, ..., \lambda_{It})$	Proportion of goods purchased at Retailer 2 on trip *t*
$Q_t = diag(q_{1t}, q_{2t}, ..., q_{It})$	Diagonal matrix of required quantities on trip *t*
q_{it}	Required quantity of item *i* in trip *t*
K	Additional fixed cost of going to Retailer 2

2.1. Risk-Neutral Cherry Picker. The risk-neutral cherry picker does not place any weight on the probability of a net gain for an individual trip, instead focusing solely on the expected gain. Her objective is to split the shopping list in a manner that minimizes expected total cost. Cherry pickers by definition shop at both retailers, and the optimality of this decision requires that

$$\sum_{i=1}^{I} q_{it} \max\{E(d_{it}), 0\} - k > 0,$$

where $d_{it} = p_{it}^{(1)} - p_{it}^{(2)}$ is the price difference between retailers for item *i* on trip *t*. Note that this maximization implicitly assigns $\lambda_{it} = 1$ for each product where $E(d_{it}) > 0$, $\lambda_{it} = 0$ otherwise. The difficult part, of course, is to specify a suitable model of the expected price differences on individual trips. We take up this problem in detail in §3.1.

Proposition 1 – The risk-neutral cherry picker's preference for deferring purchase of an item to the second retailer visited increases with the expected savings at that retailer, multiplied by the quantity.

2.2. Risk-Averse Cherry Picker. The risk-averse cherry picker places a positive weight on both the expected gain in splitting her basket and the probability of realizing such a gain. These two objectives are combined into a single objective function with the use of decision weight functions *g* and *h*. The particular choice of *g* and *h* reflects the consumer's individual risk attitude; we merely assume that they are nonnegative and strictly monotone increasing. This ensures that, all things being equal, the buyer would prefer a gamble with a higher expected return (among gambles with equal probabilities of success) and a gamble with a higher probability of success (among gambles with equal expected returns). The general mathematical program we consider is

$$\underset{\lambda_t}{Max}\ g\left(E\left\{\lambda_t^T d_t\right\}-k\right)+h\left(\text{Prob}\left\{\lambda_t^T d_t > k\right\}\right)$$

where $d_t = p_t^{(1)} - p_t^{(2)}$ is the vector of price differences for trip *t*, $E\left\{\lambda_t^T d_t\right\}-k$ is the total expected savings from deferring the purchase of λ_t to Retailer 2, and $\text{Prob}\left\{\lambda_t^T d_t > k\right\}$ is the probability of realizing a net gain by deferring the purchase of λ_t to Retailer 2.

The individual criteria can now be considered in more detail. The probability of realizing a net gain by purchasing a sublist λ_t at Retailer 2 on trip *t* can be stated mathematically as

(1) $$P\left(\lambda_t^T Q_t p_t^{(1)} - \lambda_t^T Q_t p_t^{(2)} > k\right),$$

where Q_t is the diagonal matrix of quantities. Observe that an overall better deal is only achieved if the sublist λ_t purchased at Retailer 2 more than offsets the additional cost of the trip (*k* dollars).

To analyze the probability of a better deal in greater detail, we assume that the vector of price differences on trip t, $d_t = p_t^{(1)} - p_t^{(2)}$, can be modeled by a vector equation of the general form

$$d_t = \delta_t(d_{t-1}, d_{t-2}, \ldots, d_{t-l}, w_t) + \varepsilon_t.$$ [3]

Here, $d_{t-1}, \cdots, d_{t-l}$ are the vectors of price differences over the preceding l trips, w_t is a vector of additional (known) information that affects expected savings on trip t (e.g., advertised specials, known prices, etc.), and $\varepsilon_t \sim N(0, \Sigma_t)$. The structure of the price difference equation implies

$$E(d_t) = \delta_t(d_{t-1}, d_{t-2}, \ldots, d_{t-l}, w_t),$$

thus $\delta_t(d_{t-1}, d_{t-2}, \ldots, d_{t-l}, w_t)$ is the vector of expected price differences. Although the normality assumption is used for model tractability, our primary motivation is to gain some general qualitative insights regarding the shopping behavior of risk-averse cherry pickers. It is important to emphasize that we are not assuming the marginal distributions of individual prices at each retailer are normally distributed, an assumption that would be overly optimistic in virtually all settings. We assume the shopper has "learned" about price differences empirically through shopping and can therefore approximate the expected value and variance of the various price differences on each trip. Empirical work presented in §3 suggests that a relatively simple form of the equation above will effectively approximate the price difference process.

Equation (1) is equivalent to the normalized version

(2) $$\Pr\left(\frac{\lambda_t^T Q_t d_t - \left[\lambda_t^T Q_t \delta_t\right]}{\sqrt{\lambda_t^T Q_t \Sigma_t Q_t \lambda_t}} > \frac{k - \left[\lambda_t^T Q_t \delta_t\right]}{\sqrt{\lambda_t^T Q_t \Sigma_t Q_t \lambda_t}} \right).$$

[3] If we assume the price at the first retailer is observed, then this can be rearranged to represent the conditional distribution of prices at Retailer 2 given prices at Retailer 1.

Here, δ_t is shorthand for the expected price difference vector. The probability of savings in equation (2) reduces to

(3) $$\Pr\left(z > \frac{k - \left[\lambda_t^T Q_t \delta_t\right]}{\sqrt{\lambda_t^T Q_t \Sigma_t Q_t \lambda_t}} \right)$$

where $z \sim N(0,1)$. In this case, the probability of achieving savings by purchasing λ_t at Retailer 2 is

(4) $$1 - \Phi\left(\frac{k - \left[\lambda_t^T Q_t \delta_t\right]}{\sqrt{\lambda_t^T Q_t \Sigma_t Q_t \lambda_t}} \right),$$

where $\Phi(\bullet)$ is the c.d.f. of the standard normal. Making $1 - \Phi$ large is achieved by making Φ's argument small. If we assume that the components of the errors ε_t are serially uncorrelated and uncorrelated across items, then we have the simpler form $\Sigma_t = diag(\sigma_{1t}^2, \sigma_{2t}^2, \ldots, \sigma_{It}^2)$. Observe that this does not preclude the possibility of item price correlation across retailers, as one might expect to arise through normal price competition.

At this point, we observe that the maximization posed at the outset can be formally stated as

(5) $$\underset{\lambda_t \in \Lambda}{Max} \quad g\left(\lambda_t^T Q_t \delta_t - k\right) + h\left[1 - \Phi\left(\frac{k - \left[\lambda_t^T Q_t \delta_t\right]}{\sqrt{\lambda_t^T Q_t \Sigma_t Q_t \lambda_t}} \right) \right],$$

where $\Lambda = \{\lambda : 0 \le \lambda_i \le 1,\ i = 1,2,\ldots,I\}$ is the feasible set of lambdas. The problem (5) has no closed-form solution, yet, as we shall describe next, it is still possible to determine important properties of its structure. To do so, we first define the index set $I_t^+ = \{i : \delta_{it} > 0\}$, where the i^{th} component of δ_t is δ_{it}. This set represents those items that the cherry picker expects to obtain at lower cost at Retailer 2 on trip *t*. Consequently, for items $i \in I_t^+$, δ_{it} represents the expected *savings*. Additionally, we must have $\sum_{i \in I_t^+} q_{it} \cdot \delta_{it} > k$ to ensure the constructed gamble has positive expected value. Assuming this inequality is satisfied, the optimal solution vector λ_t^* has positive elements limited to I_t^+. If this were not the case, then one could construct a better solution by setting $\lambda_{it}^* = 0, i \notin I_t^+$ while leaving all elements λ_{it}^* $i \in I_t^+$ alone. For the elements of I_t^+ we have the following result, which describes how the optimal solution is related to the mean and variance of the price difference, as well as the quantity q_{it} being purchased.

Theorem: Suppose $\sum_{i \in I_t^+} q_{it} \cdot \delta_{it} > k$. Then the optimal solution to (5) satisfies

1. $\lambda_{it}^* > 0$ for $i \in I_t^+$

2. For any $i, j \in I_t^+$ with $\lambda_{it}^* < 1, \lambda_{jt}^* < 1$, $\dfrac{\lambda_{it}^*}{\lambda_{jt}^*} = \dfrac{\delta_{it}}{\delta_{jt}} \cdot \dfrac{\sigma_{jt}^2}{\sigma_{it}^2} \cdot \dfrac{q_{jt}}{q_{it}} = \dfrac{\delta_{it} / q_{it}\sigma_{it}^2}{\delta_{jt} / q_{jt}\sigma_{jt}^2}$.

3. If $\lambda_{it}^* > \lambda_{jt}^*$, then $\dfrac{\delta_{it}}{q_{it}\sigma_{it}^2} > \dfrac{\delta_{jt}}{q_{jt}\sigma_{jt}^2}$.

Proof: See Appendix.

Part 1 of the theorem tells us that every item whose expected price is lower at Retailer 2 will be purchased in some positive proportion at Retailer 2. Part 2 of the theorem defines the relationship between two products whose purchases are both split across stores. Part 3 of the theorem addresses a more general relationship that must hold if the customer prefers purchasing proportionately more of item *i* than item *j* at Retailer 2. The necessary condition is that the ratio of the expected savings ($q_{it}\delta_{it}$) to the variance ($q_{it}^2\sigma_{it}^2$) for the purchase of q_{it} units of item *i* must exceed the same ratio for item *j*. This is an intuitively appealing condition since it incorporates both expected savings and price uncertainty in a simple way that is possible to test with empirical data. Moreover, it confirms the simple intuition that risk-averse cherry pickers defer those purchases to the second retailer that have the greatest certainty of producing savings. We observe that this ratio is similar to the Sharpe ratio used in financial portfolio theory except that our denominator uses the variance instead of the standard deviation and includes a quantity scale factor.

Proposition 2 – For risk-averse cherry pickers, the preference for deferring purchase of an item to the second retailer visited increases with the expected savings, divided by the product of the quantity and the variance of the savings.

A further consequence of the theorem is the particular structure of the *optimal proportion function*, which is the proportion of an item that should be purchased at the second retailer. We can think of this function as measuring the cherry-picker's propensity (or preference) for buying an item at the second retailer. For simplicity, assume that all items satisfy $q_{it} = 1$. By remarks made prior to the theorem, the cherry picker will not defer any part of an item-level purchase unless $\delta_{it} > 0$. Thus the optimal proportion for an item is 0 if $\delta_{it} \leq 0$. By Part 3 of the theorem, if a risk-averse cherry picker chooses to buy all of item *j* at Retailer 2, then any item *i* whose ratio satisfies

$$\frac{\delta_{it}}{\sigma_{it}^2} \geq \frac{\delta_{jt}}{\sigma_{jt}^2}$$

must also be purchased in its entirety at Retailer 2. Thus, there exists a critical ratio such that all items with higher ratios are purchased in their entirety at Retailer 2. By Part 1 and Part 2 of the theorem, the proportion purchased at the second retailer increases linearly in the ratio $\delta_{it} / \sigma_{it}^2$

prior to attaining its maximum of 1. This means that the risk-averse cherry picker's optimal proportion function is a piecewise linear sigmoidal function of the argument $\delta_{it} / \sigma_{it}^2$. A graph of this function is shown in Figure 1 below.

Price Promotions. In some cases, the price difference for an item is known with certainty due to the retailer's feature advertising. In this case, the variance of the price difference is 0. In such cases, either $\lambda_{it} = 1$ (when the item is cheaper at Retailer 2), or $\lambda_{it} = 0$ (when the product is cheaper at Retailer 1).

Routes. An important exogenous factor in our analysis is the order in which retailers are visited. For the two-store situation, there are two possible routes of return, each beginning with a different store. We do not assume the buyer begins at home, since groceries are often purchased in conjunction with other shopping activities (e.g., Dellaert, et al. 1998). However, we do assume that the home is the destination after the final grocery store, as we would anticipate the presence of at least some perishable purchases requiring prompt refrigeration. The two routes are represented in Figure 2 by solid and dashed arrows, respectively. For the bold route in Figure 2, Retailer A occupies the role of Retailer 1; for the dotted route, Retailer B occupies the role of Retailer 1. Observe that the setup costs for visiting the second retailer are not necessarily symmetric. In Figure 2, the additional setup cost of going from Retailer A to Retailer B (the route shown in bold) is greater than the additional setup cost of going from Retailer B to Retailer A (the dotted route). Retailer A is nearly "on the way home" after visiting Retailer B whereas Retailer B is "out of the way" after visiting Retailer A.

Applying equation (5) to each route results in different purchasing behaviors. This is most striking for items where there is no significant price advantage at either retailer. If an item satisfies $\delta_{it} \approx 0$ and its price is not known with certainty at the second retailer, then it should be purchased at the retailer visited first. This is because deferring its purchase to the second retailer increases the variance of the sublist and hence decreases the certainty of a net gain. For items $\delta_{it} \neq 0$, the asymmetry in additional setup costs between retailers caused by the route could affect the purchases deferred to the second retailer, λ_t. In extreme cases, cherry picking could even occur in one of the directed routes but not the other.

Proposition 3 – Compared to risk-neutral cherry pickers, risk-averse cherry pickers are more likely to purchase items with little or no expected price difference at the first retailer visited.

3. Econometric Model and Empirical Testing

We now turn to the empirical testing of our mathematical results. First, we develop an econometric model motivated by our analytical findings; then, we estimate it using panel data. A notable feature of our econometric model, shared with the analysis in §2, is that shoppers are allowed to exhibit risk neutrality or risk aversion.

The foregoing analysis allows for varying proportions of each product to be purchased at the two retailers visited. While this makes the analytical model tractable, it does not capture the discreteness of product purchases. Consider that, of the SKUs (stock keeping units) bought by cherry pickers on a given day in our dataset (see § 3.b. for details concerning the data), a single unit is purchased more than 97% of the time. In light of the discrete nature of product purchases, we specify a choice model based on random utility. Moreover, the high frequency with which a single unit is purchased allows us to safely ignore stockpiling, or purchase quantity effects, in our empirical analysis.

3.1. Econometric Model. In our econometric model, the shopper is assumed to have made the decision to cherry pick, and so may purchase items from her shopping list at any retailer that she visits. To simplify the analysis, we focus on cherry picking between the two largest grocery retailers in a duopolistic market. Thus, the shopper visits both retailers on each cherry-picking trip.[4] We model the probability that household *h* (*h* = 1, …, *H*) chooses to buy item *i* (*i* = 1, …, *I*) from her shopping list on trip *t* (*t* = 1, …, *T*) at Retailer 1, the first retailer visited on that trip:

$$\pi_{hit}^{(1)} = \frac{e^{U_{hit}^{(1)} + \varepsilon_{hit}}}{1 + e^{U_{hit}^{(1)} + \varepsilon_{hit}}}. \tag{6}$$

Here, $U_{hit}^{(1)}$ is the deterministic component of utility and ε_{hit} is a random error term. Note that, to be consistent with our analytical framework, the choice of where to purchase does not relate to a given retailer, but to the first retailer visited on the shopping trip. We assume that the errors are generalized extreme value distributed, so the resulting model is a binary logit.

To test our analytical framework and findings, we allow the deterministic component of utility to depend upon the shopper's risk attitude. As we assert in Proposition 1, the risk-neutral cherry picker's preference to defer purchase to Retailer 2 increases in the expected savings multiplied by the purchase quantity, $q_{hit}E\left(p_{it}^{(1)} - p_{it}^{(2)}\right)$. Note that the savings will be signed negative if the Retailer 1 is priced below Retailer 2. In contrast, Proposition 2 asserts that the risk-averse cherry picker's preference to defer purchase to Retailer 2 increases in the expected savings, divided by the variance of that savings and the purchase quantity, $\frac{E\left(p_{it}^{(1)} - p_{it}^{(2)}\right)}{q_{hit}Var\left(p_{it}^{(1)} - p_{it}^{(2)}\right)}$. Thus, the choices of risk-neutral and risk-averse customers are modeled as fundamentally different processes. Following Kamakura, Kim, and Lee (1996), we allow for distinct processes

[4] Again, for clarity we assume that both retailers are visited on a single shopping trip. Yet our results are not dependent upon this assumption; they hold even if the shopper returns home between visits.

using a finite mixture model that captures structural, as well as preference, heterogeneity. Specifically, we assume N risk-neutral segments and V risk-averse segments. Each segment is comprised of consumers who use the same choice process and have relatively similar preferences. The size s of each segment (which must be estimated) reflects the unconditional probability that any consumer belongs to that segment. The segments are of sizes s_n (n = 1, …, N) for the N risk-neutral segments, s_v (v = 1, …, V) for the V risk-averse segments, and $\sum_{n=1}^{N} s_n + \sum_{v=1}^{V} s_v = 1$.

The utilities of both risk-neutral and risk-averse cherry pickers also depend on the household's category-specific loyalty to Retailer 1. Thus, we specify the deterministic component of utility as:

$$U_{hit}^{(1)} = \beta_0^v + \beta_1^v \ell_{hct} + \beta_2^v r_{ht} + \beta_3^v \left(\frac{E\left(p_{it}^{(1)} - p_{it}^{(2)}\right)}{q_{hit} Var\left(p_{it}^{(1)} - p_{it}^{(2)}\right)} \right) \quad \forall \;\; h \in v, i \in c \tag{7}$$

for risk-averse cherry pickers, and

$$U_{hit}^{(1)} = \beta_0^n + \beta_1^n \ell_{hct} + \beta_2^n r_{ht} + \beta_3^n \left(q_{hit} E\left(p_{it}^{(1)} - p_{it}^{(2)}\right)\right) \quad \forall \;\; h \in n, i \in c \tag{8}$$

for risk-neutral cherry pickers; where ℓ_{hct} is household h's share-of-requirements loyalty (i.e., proportion of the household's total category purchases) to the first retailer visited for category c, and r_{ht} is an indicator variable for the *specific* retail chain visited first, which takes the value 1 if Retailer A is visited before Retailer B, 0 otherwise. Thus, r_{ht} captures an intrinsic preference for Retailer A. The category-specific loyalty variable captures non-price reasons that a household may prefer to purchase goods in a category at a particular retailer (Bell, Ho, and Tang 1998), such as a unique assortment. Category loyalty is indexed by trip because it reflects the proportion of all category purchases made by household h at Retailer 1 prior to trip t.[5] Though we are unable to test Proposition 3 *per se*, the intercept parameter, reflecting preference to purchase at the first retailer visited, should be higher for risk-averse cherry pickers than for risk-neutral cherry pickers, *ceteris paribus*.

The likelihood of the observed choice history of household h, conditional on membership in risk-neutral segment n is:

(9) $$L_{hn} = \prod_t \prod_i \left(\pi_{hit}^{(1)} \mid h \in n\right)^{y_{hit}}$$

where y_{hit} is an indicator variable which takes the value 1 if item i from household h's shopping list for trip t is purchased at Retailer 1 (the first retailer visited); 0 otherwise. Similarly, the conditional likelihood for risk-averse segment v is:

(10) $$L_{hv} = \prod_t \prod_i \left(\pi_{hit}^{(1)} \mid h \in v\right)^{y_{hit}}$$

The unconditional likelihood of the household's purchase history is an average of the segment-level conditional likelihoods, weighted by segment sizes:

(11) $$L_h = \sum_n s_n L_{hn} + \sum_v s_v L_{hv}$$

Parameter estimates are obtained by maximizing the log likelihood over all households h:

(12) $$LL = \sum_h L_h$$

---

[5] Category loyalty is computed as:

$$\ell_{hct} = \frac{\sum_{i \in c} \sum_{w=1}^{t-1} y_{hiw}}{\sum_{i \in c} \sum_{w=1}^{t-1} y_{hiw} + \sum_{i \in c} \sum_{w=1}^{t-1} \left(1 - y_{hiw}\right)}$$

where y_{hiw} is an indicator variable which takes the value 1 if item i from household h's shopping list for trip w is purchased at Retailer 1; 0 otherwise. If no prior category purchases are observed, the value of 0.5 is imputed for ℓ_{hct}, reflecting the opportunity to purchase at either retailer because both are visited on trip t.

Unfortunately, maximizing the log-likelihood using nonlinear numerical optimization (or any other technique) is subject to convergence to local maxima, rather than the global maximum. To minimize this problem, we systematically vary the starting values over a wide range of possible configurations. These configurations depend on the number of segments, as well as the proportion that are risk-neutral and risk averse.

The probability that the individual household's observed behavior is attributable to risk-neutral segment *n* is:

(13) $$\rho_{hn} = s_n L_{hn} / \left(\sum_{n'=1}^{N} s_{n'} L_{hn'} + \sum_{v=1}^{V} s_v L_{hv} \right)$$

while the probability of membership in risk-averse segment *v* is:

(14) $$\rho_{hv} = s_v L_{hv} / \left(\sum_{v'=1}^{V} s_{v'} L_{hv'} + \sum_{n=1}^{N} s_n L_{hn} \right)$$

Expected Price Savings. We now develop a specification for expected price savings. Propositions 2 and 3 are driven by uncertainty about prices at the second retailer to be visited, as the shopper is deciding which items to purchase at the first retailer and which to defer. To reduce this uncertainty, it is commonly believed that cherry-picking shoppers study retailer advertising to learn price information about feature-advertised items before shopping. We use an indicator variable, f_{it}, to capture Retailer 2's (the second retailer's) decision to feature advertise item *i* during trip *t*. Because we do not observe whether the shopper in fact studies retailer ads, we estimate this propensity with a parameter α ($\alpha \in [0,1]$). If $\alpha = 0$, then the cherry picker never studies retailer ads. If $\alpha = 1$, then the cherry picker always studies the ads of the retailers that she visits. Note that α need not always pertain to the same retailer—it applies to the second retailer on a given trip. While knowledge of retailer ads may also play a role in the shopping sequence, we can discern only whether or not advertised prices are known *a priori* for the second retailer visited. The propensity to study retailer ads, together with retailers' advertising decisions, result in the three possible alternative expectations of price savings shown below.

Case	Study Ads?	Feature Advertised?	Expectation of Price Savings
1	Yes ($\alpha = 1$)	Yes ($f_{it} = 1$)	$p_{it}^{(1)} - p_{it}^{(2)}$
2	Yes ($\alpha = 1$)	No ($f_{it} = 0$)	$p_{it}^{(1)} - p_{i,t-1}^{(2)+}$
3	No ($\alpha = 0$)	Yes ($f_{it} = 1$) or No ($f_{it} = 0$)	$p_{it}^{(1)} - \left(p_{i,t-1}^{(2)+}(1-\varphi_{it}) + p_{i,t-1}^{(2)-}\varphi_{it}\right)$

where $p_{i,t-1}^{(2)-}$ is the most recent feature-advertised price for item i at Retailer 2 prior to trip t, $p_{i,t-1}^{(2)+}$ is the most recent non-advertised price for item i at Retailer 2 prior to trip t, and φ_{it} is the probability that item i will be feature-advertised at Retailer 2 during trip t.[6] The price superscripts reflect that non-advertised prices ($^{+}$) are higher than advertised prices ($^{-}$).

In Case 1, the cherry picker studies the retailer ads, and the item is featured at the second retailer visited on the trip. Because she also knows the item price at the first retailer, the expected price savings is the true price savings. In Case 2, the shopper studies the ads, but the item is not featured at the second retailer. Knowing that the item is not featured leads the cherry picker to expect the price at the second retailer to be $p_{i,t-1}^{(2)+}$, the most recent non-advertised price, and the expected savings follows. In Case 3, the shopper does not study ads, and so does not know whether the item is feature advertised at the second retailer. Absent *a priori* knowledge of prices on trip t, the expected price at Retailer 2 is an average of the most recent advertised and non-advertised prices, $p_{i,t-1}^{(2)-}$ and $p_{i,t-1}^{(2)+}$, weighted by the probability that the item will be

[6] The probability that the item will be featured on trip t is operationalized as the proportion of the previous trips during which the item was advertised. Formally,

$$\varphi_{it} = \frac{1}{t-1}\sum_{z=1}^{t-1} f_{iz}$$

For consistency in the development of this variable, we simply average f_{it} over all weeks previous to the week of trip t.

advertised at Retailer 2, φ_{it}. Operationalizing the shopper's expectation of unknown prices using the most recent advertised and non-advertised prices offers three important advantages compared to alternatives that incorporate older price information. First, it does not assume that shoppers have detailed historical information about item prices. Second, it gives no weight to older prices, which may not reflect the retailer's latest product costs and pricing strategy. Third, our approach predicts non-advertised discounts, which typically persist for four weeks or more, better than a historical average of prices would. All cases detailed above are incorporated into the following equation for expected price savings:

$$(15)\quad E\left(p_{it}^{(1)} - p_{it}^{(2)}\right) = \alpha\left(f_{it}\left(p_{it}^{(1)} - p_{it}^{(2)}\right) + \left(1 - f_{it}\right)\left(p_{it}^{(1)} - p_{i,t-1}^{(2)+}\right)\right) + \left(1-\alpha\right)\left(p_{it}^{(1)} - \left(p_{i,t-1}^{(2)-}\varphi_{it}\right) - p_{i,t-1}^{(2)+}\left(1-\varphi_{it}\right)\right)$$

where the segment superscript of the α parameter is suppressed. While we consider equation (15) to be an intuitively appealing model of expected price differences, one might propose other plausible models to operationalize the theoretical framework of §2.

Variance of Price Savings. Equation (7) specifies the utility of risk-averse shoppers to depend on the expected price savings *divided by the variance of that savings* and purchase quantity. Recall from §2 that the variance term, $Var\left(p_{it}^{(1)} - p_{it}^{(2)}\right)$, reflects the uncertainty associated with the expected savings. We complete our specification by approximating the variance:

$$(16)\quad Var\left(p_{it}^{(1)} - p_{it}^{(2)}\right) \approx \left(\left(p_{it}^{(1)} - p_{it}^{(2)}\right) - E\left(p_{it}^{(1)} - p_{it}^{(2)}\right)\right)^2$$ [7]

As equation (16) clearly shows, the variance in price savings is approximated using prices and price expectations only for the current trip. It therefore represents the "realized uncertainty" associated with actual savings for item *i* on trip *t*, somewhat analogous to the common rational expectations assumption regarding stochastic prices. While we would prefer to incorporate

[7] Because the variance term enters the predictor as a divisor, we cannot allow it to be zero. Following convention, we instead impute a small positive value (0.01).

historical information about prices and price expectations, we are unable to do so because cherry picking shoppers do not always cherry pick. For visits on which the shopper does not cherry pick: (i) we cannot use equation (15) for expected savings, (ii) the order in which stores are visited (after the shopping list is created) is impossible to ascertain, and (iii) the shopper is far less likely to compare her expected price savings with the true savings, anyway. Thus, we use the squared error of the savings expectation on the current trip as a proxy for the uncertainty of that savings. As with our model of the expected price differences, we offer (16) as a reasonable model of the variance in price differences. Other plausible models undoubtedly exist.

3.2. Data. The model is estimated using IRI panel data from a major US market over 104 weeks between October 1995 and October 1997. The data are divided into an estimation period (weeks 1-91) and a prediction period (weeks 92-104). Panelists record the UPCs (uniform product codes) of all packaged goods products purchased on all trips to a wide variety of retailers, identifying the retailer by store chain rather than by individual store location (see Fox 1999 for a detailed description of this dataset). This purchase database is supplemented with a merchandise file containing all item prices for fifteen packaged goods categories.[8] We focus our empirical analysis on purchases in these categories.

Our conceptual framework of cherry-picking behavior allows the consumer's shopping list to be split over multiple retailers. This is not empirically observable, however, because we do not know when lists are compiled and which visits are associated with which shopping list. For example, a shopper may buy some items from her list on the way home from work on a weekday, and buy the remaining items at a different store on the weekend. Alternatively, she may shop for the items on her list at both a grocery retailer and a mass merchandiser, or supercenter. Given our inability to observe when a shopping list is developed, we apply a very conservative criterion to identify cherry picking. If the household visits two grocery retailers on the same day (i.e., without intervening consumption), we assume that the shopper is cherry picking. The primary consequence of applying this conservative test for identifying cherry-picking behavior is that many instances are not captured, and our empirical analysis is limiting to "obvious" episodes of cherry picking. As stated previously, we focus on the two largest grocery retailers in a duopolistic market. Together, these two retailers account for 64% of all grocery purchases by the panelists in the database, and 57% of all identified instances of cherry picking.

Having identified cherry picking shoppers and visits, we consider purchases of items in the fifteen categories for which item-level price data are available. We include only purchases of items available at both retailers so that the dataset is consistent with the assumptions underlying our analytical and econometric models.[9] Following Kamakura, Kim and Lee (1996), we eliminate households from the dataset that make fewer than five purchases (from the commonly available items in our fifteen categories) on cherry-picking visits over the 104 weeks, ensuring that we have enough observations to make household-level predictions. The resulting dataset contains 103 households making a total of 2144 purchases while cherry picking during the estimation period and 315 purchases while cherry picking during the prediction period. Table 1

---

[8] Available categories are beer & ale, chocolate candy, carbonated beverages, salty snacks, coffee, facial cosmetics, internal analgesics, sanitary napkins, shampoo, vitamins, cigarettes, diapers, dog food, household cleaners, and laundry detergents.

[9] We do not distinguish unplanned from planned purchases because our analysis conditions on the household cherry picking. Given that the consumer cherry picks, we would argue that both planned and unplanned purchases are generally subject to the same decision process regarding whether to purchase at the first retailer or defer. For example, the cherry picker should not make an unplanned purchase at Retailer 1 without considering the potential savings of deferring that purchase to Retailer 2. While it is possible that an item will gain the attention of the cherry picker at Retailer 2, though it could have been purchased previously at Retailer 1 at a lower price, this is unlikely to cause much measurement error. On the other hand, adding a latent variable specification to distinguish planned from unplanned purchases would substantially increase the complexity to our model. We acknowledge, however, that were we to analyze the *probability* of cherry picking (we do not), we would have to attempt to determine which items were on the *a priori* shopping list.

shows descriptive statistics for households we identify as cherry pickers. Note that the households cherry-pick the two primary retailers (i.e., visit both retailers without intervening consumption) an average of 14.2 times per year out of 41.4 and 50.3 annual visits to Retailers *A* and *B*, respectively. They spend somewhat less per visit when cherry picking ($62.70) than on the average retail visit ($74.40).

3.3. Empirical Findings

Model selection. The numbers of segments, *N* and *V*, are unknown, and must be inferred from the data. We determine the number of risk-neutral and risk-averse segments by applying a variant of the sequential addition heuristic used by Kamakura, Kim, and Lee (1996). We begin by estimating two models—one with a single risk-neutral segment, the other with a single risk-averse segment. Because both models require estimation of the same number of parameters, we compare the log-likelihoods and select the model that fits better. Next, we estimate two two-segment models, obtained by adding either a risk-neutral or a risk-averse segment to the best fitting one-segment model. Again, we directly compare the log-likelihoods of the two-segment models to determine which offers the best fit. We continue to add segments sequentially until the incremental improvement in log-likelihood is not sufficient to warrant the estimation of additional parameters. We assess the tradeoff between fit and parsimony by computing both Bayesian Information Criterion (BIC, Schwartz 1978) and Consistent Akaike Information Criterion (CAIC, Bozdogan 1987).

Applying our sequential addition heuristic, we estimate configurations of *n* and *v* for one- to six-segment solutions (see Table 2). Both CAIC and BIC are minimized in the five-segment solution, $n = 2$ and $v = 3$; hence, this number of risk-neutral and risk-averse segments represents the best balance between model fit and parsimony. The table also offers U^2 statistics, which compare the various models estimated to a simple one-segment model with intercept only. While not directly interpretable, $U^2 = 0.239$ for our preferred five-segment solution shows that our model offers substantial explanation of the phenomenon being studied.

The structural mixture model specification was estimated to investigate our theoretical findings regarding shoppers' risk attitudes and the propensity to defer purchase when cherry picking. Accordingly, we evaluate the usefulness of modeling risk attitudes by comparing the indicated five-segment structural mixture model with risk-averse and risk-neutral only specifications (see Table 3). Among possible risk-averse only specifications, a five-segment solution minimizes CAIC and BIC, while a seven-segment risk-neutral only specification minimizes the information criteria. Thus, these specifications are selected for comparison. The structural mixture model offers better in-sample fit compared with both the risk-neutral or risk-averse only model (again, based on CAIC and BIC). In particular, it fits far better than the risk-averse only model. The structural mixture model also fairs reasonably well out-of-sample, offering the same hit rate as the risk-averse only model for the 315 hold-out purchases, albeit with a slightly inferior log-likelihood. Moreover, the structural mixture dominates the risk-neutral only model in terms of in- and out-of-sample fit.

Parameter estimates. Table 4 shows the expected signs of parameters, where applicable, and the interpretations of these signs. Our investigation focuses on (i) the estimates of β_3, reflecting the expected savings to deferring purchase and expected-savings-divided-by-variance for risk-neutral and risk-averse cherry pickers, respectively, and (ii) the intercept parameter, β_0, which captures the propensity to purchase at the first store visited. In accordance with Propositions 1 and 2, we expect both β_3 parameters to be less than zero, because the cherry picker would expect to save by deferring purchase to Retailer 2. Proposition 3 leads to the expectation that β_0 will be higher for risk-averse than for risk-neutral cherry pickers. In addition, we expect the β_2 parameter for category-specific loyalty to be positive (Bell, Ho, and Tang 1998) and the α parameter for studying of the retailer ads to be closer to 1 than 0 (Urbany, Dickson, and Key 1991, Urbany, Dickson, and Sawyer 2000).

Parameter estimates and segment sizes for the five segments are shown in Table 5. No standard errors or asymptotic t-tests are reported, because our maximum likelihood estimation is constrained. Specifically, three of the segment-level α parameters are constrained at the upper bound, so their confidence intervals (and those of any variables correlated with them) violate asymptotic assumptions. Note that these violations are specific to our empirical application, and do not necessarily follow for other applications of this model. Even without standard errors, the segment-level parameter estimates are informative. Proposition 1 predicts that, for risk-neutral cherry pickers, the probability of purchasing at Retailer 1 decreases as the expected savings from deferring to Retailer 2 increases. This proposition is clearly supported, because the β_3 parameters for the risk-neutral segments (1 and 2), i.e., the support points for the discrete mixture model, are negative. Household-level β_3 parameters for risk-neutral shoppers, which must be a convex combination of the segment parameter estimates, are necessarily negative. Proposition 2 predicts that, for risk-averse cherry pickers, the probability of purchasing at Retailer 1 decreases as the expected-savings-divided-by-variance term from deferring to Retailer 2 increases. This proposition is also supported, as the β_3 parameters for risk-averse segments 3, 4, and 5 are negative. Again, household-level parameter estimates must therefore be negative. Figure 3 shows distributions of household-level parameter estimates, obtained by weighting segment parameter estimates by the individual household’s posterior probability of membership in each segment. We find that, as expected, the confidence intervals of the β_3 parameters (panels "d" and "e") do not include zero. We also observe that the confidence interval for β_1 (panel "b"), the category-specific loyalty coefficient, does not include zero. Thus, prior share-of-requirements loyalty is a significant positive predictor of the household's purchase at the first store. Both β_0 and β_2 (panels "a" and "c") include zero within the 95% confidence interval, so their mean estimates are not significantly different from zero.

Segment assignment and risk aversion. Segment sizes are also included in Table 5. Though the other segments are substantial, segment 2 represents only 4.4% of households and only two households are discretely assigned to this segment. We apply DeSarbo, et al.'s (1992) entropy measure to assess the overlap in household assignment across segments.[10] The computed entropy value of 0.74, while closer to one than zero, suggests some overlap in segment assignments. This is illustrated in Figure 4, which shows a frequency plot of the posterior probability of household assignment to a risk-averse segment. We find that 43.8% of households exhibit risk aversion (i.e., are members of risk averse segments) with probability > 0.9, 67.6% of households are risk averse with probability > 0.5 and 81.0% of households are risk averse with probability > 0.1. Thus, most cherry-picking households exhibit some degree of risk-averse purchase behavior, providing strong empirical support for our decision to incorporate risk aversion in our analytical and econometric models.

Regressing demographic variables (family size, income, home ownership, education, and working woman) on a discretized risk-aversion variable yields only one significant relationship.

[10] The degree to which segment membership is distinct can be determined using a measure of entropy due to DeSarbo, et al. (1992):

$$Entropy = 1 + \left(\sum_h \sum_n (\rho_{hn} \ln \rho_{hn}) + \sum_h \sum_v (\rho_{hv} \ln \rho_{hv}) \right) / H \ln(N+V)$$

Note that the entropy measure is bounded 0,1. The higher the entropy value, the more distinct the segment classifications; the lower the value, the more overlap in membership across segments

College education is a positive predictor of risk aversion ($p = 0.009$); other demographics are not predictive.

Recall that Proposition 3 predicts that risk-averse cherry pickers are likely to buy products with little or no expected savings at the first store visited, all other things equal. To gather additional evidence related to this proposition and associated conjectures, we compute simple correlations between household-level *risk aversion* (i.e., the posterior probability of household membership in one of the three risk-averse segments) and (i) the α parameter for *studying retail ads*, as well as (ii) the β_0 intercept parameter representing the *propensity to buy at the first store visited*. We find both correlations to be positive and highly significant ($p < 0.0001$). The correlation between risk aversion and the propensity to buy at the first store is in the predicted positive direction ($r = 0.576$). While this simple test does not distinguish between product purchases with larger or smaller expected savings, the positive correlation is consistent with Proposition 3. Interestingly, risk aversion is even more closely related to the shopper's propensity to study retail ads ($r = 0.880$). Thus, shoppers who are concerned about the probability of savings as opposed to strictly the expectation of savings are more likely to gather information prior to shopping.

4. Discussion and Managerial Implications

We return to the cherry picker's problem of deciding where to purchase the items on her shopping list, given visits to two retailers in a predetermined sequence. Both risk-averse and risk-neutral cherry pickers benefit from reducing uncertainty about prices at the second store that they visit. For the risk-neutral cherry picker, the sum of savings realized over the entire basket is maximized when prices at both stores are known, because savings is maximized for each item.

For the risk-averse shopper, the *probability* of saving on each item in the basket is also 1 when prices at the second store are known. In either case, both retailers sell only items that they price lower than their competitor.

The cherry picker's primary method of reducing price uncertainty is to study retailer ads. Consistent with previous studies (Urbany, Dickson, and Key 1991, Urbany, Dickson, and Sawyer 2000), we find that cherry pickers *do* study retailer ads (mean probability = 0.969). The retailer can also reduce uncertainty about price savings, thereby affecting where cherry pickers will buy. For items priced below competition, the retailer can maximize its sales to cherry-picking shoppers by reducing the uncertainty of the expected savings. This can be accomplished by consistently pricing a set amount below competition. Such a strategy requires that the retailer's timing and depth of promotions match competition, resulting in a high positive correlation in price between the two retailers. For products that it prices above competition, the retailer can maximize its sales to cherry pickers by inducing a high negative price correlation with competition. This can be accomplished by timing item-level promotions so that they do not coincide with competitor's promotions, and by varying promotional discount depth.[11]

5. Limitations and Future Research

The primary limitation of this research is our inability to authenticate many cherry-picking visits in the empirical data. We have applied a very restrictive criterion, i.e., visits to multiple grocery retailers without intervening consumption, which clearly understates the prevalence of cherry picking. This limitation is an obstacle to actually determining the frequency of consumer cherry picking. Moreover, our econometric study implicitly assumes that consumers behave the same

[11] Alternatively, if shoppers are able to exploit negative price correlations to effectively predict price differences, the retailer may randomize promotions on these items so as to eliminate any correlation.

way when visiting multiple grocery stores in a given day as they would if splitting their purchases with an intervening break. We have also limited our analysis to the two largest grocery chains in a duopolistic market, and therefore do not consider the possibility of cherry picking across three or more retailers. Additionally, we have implicitly assumed that consumer decision processes and preferences do not vary over time. Though this assumption is common in choice modeling, it is possible that consumers' risk profiles and preferences are non-stationary.

Though we find compelling evidence for the analysis that we have presented, the manner in which we have modeled expected price differences and their variances represents only one possible path. Our general theory addresses the ratio of an item's expected price difference to the variance of that difference. How these expectations and variances are modeled is open to competing possibilities, as well as many levels of sophistication and creativity. This represents a significant opportunity for future research.

Other opportunities remain to investigate and characterize consumer cherry picking. Of primary interest to retailers is the size of the cherry-picking segment. While it is believed that cherry pickers are less profitable than other retail customers, research in this area—particularly relating profitability to retailer pricing and promotion decisions—would be quite useful. Further, if in fact cherry pickers are systematically less profitable, identifying them using retailer customer data would represent a potentially important contribution.

References

Barnard, P. O. and D. A. Hensher (1992). The Spatial Distribution of Retail Expenditures. *Journal of Transport Economics and Policy*, (September), 299-312.

Bell, D.R., T. Ho, and C.S. Tang (1998). Determining Where to Shop: Fixed and Variable Costs of Shopping. *Journal of Marketing Research*, 35(3), 352-369.

Bozdogan, H. (1987). Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions. *Psychometrika*, 52, 345-70.

Bucklin, R. and J. M. Lattin (1992). A Model of Product Category Competition Among Grocery Retailers. *Journal of Retailing*, 68(Fall), 271-93.

Consumer Reports (1988). How Much Money Smart Supermarket Shopping Can Save. (March)158.

Dellaert, B. G. C., T. Arentze, M. Bierlaire, A. W.Borgers, and H. J. P. Timmermans (1998). Investigating Consumers' Tendency to Combine Multiple Shopping Purposes and Destinations. *Journal of Marketing Research*, 17(May), 177-88.

DeSarbo, W. S., M. Wedel, M. Vriens, and W. Ramaswamy (1992). Latent Class Metric Conjoint Analysis. *Marketing Letters*, 3(3), 273-88.

Dreze, X. (1995). Rehabilitating Cherry-Picking. Working paper. Los Angeles, CA: UCLA.

Fox, E.J. (1999). *Analysis of Household Shopping Behavior Across Retail Formats*. Unpublished Ph.D. Dissertation, Wharton School, Philadelphia, PA.

Fox, E. J., R. Metters, and J. Semple (2002). Every House a Warehouse: An Inventory-Theoretic Model of Retail Shopping Behavior. Working paper. Dallas, TX: Southern Methodist University.

Grover, R. and V. Srinivasan (1992). Evaluating the Multiple Effects of Retail Promotions on Brand Loyal and Brand Switching Segments. *Journal of Marketing Research*, 24(February), 76-89.

Kahn, B. E. and L. McAlister (1997). *Grocery Revolution: The New Focus on the Consumer*. ed. Joel Steckel. Reading, MA: Addison-Wesley.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(March), 263-91.

Kamakura, W. A., B. D. Kim, and J. Lee (1996). Modeling Preference and Structural Heterogeneity in Consumer Choice. *Marketing Science*, 15(2), 152-72.

Kumar V. and Leone, R. V. 1988. Measuring the Effect of Retail Store Promotions on Brand and Store Substitution. *Journal of Marketing Research*, 25(May), 178-85.

Lal, R. and R. C. Rao (1997). Supermarket Competition: The Case of Everyday Low Pricing. *Marketing Science*, 16(1): 60-80.

Mayhew, G. E. and R. S. Winer (1992). An Empirical Analysis of Internal and External Reference Prices Using Scanner Data. *Journal of Consumer Research*, 19(June), 62-70.

Mogelonsky, M (1995). *Everybody Eats: Supermarket Consumers in the 1990s*. American Demographic Books: Ithaca, NY.

Mogelonsky, M (1994). Please Don't Pick the Cherries. *Marketing Tools*. (September/ October), 10.

MMR (2002). Wal-Mart's Challenge. 19(5), 5.

Schwartz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* 6, 461-4.

Urbany, J. E., P. R. Dickson, and A. G. Sawyer (2000). Insights into Cross- and Within-Store Price Search: Retailer Estimates Vs. Consumer Self-Reports. *Journal of Retailing*, 76(2), 243-58.

Urbany, J. E., P. R. Dickson, and R. Key (1991). Actual and Perceived Consumer Vigilance in the Retail Grocery Industry. *Marketing Letters*, 2(1), 15-25.

Walters, R. G. (1991). Assessing the Impact of Retail Price Promotions on Product Substitution, Complementarity, Purchase and Interstore Sales Displacement. *Journal of Marketing* (April): 17-28.

Walters, R. G., and MacKenzie, S. 1988. A Structural Equations Analysis of the Impact of Price Promotions on Store Performance. *Journal of Marketing Research* (February): 51-63.

Walters, R. G., and Rinne, H. J. 1986. An Empirical Investigation into the Impact of Price Promotions on Retail Store Performance. *Journal of Retailing* 62(3): 237-66.

Figure 1
The Optimal Proportion of an Item to Buy at the Second Retailer.

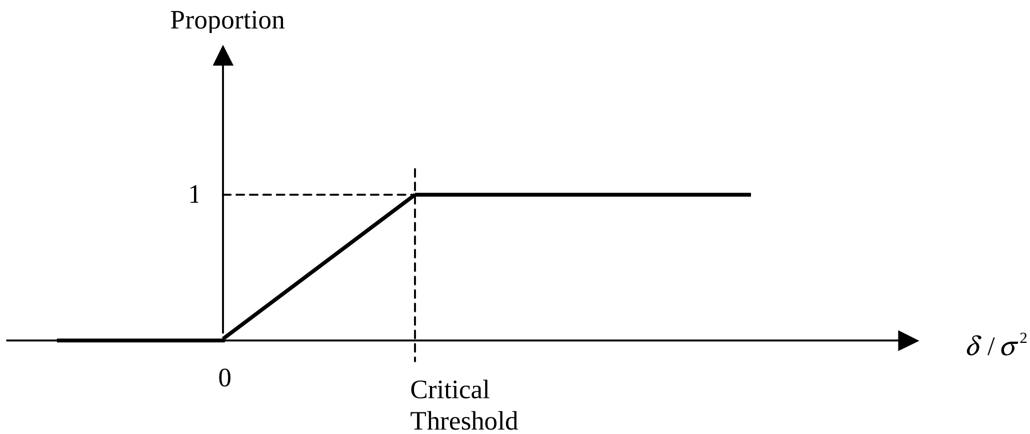


Figure 2
Routes of Return Beginning with Different Retailers

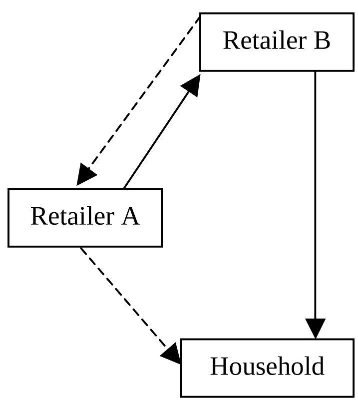
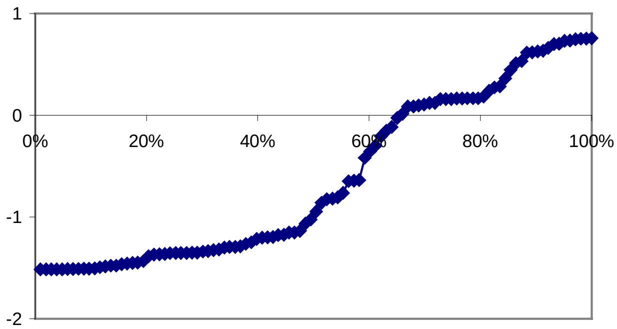


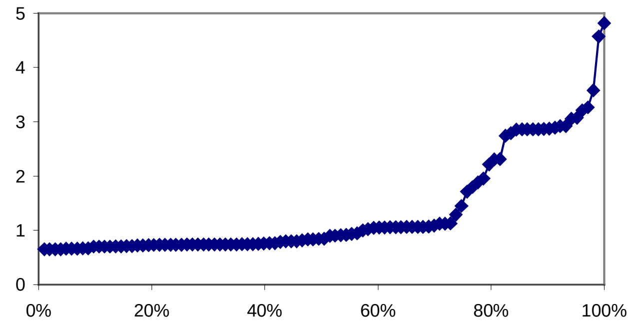
Figure 3

Distributions of Household-Level Parameter Estimates

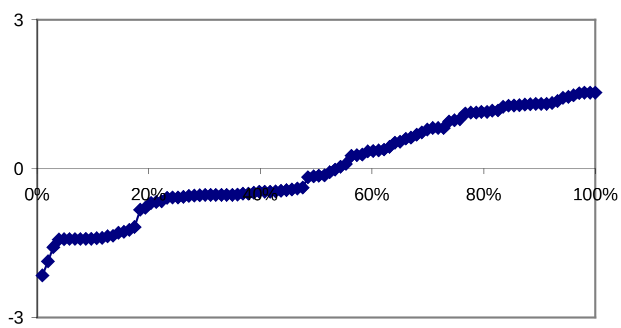
a. Preference for First Retailer Visited



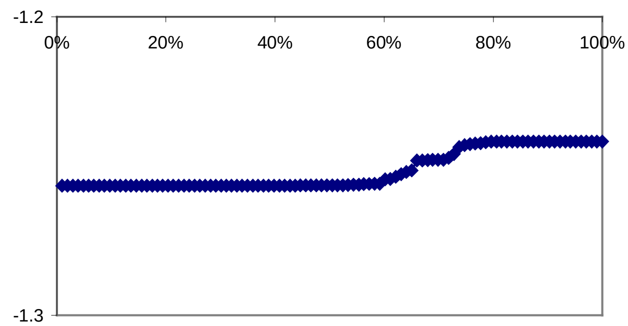
b. Category-Specific Loyalty



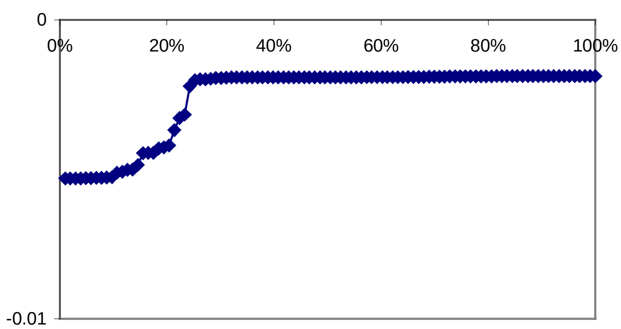
c. Preference for Retailer "A"



d. Expected Savings (Risk-Neutral)



e. Expected Savings / Variance (Risk-Averse)



f. Study of Retailers' Feature Ads

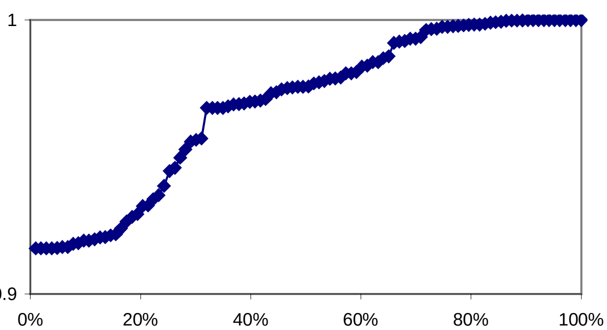


Figure 4

Posterior Probability of Risk-Aversion

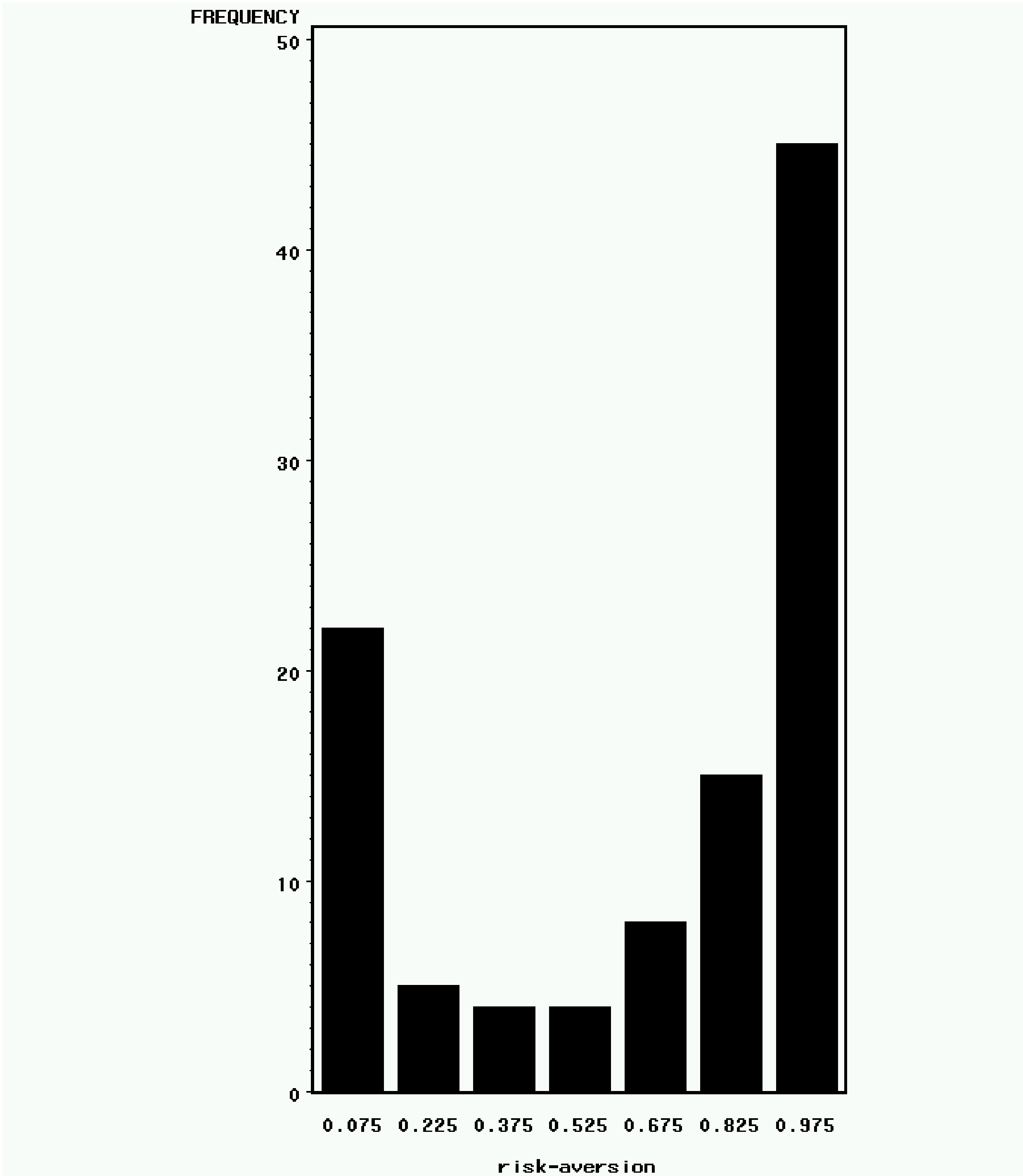


Table 1
Descriptive Statistics for Cherry-Picking Households

		N	Mean	StdDev
Store Visits				
	Retailer A (per year)	206	41.4	33.6
	Retailer B (per year)	206	50.3	30.5
	Spending per Visit ($)	21611	74.4	74.3
Cherry Picking Store Visits				
	Retailer A (per year)	206	14.2	10.7
	Retailer B (per year)	206	14.2	10.7
	Spending per Visit ($)	3376	62.7	61.6
Shopper Demographics*				
	Family Size (#)	100	3.12	1.42
	Income (x $1000)	100	55.3	25.3
	Working Woman (%)	100	63.0%	48.5%
	College Educated (%)	100	24.0%	42.9%
	Homeowner (%)	100	90.0%	30.2%

* Demographics were available for only 100 of the 103 cherry-picking households

Table 2
Comparison of Structural Mixture Models

Total Number of Segments	Structure N	Structure V	Log-Likelihood	CAIC	BIC	Number of Parameters	u^2
1	1	0	-1349.3	2713.7	2715.3	5	0.093
2	1	1	-1220.3	2470.8	2473.9	10	0.179
3	1	2	-1180.2	2405.5	2410.3	15	0.206
4	1	3	-1154.5	2369.2	2375.6	20	0.224
5	2	3	-1138.7	2352.6 *	2360.6 *	25	0.234
6	2	4	-1131.5	2353.3	2362.9	30	0.239

* Lowest information criterion

Table 3
Comparison of Structural Mixture, Risk-Neutral Only, and Risk-Averse Only Models

Description	# of Segments N	V	Parameters	In Sample Log-Likelihood	CAIC	BIC	Out of Sample Log Likelihood	Hit Rate
Structural Mixture	2	3	25	-1138.7	2352.6	2360.6	-37.4	0.61
Risk-Neutral Mixture	7	0	35	-1130.4	2366.2	2377.4	-42.1	0.55
Risk-Averse Mixture	0	5	25	-1165.0	2405.4	2413.4	-36.8	0.61

Table 4
Parameter Descriptions, Expected Signs and Interpretations

Parameter	Description	Expected Sign	Interpretation of Positive (+) Sign
β_0	Intercept	Risk Averse > Risk-Neutral	Cherry picker prefers to buy at the first store visited
β_1	Category-Specific Loyalty	+	Prior loyalty to Retailer 1 for products in that category increases utility for purchase at Retailer 1
β_2	Indicator for Retailer A		Cherry picker prefers to purchase at Retailer A vs. Retailer B, regardless of which is visited first
β_3 Risk-Neutral	Expected Price Savings	-	Utility and purchase probability decrease in expected price savings from deferring purchase
β_3 Risk-Averse	Expected Price Savings Divided by Variance	-	Utility and purchase probability decrease in expected-price-savings-divided-by-variance from deferring purchase
α	Probability of Studying Retailers' Ads	+	Cherry picker has a positive probability of studying retailers' feature ads

Table 5
Parameter Estimates and Segment Sizes for (2,3) Structural Mixture Model

	Risk-Neutral		Risk Averse			(Weighted) Average
	Seg 1	Seg 2	Seg 3	Seg 4	Seg 5	
Parameter Estimate						
β_0	-1.515	0.240	0.167	-1.351	0.756	-0.612
β_1	0.738	4.819	1.061	0.652	2.860	1.337
β_2	-0.526	-2.154	-1.418	1.529	1.306	0.012
β_3 Risk-Neutral	-1.257	-1.242				-1.255
β_3 Risk-Averse			-0.002	-0.002	-0.005	-0.003
α	0.917	1.000	1.000	1.000	0.968	0.969
Segment Size						
Probabilitistic Assignment (% of hh)	30.1%	4.4%	23.6%	24.7%	17.1%	16.7%
Discrete Assignment (of 103 hh)	32	2	23	24	22	20.6

Appendix

Proof of Theorem 1. Since $\sum_{i\in I^+} q_{it}\cdot\delta_{it} > k$, at least one component of the optimal solution is positive. Let this component be denoted by λ_{it}^*. Now suppose $\lambda_{jt}^* = 0$ for some item $j \in I^+$. Consider a new solution with infinitesimal perturbations in two components: $\lambda_{it} = \lambda_{it}^* - \varepsilon$, $\lambda_{jt} = \frac{q_{it}\cdot\delta_{it}}{q_{jt}\cdot\delta_{jt}}\varepsilon$ ($\varepsilon > 0$). The new solution preserves the value of g as well as the numerator in the argument of Φ. The net change in $\Phi's$ denominator is

(A1) $$\left(-2\lambda_{it}^*\varepsilon + \varepsilon^2\right)q_{it}^2\sigma_{it}^2 + \left(\frac{q_{it}\delta_{it}}{q_{jt}\delta_{jt}}\varepsilon\right)^2 q_{jt}^2\sigma_{jt}^2.$$

For sufficiently small $\varepsilon > 0$, (A1) is negative, whereupon the term $1-\Phi$ would increase, thus increasing h as well (recall h is strictly increasing), thus violating the optimality of λ_t^*. This proves part 1.

To prove part 2, we consider a similar infinitesimal perturbation: $\lambda_{it} = \lambda_{it}^* - \varepsilon$ and $\lambda_{jt} = \lambda_{jt}^* + \frac{q_{it}\delta_{it}}{q_{jt}\delta_{jt}}\varepsilon$ for $i, j \in I^+$. Again, this maintains the value of g as well as the numerator in Φ's argument. Consequently, the denominator of Φ's argument must increase or stay the same so that optimality of λ_t^* is preserved. Since we may take ε to be either positive or negative in this case, we must have

(A2) $$\left(-2\lambda_{it}^*\varepsilon + \varepsilon^2\right)q_{it}^2\sigma_{it}^2 + \left[2\lambda_{jt}^*\frac{q_{it}\delta_{it}}{q_{jt}\delta_{jt}}\varepsilon + \left(\frac{q_{it}\delta_{it}}{q_{jt}\delta_{jt}}\varepsilon\right)^2\right]q_{jt}^2\sigma_{jt}^2 = 0.$$

Dividing (A2) by $\varepsilon \neq 0$, taking the limit as $\varepsilon \to 0$, and rearranging the remaining terms yields the expression in part 2.

To prove part 3, consider the same perturbation used in part 2, with the exception that now $\lambda_{it}^* \leq 1$, which limits perturbations to one-sided changes $\varepsilon > 0$. The optimality of λ_t^* implies

(A3) $$\left(-2\lambda_{it}^*\varepsilon + \varepsilon^2\right)q_{it}^2\sigma_{it}^2 + \left[2\lambda_{jt}^*\frac{q_{it}\delta_{it}}{q_{jt}\delta_{jt}}\varepsilon + \left(\frac{q_{it}\delta_{it}}{q_{jt}\delta_{jt}}\varepsilon\right)^2\right]q_{jt}^2\sigma_{jt}^2 \geq 0.$$

Dividing (A3) by $\varepsilon > 0$, taking the limit as $\varepsilon \to 0$, and rearranging the remaining terms yields

$$\frac{\lambda_{jt}^*}{\lambda_{it}^*} \geq \frac{q_{it}^2\sigma_{it}^2}{q_{jt}^2\sigma_{jt}^2}\cdot\frac{q_{jt}\delta_{jt}}{q_{it}\delta_{it}}.$$

But $\frac{\lambda_{jt}^*}{\lambda_{it}^*} < 1$, hence $1 > \frac{q_{it}^2\sigma_{it}^2}{q_{jt}^2\sigma_{jt}^2}\cdot\frac{q_{jt}\delta_{jt}}{q_{it}\delta_{it}}$, which implies part 3.