ESSAYS ON THE INTERACTION OF PRODUCT CHARACTERISTICS, CONSUMER HETEROGENEITY, AND INFORMATION

By

ROGER ALLEN BAILEY

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

December, 2013

Nashville, Tennessee

Approved:

Andrew F. Daughety, Ph.D.

Jennifer F. Reinganum, Ph.D.

Tong Li, Ph.D.

Jeffrey P. Dotson, Ph.D.

ACKNOWLEDGMENTS

First and foremost, I express my gratitude to my Lord and Savior, Jesus Christ, for providing me with a determined, inquisitive spirit, and for being an ever-present help in all pursuits. I also express my great appreciation for the advice and support provided by my co-advisors, Professors Andrew Daughety and Jennifer Reinganum, throughout the research in this dissertation. Additionally, I thank Professor Tong Li and Professor Jeffery Dotson for their advice with my third and fourth chapters, and for their work on my dissertation committee. I gratefully acknowledge the financial support provided by NSF grant SES-0814312, as well as that provided by the Harold Sterling Scholarship and other sources from Vanderbilt University. I also appreciate the hard work and incite from PJ Glandon, the coauthor of the third chapter of this dissertation. Finally, I am very grateful to Jane, my wonderful wife, and to my children, Annabelle and Nathan, for their continuous and unwavering support of my academic pursuits.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

My primary research interest lies in applying micro-economic theory and structural modeling techniques to the fields of industrial organization and marketing. More specifically, I am interested in the interaction of product characteristics and consumer heterogeneity, and the role of information about products and individuals in the decision-making processes of consumers and firms. To this end, the chapters in this dissertation apply a wide range of techniques to investigate models in which product characteristics strike different consumers in different ways, analyzing how information is used or could be used by the consumers and firms in such markets. Chapter 2 employs methods from the theoretical literature on quality and price signaling to investigate a model wherein a firm has private information about its product that consumers can incur a cost to learn before making a purchase decision. Chapter 3 uses a structural model to empirically investigate the importance of a specific type of consumer information, and Chapter 4 provides a Bayesian technique for investigating household-level heterogeneity in purchase decisions. This dissertation investigates related questions from very different angles, demonstrating some of the vast array of tools at the disposal of economists and marketers, and applying these methods to advance knowledge about the interaction of product characteristics, consumer heterogeneity, and information.

The second chapter of the dissertation seeks to analyze the role of price signaling and consumer audit in a model where consumers have heterogeneous tastes over a monopolist's product. In this model, consumers are not only heterogeneous in their base taste levels for the product itself, they are also heterogeneous in the way that the product's single characteristic affects their utility; some consumers desire a higher level of the characteristic, while others desire a lower level of the characteristic. One example of such a characteristic could be the proportion of foreign-produced inputs of a product, wherein some consumers prefer domestically-produced inputs while others do not. Another example could be the fat content of a product, for which some consumers prefer the improved taste of a high fat content while others prefer low-fat, healthier products. Such differences in a product's characteristics may also lead to substantial differences in production costs, a consideration which is included in the construction of the model. In short, this chapter develops an asymmetric information model of a monopolistic firm wherein consumers have heterogeneous preferences over the firm's product. The firm has ability to signal via price, but also faces the possibility of consumers choosing to incur an audit cost to acquire this information before purchase. As a result, the firm must choose whether to signal via price or to leave the decision (and burden) of becoming informed to the consumer. This model is valuable to the field of industrial organization since a novel product, about which consumers have limited information, will often have attributes that strike different consumers in very different ways. Given the vast amount of available product information via published product specifications, online product reviews, and other sources, these consumers also face a choice of whether or not to incur a cost (monetary or in effort) to become fully informed about a product before they make a purchasing decision. Therefore, firms must account for both consumer heterogeneity and information acquisition when choosing price, and understanding this dynamic is important for both economists and policymakers.

The resulting equilibria of the model are analyzed with and without the possibility of consumer verification. As a result of the structure of the model, expected aggregate demand will either be increasing or decreasing in the level of the characteristic, depending on the relative proportion of consumer types and the weight that the respective types put upon the level of the characteristic. Therefore, expected aggregate demand can be used to determine which of the firm's types is "strong" and which type is "weak." When expected aggregate demand is increasing in the level of the characteristic, the high type is strong (preferred on average), and when expected aggregate demand is decreasing in the level of the characteristic, the low type is strong. In the case where consumer verification is not possible, it is found that there exists a unique Perfect Bayesian equilibrium satisfying the Intuitive Criterion of Cho and Kreps (1987) in which the level of the product's characteristic is revealed through price signaling. When the low type is strong, the level of the characteristic is signaled with an upward-distortion in the low-type's equilibrium price, and when the high type is strong, the level of the characteristic is signaled with a downward-distortion in the high-type's equilibrium price.

When the possibility of consumer verification is added to the model, so that after learning the product's price a consumer can choose to incur an audit cost in order to learn the value of the product's characteristic, there exist conditions on the parameters of the model such that there will exist a second equilibrium in addition to the aforementioned separating equilibrium. In this second equilibrium type, the monopolistic firm charges the same price regardless of the level of the characteristic, and all consumers choose to verify the level of the characteristic. That is, the firm does not reveal the level of the characteristic through signaling, but shifts the burden of becoming informed onto consumers.

As it is possible for two equilibria to survive refinement in the case when consumers are able to audit, I conduct a welfare analysis of the two equilibria, and show that either equilibrium may be more efficient, depending on the values of the parameters. I concluded that in order to maximize social welfare any efforts by a policymaker to make product information more accessible to consumers should be selective in the products they target, as lowering the cost of consumer audit can actually result in a less efficient equilibrium in some cases.

Chapter 3 is a paper co-authored by P.J. Glandon, wherein we analyze the value of information regarding how much a consumer cares about their health (consumer health conscientiousness) in the estimation of demand for products wherein a brand's "healthiness" may be of concern. As disentangling the effects heterogeneity in healthy vs unhealthy eating habits could be valuable in determining a firm's optimal strategy across markets in which consumer preferences over "healthy" vs "unhealthy" products differ, this investigation is important to both economists and marketers. Given the active policy debate over nutrition and health, the results of our analysis are also valuable to policymakers.

As a consumer's determination of a product's "healthiness" is a matter of perception, which could be argued to be a function of any number of observable and unobservable product characteristics together with the consumer's level of health conscientiousness, one could attempt to perform an analysis that controls for heterogeneity in the perception of a product's "healthiness." However, as our goal is to analyze the value of information regarding consumer health conscientiousness in the estimation of demand, such an analysis is unnecessary (and would almost necessarily be confined to an experimental setting). Therefore, this chapter investigates the significance of the coefficients on the interactions of a variable measuring health conscientiousness with observable product characteristics in the determination of demand across a category wherein a brand's "healthiness" is important to some consumers.

The ideal dataset for our study would be a panel of individual (household) purchases, complete with demographic information including a variable serving as a measure of health conscientiousness. However, to our knowledge no such panel currently exists. Therefore, we apply the discrete choice model presented in Nevo (2001) to a simulated population of households by combining data from both the Current Population Survey (henceforth CPS) and the Behavioral Risk Factor Surveillance System (henceforth BRFSS) in order to associate each household with a dummy variable for smoking. Given the well known and very serious health implications of smoking, the result is a simulated panel of households for each city that includes a proxy variable for consumer health conscientiousness. This dataset is then used to investigate whether or not the coefficients on the interaction of the smoking dummy and health-related product characteristics are statistically significant in the estimation of demand in the ready-to-eat cereal category. A category chosen for the substantial variation in the characteristics associated with healthy eating, such as sugar, fat, and fiber content, and for the clear presence of the targeting of health conscious consumers with specific brands.

The estimation proceeds by using the aforementioned simulated population of households to predict the shares of each cereal in a given city, conditional on the value of the model parameters including the values of the coefficients on the interaction of the smoking dummy variable and the amount of sugar, fat, and fiber in each cereal. Estimation is completed by finding the vector of parameters that minimizes the distance between the observed shares and the resulting simulated shares. Our results suggest that the interaction of the smoking behavior and the sugar content of a cereal may be important in the estimation of demand within the ready-to-eat cereal category, and may therefore be a worthwhile addition to the commonly collected demographic information in household panel datasets.

Chapter 4, the final chapter of this dissertation, presents a marketing model for handling a commonly overlooked problem in the estimation of household-level preferences: within-household heterogeneity. A significant challenge in marketing analysis is to construct models that handle consumer heterogeneity in such a way as to draw inference about "average" consumer preferences while still allowing for the recovery of actionable information about how these preferences differ across the consumer base. This goal is complicated when the "individual" in a dataset is a household, a fairly common situation given the prevalence of household-level panel datasets in marketing analysis. It is widely accepted that the decisions of a household are affected by the interactions of the individuals within the household, and that a single set of preferences is insufficient to control for the amalgamation of the several sets of preferences that could underly observed purchasing behavior. Though there are many methods for handling heterogeneity across households, there are few models designed to handle heterogeneity within households. As a result, most models in the extant literature on the estimation of consumer preferences either treat the household as having a single set of preferences, or ignore the household structure altogether and assume that individuals act independently. These approaches are somewhat defensible in much of the econometrics literature, as the parameters describing individual consumer preferences are typically considered nuisance parameters in the estimation of the aggregate-level parameters used in market analysis and/or policy; however, given the aforementioned goals, such models are inadequate for the purposes of marketing. The few marketing models that do attempt to include within-household heterogeneity are limited in the datasets to which the model can be applied and/or yield little to no actionable information about how households differ in their preferences over products and product characteristics.

By comparison, the model presented in chapter 4 uses Bayesian techniques to fit a finite mixture model. In a typical finite mixture model, the preferences of an individual household are assumed to be described by one of finitely many "mixing distributions" or segments. I adapt the method of Allenby, Arora, and Ginter (1998) to allow each purchase of a household to be the result of a mixture of such segments, using the assumption that the choices of demographically similar households will tend to have similar probabilities of belonging to a given segment. This allows my model to handle within-household heterogeneity as well as the multiple discreteness/variety seeking present in complex markets such as ready-to-eat cereal and soft drinks. Moreover, within-household heterogeneity is modeled in such a way as to estimate the distribution of preferences within each segment while providing useful information about the how the choices of specific household demographic types are distributed across the segments. As a result, the model provides both aggregate level information about preferences and the more granular information about individual households that is valuable in modern marketing. My model accomplishes this while being both logically and computationally suitable for larger household panel datasets, and therefore has substantial value in the field of marketing.

To demonstrate the ability of the model to retrieve the underlying parameters of the components of household preferences, the model is applied to two simulated household-level panel datasets.

The simulated purchases of each household in these datasets arise from a mixture of multiple sets of preferences, as described above, wherein the true values of the parameters of these preferences are known. The results of these simulations are presented, and in both cases the posterior draws of the model do well in navigating to the true underlying parameters used to generate the data. The chapter is then concluded by discussing how a practitioner could apply the model, and the common methods for handling potential challenges in the estimation process.

Taken as a whole, this dissertation presents a clear in-depth analysis of the interaction of product characteristics, consumer heterogeneity, and information. Three important questions/problems in industrial organization and marketing, related via my stated research interests, are analyzed using three different techniques. The results of each chapter contribute to the corresponding literature, and demonstrate the value in being able to approach a problem using a wide array of techniques.

CHAPTER II

QUALITY SIGNALING WITH COSTLY VERIFICATION BY HETEROGENEOUS CONSUMERS

II.1 Introduction

The goal of this paper is to analyze the role of price signaling and consumer verification in a model where consumers have heterogeneous tastes over a monopolist's product; that is, to investigate the signaling behavior of a monopolist when information about the product is available to the consumer at a cost (an audit cost). Such a model is interesting as consumers naturally differ in the way they perceive a specific characteristic of a good, thereby affecting each consumer's incentive to become informed, as well as the firm's incentive to signal the value of this characteristic. For example, consider a good that is produced using both foreign and domestic inputs (parts and labor). One consumer may consider a high proportion of foreign inputs to be a negative characteristic,[1] while another consumer may have no preference whatsoever, or may even prefer a high proportion of foreign inputs. Another example is the recycled content of a product, about which some consumers care greatly while others are indifferent. Even the fat content of a food product serves as an example, as some consumers may enjoy the additional flavor provided by a high-fat content, while other consumers would prefer a low-fat content despite a decrease in flavor. If the value of the characteristic in question (proportion of foreign inputs, recycled content, fat content) is the firm's private information, it is plausible that a consumer would not have to rely on the actions of the firm to become informed. To illustrate this, consider the example of a new product with foreign and domestic inputs. If this is a novel product, it is reasonable to assume that there are no close substitutes and that consumers will not know the proportion of the product's inputs that are foreign. While the firm may signal this information to the consumer via price, a consumer who cares about the proportion of foreign inputs may be able to access this information via the internet (or elsewhere) by incurring

[1]Effects of the consumer's purchase on his or her own income are assumed to be negligible; that is, the consumer's reasons for preferring domestically produced goods are not based on monetary benefits received due to the purchase. In this particular example, the consumer may just have a "made in the USA" preference.

some cost of effort. Therefore, the model is one of asymmetric information in that the level of the characteristic in question is the private information of the monopolistic firm that produces the product; however, this information may be available to consumers that are willing to incur an audit cost.

As consumers in this model are assumed to respond differently to the product characteristic in question, some consumers' demand curves will be increasing in the level of this characteristic, while the demand curves of others may be constant or even decreasing in the level of the characteristic. If, however, consumers' demand curves are aggregated, the resulting aggregate demand curve will be either increasing or decreasing in the level of the characteristic, and aggregate demand can then be used to vertically differentiate products that differ only in the value of this characteristic. To illustrate this using a more classic language as a convention, assume that X denotes the value of the characteristic in question. If aggregate demand is increasing in X, then the firm is "stronger" when it produces a product with a greater X, and if aggregate demand is decreasing in X, then the firm is "weaker" when it produces a product with a greater X. One could use a similar convention in comparing the model to the existing quality literature: a greater X results in a "higher-quality" product when the aggregate demand is increasing in X, and a greater X results in a "lower-quality" product when the aggregate demand curve is decreasing in X. Therefore "quality" could be increasing or decreasing in the level of the product's characteristic, depending on the preferences of different consumer types and the relative proportion of these consumer types within the population (consumer demographics). Consider again the example of a product with some proportion of foreign inputs. Naturally, as the proportion of these inputs increases, the production cost (in most cases) would be expected to decrease. Similarly, the production cost in this model is assumed to be decreasing in the level of the characteristic;[2] however, the aforementioned notion of quality in this model allows a higher-quality product to be more expensive or less expensive to produce than a lower-quality product, depending on consumer demographics. The model assumes that while the firm is unable to identify the preferences of individual consumers, the firm does know the relevant demographics of the consumer base.

In short, the paper develops a model of a monopolistic firm with private information about a

[2]While production costs are assumed to be decreasing in the level of the characteristic for convenience, the model can be reformulated with production costs that are increasing in the characteristic with similar results.

characteristic of its product (over which anonymous consumers have heterogeneous preferences) that has ability to signal via price and faces the possibility of consumers choosing to incur an audit cost to acquire this information before purchase. In addition, the firm will face production costs that are decreasing in the level of the product's characteristic. As a result of the structure, the firm must choose whether to signal via price, or to leave the decision (and burden) of becoming informed to the consumer.

The resulting equilibria of the model are analyzed with and without the possibility of consumer verification. When consumer verification is not possible, it is found that there exists a unique Perfect Bayesian equilibrium satisfying the Intuitive Criterion Cho and Kreps (1987) in which the level of the product's characteristic is revealed through price signaling. As previously discussed, production cost is decreasing in the value of the characteristic, and a higher value of the characteristic may result in a higher-quality or lower-quality product depending upon consumer demographics. When lower levels of the characteristic (and thus higher production costs) result in a higher-quality product, the value of the characteristic is signaled with an upward-distortion in price. When higher levels of the characteristic (and thus lower production costs) result in a higher-quality product, the value of the characteristic is signaled with an downward-distortion in price.

When the possibility of consumer verification is added to the model, so that after learning the product's price a consumer can choose to incur an audit cost in order to learn the value of the product's characteristic, there exist conditions on the parameters of the model such that there will exist a second equilibrium (in addition to the separating equilibrium above) wherein the monopolistic firm charges the same price regardless of the value of the characteristic, and all consumers choose to verify the value of the characteristic. That is, the firm does not reveal the value of the characteristic through signaling, but shifts the burden of becoming informed onto consumers. Welfare analysis of the two equilibria show that either equilibrium may be more efficient, depending on the values of the parameters. It is concluded that in order to maximize social welfare, any efforts by a social planner to make product information more accessible to consumers should be selective in the products they target, as lowering the cost of consumer audit can actually result in a less efficient equilibrium in some cases.

II.1.1 Outline

Section 2 of this paper constructs a full information model of a monopolistic firm and heterogeneous consumers as described above. Section 3 analyzes the model under asymmetric information where price signaling can occur, and Section 4 adds the possibility of consumer audit to the model with asymmetric information. Section 5 discusses policy implications of the model, and Section 6 concludes the paper.

II.1.2 Related Literature

There is a substantial literature devoted to analyzing the ability of a firm to signal quality via price.[3] Bagwell and Riordan (1991) construct a two-type model in which production costs are greater for the higher quality type, and find that the higher quality type will distort price upward in order to separate in equilibrium. The model presented here is similar as there are two types for the firm, however, these types are not differentiated by quality, but rather by the level of a specific characteristic of the product. Additionally, there are two types of consumer in the model of this paper, and the varying preferences and proportions of each consumer type are what determines the firm type that is viewed as the "high-quality" or "strong" type. Furthermore, though production costs in the model presented here are decreasing in the level of the product's characteristic, these costs may be greater or less for the high-quality type (as previously discussed). Since it is the high-quality firm type that chooses to distort its price to separate, the price distortion by this type may be upward or downward, depending upon consumer demographics.

A more similar model (in terms of production cost/quality relationship and the direction of the equilibrium price distortion) is that of Daughety and Reinganum (1995), wherein a monopolistic firm that is one of a continuum of types faces full unit costs that are the sum of pure production costs and liability costs. Though the cost of production is assumed to be increasing in quality[4] similar to Bagwell and Riordan's model, the liability costs depend on the quality of the product,

[3]In addition to the papers discussed here, see Bagwell (1992), Daughety and Reinganum (1997, 2005), and Ellingsen (1997) for monopoly models of price signaling; see Daughety and Reinganum (2007, 2008b), and Jansen and Roy (2010) for oligopoly models of price signaling.

[4]Daughety and Reinganum interpret quality as safety, with the measure of safety being the probability of product failure that may or may not injure the consumer.

as well as the liability system, therefore allowing the full unit cost to be increasing or decreasing in quality depending on the firm's expected liability for injuries caused by its product. Daughety and Reinganum find that a higher quality firm distorts price in order to separate, and that the distortion is upward when the full unit cost is increasing in quality and downward when full unit cost is decreasing in quality. While these results are similar to those found in the analogous cases of this paper (those without consumer audit), there are significant differences in the models that will be discussed shortly. The model that is closest in structure to that presented here is Daughety and Reinganum (2008a), wherein a monopolistic firm that is one of two types, with a cost structure similar to Daughety and Reinganum (1995), may signal its type via price or may incur some cost to credibly disclose its type to consumers. The model presented here is similar to Daughety and Reinganum (1995 and 2008a) in the quality/production cost structure and in using quadratic consumer utility; however, there are important differences. Daughety and Reinganum do not allow for consumers to be heterogeneous in how they view quality, and in addition, Daughety and Reinganum's model does not consider the possibility of consumers paying an audit cost to become fully informed about the product.

A recent subset of the literature analyzes the interaction between price signaling and direct disclosure to the consumer, assuming the firm has the ability to do either (or both).[5] For example, Fishman and Hagerty (2003) and Daughety and Reinganum (2008b) construct models in which a firm can use signaling or disclosure (or both) to inform consumers about product quality. The possibility of disclosure is analyzed in the model presented here; however, in the interest of brevity these results are reported very simply at the end of Section 3 in order to maintain the focus on consumer audit.

Other recent papers, such as Bar-Isaac, Caruana, and Cuñat (2011, 2010) construct models with the possibility of consumer audit, but like the aforementioned signaling models, there are important differences between these recent audit models and the model presented here. In Bar-Isaac, Caruana, and Cuñat (2011), the authors analyze a model in which a monopolist sells a product with two characteristics at an exogenously determined price; one of these characteristics is fixed at

[5]For early models of costless disclosure and subsequent "unraveling," see Grossman (1981) and Milgrom (1981). For an early model of costly disclosure see Jovanovic (1982). Though there is a substantial literature on disclosure of information, the main focus of this model is price signaling and thus a thorough review of this literature is omitted for brevity. For a recent review of both the theoretical and empirical literature on disclosure see Dranove and Jin (2010).

high quality, while the other has quality that is stochastically affected by the firm's chosen level of investment in quality. Though consumers in Bar-Isaac et al. (2011) are heterogeneous in which of the two characteristics they view as being more important, all consumers agree on what constitutes high quality, and are able to incur a cost (if they so choose) to evaluate the quality of one of the two characteristics. In Bar-Isaac, Caruana, and Cuñat (2010), the authors analyze a model in which the value of a monopolist's product to individual consumers is stochastically affected by her level investment in the product; this investment increases the probability of a "good" match over a "bad" match for each consumer. Consumers are also heterogeneous in their base taste for the product, and can incur a cost (if they so choose) to evaluate whether the product is good or bad match for them individually. The firm chooses the aforementioned investment, the price of the product, and a "marketing strategy" that determines the cost to the consumer of evaluating the product's quality. Though the model of this paper has a similar structure to the above papers in that a monopolist is selling a product for which she has asymmetric information, and consumers can incur a cost to learn that information, there are important differences. First, consumers in the model presented here are not only heterogeneous in their base taste for the product, consumers also have opposing views on what represents high quality (some prefer a high level of the characteristic, while others prefer a low value of the characteristic). This is clearly different from Bar-Isaac et al. (2011), but is also fundamentally different from Bar-Isaac et al. (2010). To see this, note that while the product in Bar-Isaac et al. (2010) may be a "bad" match for some consumers and a "good" match for others, it is also the case that an increase in the level of investment makes the product more likely to be a good match for all consumers. In the model of this paper, the characteristic in question affects consumer types in opposing ways, thus any change in the characteristic makes the product less attractive to one type of consumer (the product cannot be a "good match" for both consumer types). Another important difference is that the production costs in both Bar-Isaac et al. (2010) and Bar-Isaac et al. (2011) are assumed to be the same across firm types, while production costs in the model of this paper are decreasing in the value of the characteristic. Most importantly, neither paper by Bar-Isaac, Caruana, and Cuñat consider price signaling as a means for the firm to convey its type to consumers.

To my knowledge, there is only one other model, Bester and Ritzberger (2001), that combines price signaling with the possibility of consumer audit. In this paper Bester and Ritzberger analyze a market in which a monopolist sells a product that is either high or low quality. Bester and Ritzberger assume that consumers are homogeneous in their preferences for the good (that is, all consumers agree on the type that represents higher quality), but are heterogeneous in their outside options. The differences in consumers' outside options yields downward-sloping demand for the monopolist's product, depending upon consumers' beliefs about the product's quality. Bester and Ritzberger also assume that production costs are the same for both the high- and low-quality product, and as production costs are assumed to be type-independent, the firm's profits depend solely on consumer beliefs about quality and not directly on the firm's type. Therefore, in the base model without consumer verification the usual equilibria exist (separating, pure-pooling, partial-pooling), but the Intuitive Criterion does not refine the set of equilibria (as any price deviation is equally attractive to both of the firm's types). These same features hold in the model with consumer verification, and so Bester and Ritzberger propose a "modified" version of the Intuitive Criterion that eliminates the pure pooling equilibrium (wherein both firm types charge the same price and some fraction of consumers engage in audit), and depending on the cost of consumer verification, sometimes eliminates the separating equilibrium. The result is that for a sufficiently small cost of consumer verification, the refinement selects an equilibrium that involves partial pooling in prices and a fraction of consumers choosing to engage in audit.

By contrast, my base model is more similar to the standard models of price signaling discussed earlier in this literature review, wherein production costs differ across firm types, firm profits depend directly on the firm's type as well as on consumer beliefs, and wherein refinement by the Intuitive Criterion selects a (unique) separating equilibrium. The model presented here also differs from that of Bester and Ritzberger in that consumer heterogeneity results from disagreement about which product type is preferable; some consumers prefer one type, while other consumers prefer the other type. The possibility of consumer audit is then added to the base model, and it is shown that a separating equilibrium (analogous to that of the base model) survives refinement by the Intuitive Criterion; however, a new equilibrium also survives refinement, wherein both of the firm's types charge the same price and consumers engage in audit to verify the firm's type. Though this second "single-price" equilibrium[6] survives refinement by the Intuitive Criterion, all hybrid (partial-pooling) equilibria are eliminated via refinement by the Intuitive Criterion. As there are

[6] The term "single-price" equilibrium is used over the typical "pooling" equilibrium terminology for reasons discussed in Section 4.

cases in which both the single-price equilibrium and the separating equilibrium exist (after refinement), the welfare properties of these two equilibria are compared and the results used to assess the social benefit of decreasing the cost of consumer audit.

II.2 The Model

II.2.1 The Firm

Consider a monopolistic risk-neutral firm which has developed a new product. The firm is assumed to be one of two types, denoted by $\tau \in \{H, L\}$ (high or low). These types differ only in the value of a specific characteristic of the firm's product, denoted by $X_\tau \in \{X_H, X_L\}$. For simplicity, the value of X_L is assumed to be zero, with $X_H > 0$. The marginal cost of production for the firm's high type is assumed to be zero, while the marginal cost of production for the low type is $c > 0$; a simple and intuitive application of such a structure is the aforementioned example of a product with some proportion of foreign produced inputs. If this proportion is the characteristic in question, it is reasonable to assume (in most cases) that a firm must pay more to produce a product with a lower proportion of foreign inputs. It is assumed that nature determines the firm's type,[7] with the probability of the firm being the high type denoted by λ ($\lambda = \Pr(\tau = H)$) where $\lambda \in (0,1)$. The firm sets a price to maximize profits, and produces to fill demand.

II.2.2 Consumers

All consumers are risk-neutral and may purchase multiple units of the good. Consumers have heterogeneous preferences in two ways. First, each consumer has some level of "taste" for the product, given by θ. The value of θ is uniformly distributed over $[\underline{\theta}, \underline{\theta}+1]$, and thus $\underline{\theta}$ represents the minimum value for a consumer's taste. This continuum of consumers is assumed to have a mass equal to one. In addition to this, consumers are of two types in the way they perceive the level of X. Some consumers are "X-seekers" (type S) while others are "X-avoiders" (type A). A consumer's

[7] Though the probability of each type is assumed to be exogenous for simplicity, the model could be extended to include a preliminary stage in which the firm engages in some activity to affect the probability of each type, as in Daughety and Reinganum (1995).

type, denoted by $T \in \{S,A\}$, is determined by nature and is independent of the consumer's taste. The probability of a consumer being type A is given by α ($\alpha = \Pr(T = A)$) where $\alpha \in (0,1)$. Each consumer is therefore characterized by the pair (θ, T), and this pair is assumed to be the consumer's private information. The consumer's type determines how the level of X_τ affects his or her utility. That is, a greater X_τ decreases the utility of X-avoider consumers by some amount, but increases (or has no effect on) the utility of X-seeker consumers. To model this, the scalar γ_T is used to convert units of X into positive (or negative) utils, with $\gamma_S \geq 0$, and $\gamma_A < 0$. The structure can again be demonstrated using the example of a product with some proportion of foreign inputs. While all consumers have different levels of taste for the product, θ, they also differ in the effect that the proportion of foreign inputs has on their utility. Some consumers (X-seekers) may not care about where the content of the product is produced ($\gamma_S = 0$), or may prefer the product's content be produced elsewhere ($\gamma_S > 0$), while others (X-avoiders) dislike products with a higher proportion of foreign inputs ($\gamma_A < 0$) and will therefore, *ceteris paribus*, purchase less of a product if the proportion of foreign inputs is larger.

All of the above characteristics can now be captured by the following utility function of a type T consumer with a taste for the good given by θ, purchasing q units of a type τ product at price p:

$$U(q,p,\tau,\theta,T) \quad = \quad (\theta + \gamma_T X_\tau)q - .5q^2 - pq. \tag{II.1}$$

Note what equation (II.1) implies about the effects of the level of X on consumer utility.[8] A product with $X_\tau = X_L$ will result in both types of consumers receiving similar utility for the same quantity since $X_L = 0$. If, however, $X_\tau = X_H$, then X-seekers may have much higher utility than X-avoiders for consuming the same quantity of the product. Naturally, this will induce X-seekers and X-avoiders to demand different quantities of the same product (in general).

In order to simplify the analysis it is assumed that every consumer (except possibly those on the margin with the lowest possible level of taste, $\underline{\theta}$) will choose to purchase a positive amount of the firm's product at equilibrium prices. Under full information, this is achieved by the following assumption:

[8]This representation of utility is the reduced form of a quasilinear consumer utility function where a composite of all other goods is the numeraire and the price of the numeraire is normalized to be 1. Income is omitted from the equation for notational convenience, as it has no effect on the results.

A.1 $\underline{\theta} \geq \max[.5 - 2\gamma_A X_H + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_H, .5 + c]$.

The consumer is assumed to be fully informed in this section of the paper. As a result, the consumer bases the purchasing decision directly on the firm's type. Maximizing equation (II.1) over q yields the following first order condition:

$$(\theta + \gamma_T X_\tau) - q - p = 0.$$

Therefore, given the firm type τ and price p, the demand of a type T consumer with taste level θ will be:[9]

$$q(p, \tau, \theta, T) = \theta + \gamma_T X_\tau - p.$$

Since α is the probability of a consumer being type A, and α is independent of the consumer's taste, the expected demand of a consumer with taste θ is:

$$\begin{aligned} q(p, \tau, \theta, \cdot) &\equiv \alpha(\theta + \gamma_A X_\tau - p) + (1-\alpha)(\theta + \gamma_S X_\tau - p) \\ &= \theta + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_\tau - p. \end{aligned}$$

Note the constant $(\alpha\gamma_A + (1-\alpha)\gamma_S)$ in this equation (as well as in assumption A.1). This is a weighted average of the γ's that converts X_τ into utils, and in essence reflects the average consumer's preference over the characteristic X. If this value is positive, the proportion of X-seekers and X-avoiders, weighted by the corresponding γ's, yields higher expected demand for products with a higher level of X. For notational simplicity, henceforth we will use ω to represent this value, that is:

$$\omega = (\alpha\gamma_A + (1-\alpha)\gamma_S).$$

Finally, by integrating over all possible taste levels, and thus aggregating over all possible tastes, the

[9] Though demand is actually given by $q(p, \tau, \theta, T) = \max[\theta + \gamma_T X_\tau - p, 0]$, the maximization is omitted given assumption A.1.

aggregate expected demand for a type τ product at price p is given by:

$$\begin{aligned} Q(p,\tau) &\equiv \int_{\underline{\theta}}^{\underline{\theta}+1} (\theta + \omega X_\tau - p) d\theta \\ &= .5 + \underline{\theta} + \omega X_\tau - p. \end{aligned}$$

The model uses a single-period market, where interaction occurs over three stages: Nature chooses the firm's type and each consumer's type, the firm observes its type and sets a price p, and each consumer then decides the quantity of the product he or she will purchase. As noted above, the firm produces to meet demand, and profits are realized at the end of the last stage.

II.2.3 Full-Information Equilibria

Given the expected demand for each of the firm's types, the full-information expected profit functions are defined as follows:

$$\begin{aligned} \Pi(p,H) &= pQ(p,H) = p(.5 + \underline{\theta} + \omega X_H - p). \\ \Pi(p,L) &= (p-c)Q(p,L) = (p-c)(.5 + \underline{\theta} - p). \end{aligned}$$

Maximizing these expressions yields the following full information prices:

$$\begin{aligned} P_H^{full} &= \frac{.5 + \underline{\theta} + \omega X_H}{2}, \\ P_L^{full} &= \frac{.5 + \underline{\theta} + c}{2}. \end{aligned}$$

Full-information expected profits are then given by:

$$\begin{aligned} \Pi(p_H^{full},H) = \Pi_H^{full} &= \left[\frac{.5 + \underline{\theta} + \omega X_H}{2}\right]^2, \\ \Pi(p_L^{full},L) = \Pi_L^{full} &= \left[\frac{.5 + \underline{\theta} - c}{2}\right]^2. \end{aligned}$$

Recall that c denotes the production cost for the low-type firm, and that ωX_H denotes the average increase (or decrease) in consumer utility due to a high level of X. As these quantities capture the

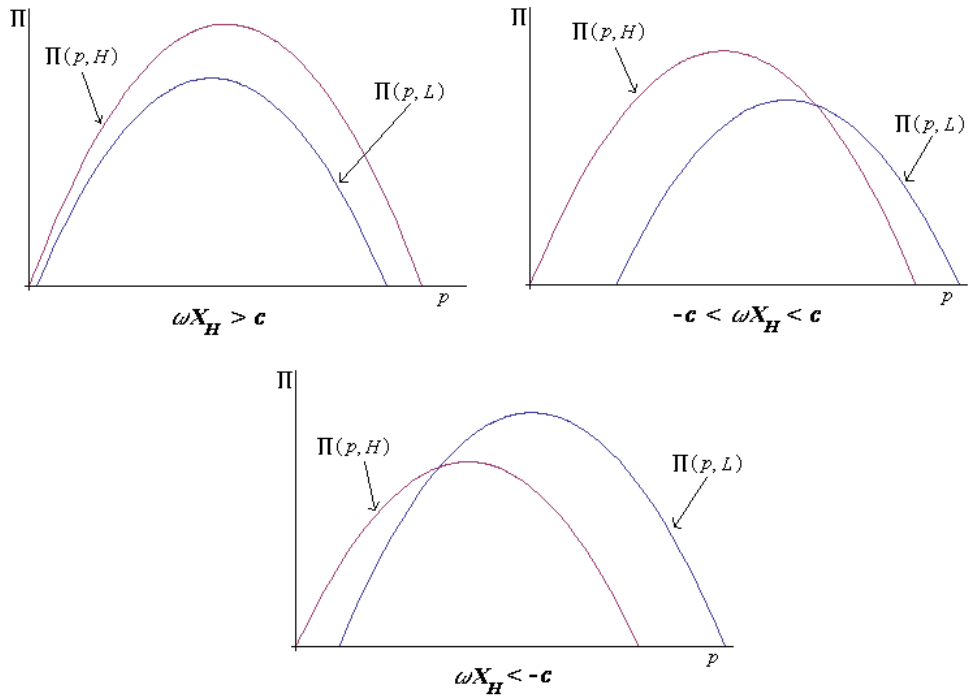

Figure II.1: Comparison of low-type and high-type profit curves

effects of the differences between the firm types, the ordinal relationship between c and ωX_H will determine which firm type will have the greater full-information profits and full-information price; there are three possibilities, illustrated by Figure 1. First, when $\omega X_H > c$, the full information price will be higher for the firm's high type ($p_H^{full} > p_L^{full}$), and profits for the firm's high type will also be higher ($\Pi_H^{full} > \Pi_L^{full}$). The top left diagram illustrates this case. The second possibility is that $-c < \omega X_H < c$, in which case ($p_H^{full} < p_L^{full}$) and ($\Pi_H^{full} > \Pi_L^{full}$), illustrated by the top right diagram. Finally, if $\omega X_H < -c$, then ($p_H^{full} < p_L^{full}$) and ($\Pi_H^{full} < \Pi_L^{full}$), illustrated in the lower diagram above. Note that it is impossible for ($p_H^{full} > p_L^{full}$) and ($\Pi_H^{full} < \Pi_L^{full}$).

II.3 Asymmetric Information

Assume that the game (the payoffs, extensive form, and priors) is common knowledge, but the firm's type (τ) and each consumer's taste/type pair (θ, T) is private information. Also assume that the firm cannot credibly disclose its type. Therefore, the only information available to the consumer about the firm's type is that which can be inferred from the product's price.

As a result, consumers will base their purchasing decisions on their beliefs about the product's type given the product's price, $b(p)$, where b represents the consumer's perceived probability that the firm is the high type ($b(p) \equiv \Pr\{\tau = H|p\}$), and firms will rely on the aggregate demographic information about consumers given by α. Note that b will sometimes be used in place of $b(p)$ for notational convenience, and is intended to denote $b(p)$ unless stated otherwise.

Given that price signaling by the firm can result in upward (or downward) distortions in price, the following adjustment is made to assumption A.1 for cases when $\omega < 0$ in order to guarantee that every consumer (except possibly those with the lowest possible taste level) will choose to purchase a positive amount of the product at equilibrium prices.

$$\text{A.1}' \quad \underline{\theta} \geq \begin{cases} \max[.5 - 2\gamma_A X_H + \omega X_H, .5 + c, .5 + z_1] & \text{if } \omega < 0, \\ \max[.5 - 2\gamma_A X_H + \omega X_H, .5 + c] & \text{if } \omega \geq 0, \end{cases}$$

where $z_1 \equiv \sqrt{-2\omega X_H(.5 + \underline{\theta}) - (\omega X_H)^2}$. Note that z_1 is real when $\omega < 0$ by the original assumption A.1 (It will be shown in the following analysis that z_1 characterizes the upward distortion of the firm's low type when it is optimal for the low type to distort its price to separate).

II.3.1 Demand and Firm Profits

Given consumer beliefs, $b(p)$, together with the two possible firm types H and L, the expected utility of a type T consumer with taste for the good given by θ, purchasing q units of uncertain product type at price p will be:[10]

$$\begin{aligned} E[U(q,p,\tau,\theta,T)|b(p)] = b(p)[(\theta + \gamma_T X_H)q - .5q^2 - pq] + \\ (1 - b(p))[(\theta + \gamma_T X_L)q - .5q^2 - pq] \\ = (\theta + \gamma_T b(p) X_H)q - .5q^2 - pq. \end{aligned}$$

Where the latter expression employs the normalization that $X_L = 0$. Maximizing this expected utility

[10] Note that income has again been omitted from the expression for convenience.

over q yields the following first order condition:

$$(\theta + \gamma_T b(p) X_H) - q - p = 0.$$

Note that the firm's true type is not found in the above expression, as consumers base their purchasing decision on their beliefs about the firm's type given the product's price. As a result, the notation for quantity in the asymmetric-information case will be changed from that of the full-information case by replacing the argument τ (the firm's type) with $b(p)$ (consumer beliefs about the firm's type). Therefore, given beliefs $b(p)$ and price p, the demand of a type T consumer with taste level θ will be:

$$q(p, b(p), \theta, T) = (\theta + \gamma_T b(p) X_H - p).$$

By taking the expectation of demand over consumer types and integrating (aggregating) this over all possible consumer taste levels, the aggregate expected demand for a product of price p under consumer beliefs $b(p)$ is given by:

$$\begin{aligned} Q(p, b(p)) &\equiv \int_{\underline{\theta}}^{\underline{\theta}+1} (\alpha(\theta + \gamma_A b(p) X_H - p) + (1-\alpha)(\theta + \gamma_S b(p) X_H - p)) d\theta \\ &= .5 + \underline{\theta} + \omega b(p) X_H - p. \end{aligned} \tag{II.2}$$

Note that for an arbitrary value of b, $\partial q / \partial b = \omega X_H$, that is, the effect of beliefs on aggregate demand will hinge on the sign of ω.

Unlike quantity and aggregate quantity derived above, the profits for each firm type will depend on both consumer beliefs ($b(p)$) and the firm's type (τ). Therefore, the notation for the profit functions in the asymmetric-information case will be changed from that of the full-information case by adding beliefs as an argument. As a result, the profit functions in the asymmetric-information case are defined as follows:

$$\Pi(p, H, b(p)) \equiv pQ(p, b(p)) = p(.5 + \underline{\theta} + \omega b(p) X_H - p).$$

$$\Pi(p, L, b(p)) \equiv (p-c)Q(p, b(p)) = (p-c)(.5 + \underline{\theta} + \omega b(p) X_H - p).$$

II.3.2 Separating Equilibria

To begin, consider the following characterization of a separating equilibrium in the context of this model.

Def$^{\mathbf{n}}$. *A price pair (p_H^{sep}, p_L^{sep}), together with consumer beliefs $b(p)$, constitutes a separating equilibrium if $p_H^{sep} \neq p_L^{sep}$ and the following incentive compatibility constraints are met:*

$$\Pi(p_H^{sep}, H, 1) \geq \max_{p \geq 0} \ \Pi(p, H, b(p)) \tag{II.3}$$

$$\Pi(p_L^{sep}, L, 0) \geq \max_{p \geq c} \ \Pi(p, L, b(p)) \tag{II.4}$$

$$b(p_H^{sep}) = 1 \tag{II.5}$$

$$b(p_L^{sep}) = 0. \tag{II.6}$$

That is, each of the firm's types must choose its price to maximize profits given beliefs, and beliefs must be correct at the equilibrium prices. It will be useful to define the following type- and belief-specific profits, as these functions provide bounds for the true profit functions that use arbitrary consumer beliefs:

$$\Pi(p, H, 1) = p(.5 + \underline{\theta} + \omega X_H - p) \tag{II.7}$$

$$\Pi(p, H, 0) = p(.5 + \underline{\theta} - p) \tag{II.8}$$

$$\Pi(p, L, 0) = (p - c)(.5 + \underline{\theta} - p) \tag{II.9}$$

$$\Pi(p, L, 1) = (p - c)(.5 + \underline{\theta} + \omega X_H - p). \tag{II.10}$$

Function (II.7) above is simply the "full information" profit function for the high type; that is, the profit function for the firm when consumers correctly identify it as the high type. This will be called the high type's full information profit function. Function (II.8) above will be called the high type's "mimicking profit function;" that is, this function provides the profits for the high type firm, at price p, when the firm is incorrectly believed to be the low type. Similarly, function (II.9) is the low type's full-information profit function, while function (II.10) is the low type's mimicking profit function. Figure 2 shows the full-information and mimicking profit curves for both types for

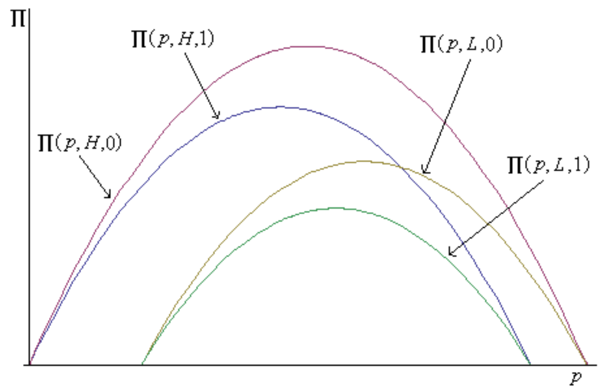

Figure II.2: Profit functions with $-c < \omega X_H < c$

a particular parameter vector wherein $-c < \omega X_H < c$.

The following lemma is immediate from the inspection of equation (II.2) and will be useful in working through the subsequent cases.

Lemma 1. *If all other parameter values are held constant, the values of* $\Pi(p,H,b)$ *and* $\Pi(p,L,b)$ *are strictly increasing in b if* $\omega > 0$*, constant in b if* $\omega = 0$*, and are strictly decreasing in b if* $\omega < 0$*, at any given price p. Thus, (i) if* $\omega < 0$ *the firm's high type will be tempted to mimic the low type, while the low type will wish to be identified, (ii) if* $\omega = 0$ *then beliefs have no effect on profits and neither of the firm's types will be tempted to mimic the other, and (iii) if* $\omega > 0$*, the firm's low type will be tempted to mimic the high type, while the high type will wish to be identified.*

In other words, the type that will be the high-quality or strong type depends on ω; if $\omega < 0$ then the low type is strong while if $\omega > 0$ then the high type is strong.

II.3.2.1 Case 1: The Low Type is Strong ($\omega < 0$)

Recall that $\omega < 0$ implies that a higher value of X, on average, has a negative effect on consumer demand. By Lemma 1, if p' yields positive profits for the firm's low type and $0 < b(p') \leq 1$, then the low type's profits will be less than if the low type is identified ($b = 0$) at p'; that is, $\Pi(p',L,0) > \Pi(p',L,b(p'))$ for all $b(p') \in (0,1]$. Conversely, if p' yields positive profits for the firm's high type and $0 \leq b(p') < 1$, the high type's profits will be greater than if the high type is identified ($b = 1$) at

p' ; that is, $\Pi(p',H,1) > \Pi(p',H,b(p'))$ for all $b(p') \in [0,1)$. As a separating equilibrium requires $b(p_H^{sep}) = 1$, and p_H^{full} maximizes the strictly concave function $\Pi(p,H,1)$, Lemma 1 implies that $p_H^{sep} = p_H^{full}$ in any such equilibrium, otherwise the high type would have incentive to deviate to p_H^{full} regardless of the value of $b(p_H^{full})$, yielding the following Lemma.

Lemma 2. *If $\omega < 0$, then in any separating equilibrium it must be the case that $p_H^{sep} = p_H^{full}$.*

An implication of this lemma is that when $\omega < 0$, supporting beliefs for a separating equilibrium must be constructed in such a way that $b(p_H^{full}) = 1$ and $\Pi(p_H^{full},H,1) \geq \Pi(p,H,b(p))$ for all $p \geq 0$. What remains to be shown is that there exists a set of beliefs, $b(p)$, and a price for the firm's low type, p_L, that satisfy this requirement as well as conditions (II.4) and (II.6). Furthermore, these beliefs must survive refinement by the Intuitive Criterion. The following proposition is proved in the appendix:

Proposition 1. *Let $z_1 \equiv \sqrt{-2\omega X_H(.5+\underline{\theta}) - (\omega X_H)^2}$. For $\omega < 0$, the unique (refined) separating equilibrium is $p_H^{sep} = p_H^{full}$, and*

i) $p_L^{sep} = p_L^{full}$ *if* $z_1 \leq c$;

ii) $p_L^{sep} = p_L^{full} + \left(\frac{z_1 - c}{2}\right)$ *if* $z_1 > c$.

Supporting beliefs for both cases are given by $b(p) = 1$ when $p < p_L^{sep}$ and $b(p) = 0$ otherwise.

Figure 3 illustrates both types of separating equilibria described in Proposition 1. The figure on the left has parameters such that both firm types are induced to charge the full-information price in equilibrium (as in Proposition 1, part (*i*)), while the figure on the right illustrates a case in which the low type must distort upward from the full-information price in order to prevent mimicry by the high type (as in Proposition 1, part (*ii*)). Note that in any case for which the low type must distort its price, it will be an upward distortion.

II.3.2.2 Case 2: The Trivial Case ($\omega = 0$)

If $\omega = 0$, then the full-information profit function is identical to the mimic profit function for both types of firm. Therefore, there is no incentive for either firm to mimic and this case reduces to the full-information case.

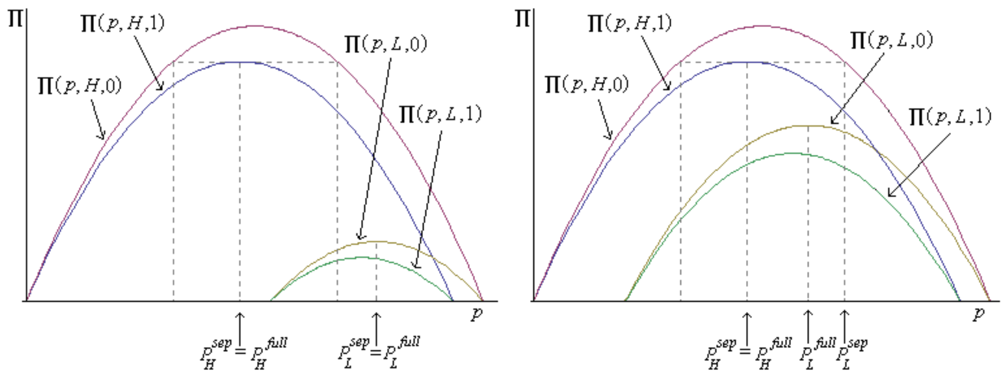

Figure II.3: Separating equilibia when the low type is strong.

Proposition 2. *For $\omega = 0$ there exists a unique separating equilibrium in which $p_H^{sep} = p_H^{full}$ and $p_L^{sep} = p_L^{full}$.*

II.3.2.3 Case 3: The High Type is Strong ($\omega > 0$)

Assuming $\omega > 0$, the firm's high type prefers to be identified, while the firm's low type is tempted to mimic the firm's high type. The following Lemma results from an argument analogous to that preceding Lemma 2.

Lemma 3. *If $\omega > 0$, then in any separating equilibrium it must be the case that $p_L^{sep} = p_L^{full}$.*

By Lemma 3, any supporting beliefs for a separating equilibrium must be constructed in such a way that $b(p_L^{full}) = 0$ and $\Pi(p_L^{full}, L, 1) \geq \Pi(p, L, b(p))$ for all $p \geq 0$. What remains is to show that there exists a set of beliefs, $b(p)$, and a price for the firm's high type, p_H, that satisfy this requirement as well as conditions (II.3) and (II.5). The proof of the following proposition is analogous to that of Proposition 1:

Proposition 3. *Let $z_2 \equiv \sqrt{2\omega X_H(.5 + \underline{\theta} - c) + (\omega X_H)^2}$. For $\omega > 0$, the unique (refined) separating equilibrium is $p_L^{sep} = p_L^{full}$, and*

i) $p_H^{sep} = p_H^{full}$ *if* $z_2 \leq c$;

ii) $p_H^{sep} = p_H^{full} - \left(\frac{z_2 - c}{2}\right)$ *if* $z_2 > c$.

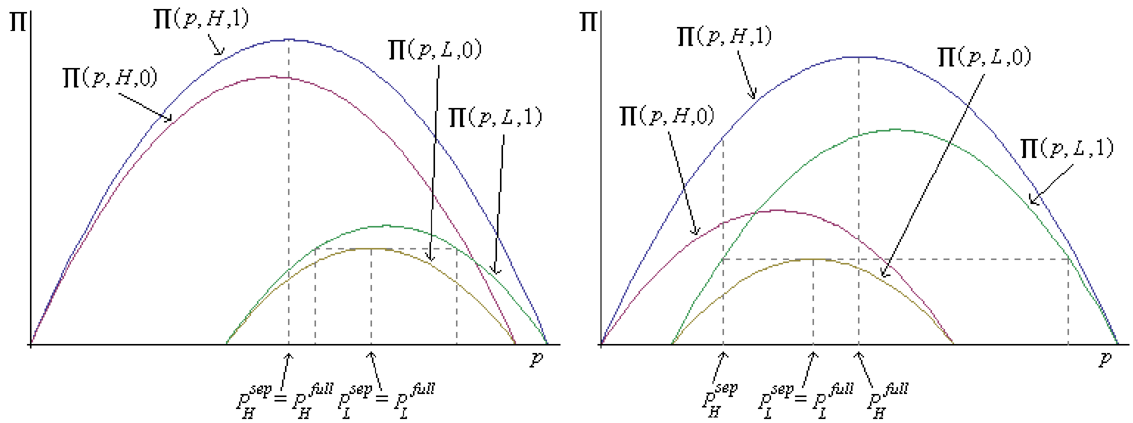

Figure II.4: Separating equilibia when the high type is strong.

Supporting beliefs for both cases are given by $b(p) = 1$ *when* $p \leq p_H^{sep}$ *and* $b(p) = 0$ *otherwise.*[11]

Figure 4 illustrates both types of separating equilibria described in Proposition 3. The figure on the left has parameters such that both firm types are induced to charge the full-information price in equilibrium (as in Proposition 3, part (i)), while the figure on the right illustrates a case in which the high type must distort downward from the full-information price in order to prevent mimicry by the low type (as in Proposition 3, part (ii)). Note that in any case for which the high type must distort its price, it will be a downward distortion.

II.3.3 Pooling and Partial-Pooling Equilibria

In a pooling equilibrium each type chooses the same price; that is, $p_H^{pool} = p_L^{pool}$. In a partial-pooling equilibrium at least one of the firm's types plays a mixed pricing strategy wherein some (but not all) information about the firm's type is passed to consumers via equilibrium prices. For example, one possibility would be the strong type charging price p_S in equilibrium, while the weak type mixes between charging p_S and another price $p_W \neq p_S$ in equilibrium. In this case, consumers would be certain the firm is the weak type if p_W is the observed price, but would be uncertain of

[11] Note that while it is possible that $p_H^{full} \geq p_L^{full}$, these beliefs will still support the separating equilibrium as p_H^{sep} is distorted downward below p_L^{full} whenever $p_H^{full} \geq p_L^{full}$.

the firm's type if p_S is observed.[12] While there are many other ways a partial-pooling equilibrium could occur, the characteristic of importance to the proposition below is that in any partial-pooling equilibrium at least one equilibrium price is a pooling price, and thus there is still ambiguity about the firm's type at any such price. As is often the case, the Intuitive Criterion eliminates all pooling equilibria and partial-pooling equilibria in the asymmetric-information model of this section; this is proved in the appendix.

Proposition 4. *There exist no pooling or partial-pooling equilibria satisfying the Intuitive Criterion in the asymmetric-information case.*

II.3.4 Disclosure

As mentioned in the literature review of this paper, there are recent models that combine price signaling with costly disclosure as two methods by which a firm can communicate its quality to consumers, for example, Daughety and Reinganum (2008a, 2008c). Although the focus of this paper is costly consumer audit, the base model of this section can easily be extended to allow disclosure; however, since the subsequent results are in line with the aforementioned papers, a rigorous treatment of disclosure is foregone for the sake of brevity. With that said, the basic structure and results are presented below.

If a firm is allowed to disclose its type directly to consumers at cost k, the advantage of disclosure is that it allows the firm to charge its full-information price and earn full-information profits. As some separating equilibria of the model result in full-information prices being charged by both firms, it is easy to see that neither firm has incentive to incur k to disclose in such cases; the strong type could not gain more profit through disclosure (and thus would not incur k), and the only way for the weak type to increase profits would be through mimicry (convincing consumers it is the strong type) which cannot be facilitated by disclosure (and thus the weak type would not incur k). In other cases of the model, the strong type must distort its price in order to separate, and in these cases the strong type will only have incentive to incur k and disclose if the loss in profits due to distortion (from those of full information) is greater than k. Thus if k is large enough, neither type

[12] Note that the true probability of a given type conditional on p_S being observed would depend upon λ and the mixing strategy of the weak type.

has incentive to disclose, but for small enough k the strong type will choose to incur k and charge its full information price. To see that the weak type will not have incentive to mimic the strong type in such a case, note that as the strong type has incentive to disclose consumers would interpret the lack of disclosure as proof that the firm is the weak type, and so the weak type will fail in any attempt to mimic the strong type; that is, the weak type will have no incentive to deviate from its full information price.

In summary, for large k the equilibria of the model with disclosure are the same as those separating equilibria of the model without disclosure; while for small enough k, the strong type will incur k to disclose and charge its full-information price and the weak type will still charge its full information price.

II.4 Consumer Audit

In this section, it is assumed that consumers have access to information about the firm's type; however, there is a one-time audit cost incurred by the consumer to learn this information, denoted by a. Additionally, assume that the firm cannot credibly disclose this information directly to the consumer. As a result, after learning the price of the product, the consumer may choose to pay the audit cost and access the information before making a purchase, or may choose not to access the information and use his or her posterior beliefs about the product's type based only upon its price. In the aforementioned example of a product with some proportion of foreign inputs, this could be analogous to product information available on the internet, where the audit cost a is the cost to a consumer of seeking out, reading, and verifying this information. Other examples of costly sources of information include consumer publications that require a purchase (e.g. Consumer Reports), and other information available online in the form of expert reviews, that require some cost of effort. Even information made available by the firm could be considered in the consumer audit framework of this section, as a (possibly mandated) disclosure by the firm could be in a format that requires a significant cost of effort on the part of the consumer in order to understand the disclosure (e.g., a booklet containing large amounts of technical product information). Since consumers may still glean information from the price of the good, price signaling by the firm is still possible. The same cost structure as in Section 3 is used in this model for simplicity; that is, the cost of production for the high type is assumed to be zero while the cost of production for the low type is assumed to be $c > 0$.

II.4.1 The Consumer's Decision

The only change from the model of Section 3 without disclosure is the cost faced by consumers, $a > 0$, incurred only if the consumer chooses to verify the product's type. As this may result in some consumers incurring a while others do not, this raises the possibility of free-riding. To avoid this, it is assumed that consumers cannot use the observed purchasing decisions of others to deduce the firm's type without incurring a cost that is greater than the audit cost a; that is, it would cost more (in effort) to deduce firm type by observing others than to simply incur a and verify it directly.

Another important issue is how consumers use beliefs in this model. In addition to being used in the calculation of the optimal purchasing decision, consumer beliefs are now used in making the decision as to whether or not to incur the audit cost a. The firm sets the price first, and the consumer observes this price before making the decision to audit. At any observed price, the consumer will only choose to incur the audit cost if beliefs indicate a positive probability for both firm types. If beliefs are 1 or 0 at the observed price, then no consumer will choose to incur a as all consumers are already "sure" of the firm's type at that price. This decision is captured by the function $\delta^T(p,b(p))$, where $\delta^T(p,b(p)) = 1$ if a type T consumer chooses to audit at price p, and $\delta^T(p,b(p)) = 0$ otherwise.[13] Now consider the determinants of the consumer's audit decision. A type T consumer with taste θ that chooses not to audit will base their purchasing decision upon beliefs, thus the utility maximizing quantity will be $q(p,b(p),\theta,T) = (\theta + \gamma_T b(p)X_H - p)$, as in Section 3.1. Therefore, if $U^{\delta^T}(q,p,\tau,\theta,T)$ is used to denote consumer utility when audit is possible, the expected utility for a consumer who chooses not to audit is given by:[14]

$$E[U^0(q,p,\tau,\theta,T)|b(p)] \equiv \frac{(\theta + \gamma_T b(p)X_H - p)^2}{2}.$$

[13] It will be shown that the consumer's taste level does not affect the auditing decision, and therefore θ is not an argument of the function δ^T.

[14] Note that assumptions $A.1$ and $A.1'$ are again sufficient allow the omission of $\max[\theta + \gamma_T b(p)X_H - p, 0]$ in the derivation of consumer demand in this section; that is, these assumptions guarantee that $\theta + \gamma_T b(p)X_H - p \geq 0$ for all consumers types and all $b(p) \in [0,1]$.

The same consumer who incurs a to know the firm's type will choose $q(p,1,\theta,T) = \theta + \gamma_T X_H - p$ when the firm is the high type and $q(p,0,\theta,T)) = \theta - p$ when the firm is the low type. Thus the expected utility for this consumer is:

$$\begin{aligned}
E[U^1(q,p,\tau,\theta,T)|b(p)] &\equiv b(p)(\frac{(\theta+\gamma_T X_H - p)^2}{2} - a) + (1-b(p))(\frac{(\theta-p)^2}{2} - a) \\
&= \frac{(\theta+\gamma_T b(p) X_H - p)^2}{2} + \frac{(1-b(p))b(p)\,(X_H)^2\,\gamma_T^2}{2} - a \\
&= E[U^0(q,p,\tau,\theta,T)|b(p)] + \frac{(1-b(p))b(p)\,(X_H)^2\,\gamma_T^2}{2} - a.
\end{aligned}$$

Therefore, since the consumer makes the audit decision by comparing these two expected utilities, the following Lemma results.

Lemma 4. *At any price p, a consumer of type T with beliefs $b(p)$ will choose to incur the cost a in order to know the firm's type if and only if:*

$$a < \frac{b(p)(1-b(p))\,(X_H)^2\,(\gamma_T)^2}{2}.$$

Note that the audit decision does not depend on the consumer's base level of taste, θ, as the contribution to utility from non-X-related taste does not depend upon the firm's type. In addition, the benefit for auditing depends upon the beliefs at the given price, but not the price itself. If the beliefs are close to 0 or 1 (the consumer is fairly sure the firm is a particular type), there is less benefit from auditing (or, alternatively, a lower cost of basing the purchasing decision on beliefs), and therefore the consumer will choose not to audit. This leads to the following characterization[15] of $\delta^T(p,b(p))$:

$$\delta^T(p,b(p)) = \begin{cases} 1 & \text{if } b(p) \in (0,1) \text{ and } a < b(p)(1-b(p))(X_H)^2(\gamma_T)^2 \\ 0 & \text{otherwise.} \end{cases}$$

[15] For simplicity, it is assumed that consumers in the model play pure strategies, and that every consumer of a given type will make the same audit decision. It can be shown that the equilibria of the model with pure auditing strategies are also equilibria of the model with the possibility of consumers playing mixed auditing strategies, and moreover, additional equilibria are only possible if $\lambda(1-\lambda)$ is exactly equal to $\frac{2a}{(X_H)^2(\gamma_T)^2}$.

II.4.2 Separating Equilibria

As in the model of Section 3 (without consumer audit), a separating equilibrium consists of equilibrium prices for both of the firm's types, p_L^{sep} and p_H^{sep}, together with consumer beliefs, $b(p)$, such that $p_L^{sep} \neq p_H^{sep}$, and no party wishes to deviate from the equilibrium. As the consumer knows the price before making the audit decision, and the supporting beliefs for the separating equilibria of Section 3.2 take values of only 0 or 1 (including out-of-equilibrium beliefs), there will be no consumer audit and these same beliefs support the same separating equilibria in the current model. Therefore, the unique (refined) separating equilibrium for each case in Section 3.2 is also a separating equilibrium for the corresponding case when the consumer has the option to audit. In addition to this, note that Lemma 2 and Lemma 3 will still hold, as consumers will never choose to audit in a separating equilibrium; thus the weak firm type will be fully identified and will have incentive to deviate from any price other than the full-information price. Furthermore, there exists no separating equilibrium in which the strong firm type chooses a less-distorted price, as consumers do not audit in a separating equilibrium and the weak firm type would then have incentive to deviate. Finally, all separating equilibria in which the strong firm type chooses a more distorted price will not satisfy the Intuitive Criterion; this due to the fact that the Intuitive Criterion compares profits where beliefs equal 1 or 0, and is thereby unaffected by consumer audit. Therefore, the only separating equilibrium that is possible when consumers have the choice of auditing occurs at the same prices as the unique (for each case) separating equilibrium from Section 3.2, yielding the following result.

Proposition 5. *The unique (refined) separating equilibrium for each case of the model without consumer audit is also the unique (refined) separating equilibrium for the corresponding case of the model with consumer audit.*

II.4.3 Single Price Equilibria

Although the unique separating equilibrium from the base model without consumer audit is still an equilibrium of the model with the possibility of consumer audit, there is different type of equilibrium, analogous to the pooling equilibria discussed in Section 3, that must be considered.

Def$^{\mathbf{n}}$. *A "single-price" equilibrium consists of an equilibrium price pair,* (p_L^{SP}, p_H^{SP})*, together with consumer beliefs,* $b(p)$*, and the audit decision function for each consumer type,* $\delta^A(p, b(p))$ *and* $\delta^S(p, b(p))$*, such that* $p_L^{SP} = p^{SP} = p_H^{SP}$*,* $b(p^{SP}) = \lambda$ *, and both expected profits and expected utility are maximized.*

While this definition may seem identical to that of the classical pooling equilibrium because both firm types charge the same price, there is an important difference in the consumer's information at the time of purchase. In a classic pooling equilibrium, the consumer would use prior beliefs (the same beliefs as before price was observed) about the firm's type in order to make a purchasing decision. In this model, however, some or all of the consumers may choose to become fully informed via audit before the purchase is made, and therefore the resulting equilibrium is distinct from the classic pooling equilibrium. While all of the included single-price equilibria are pooling equilibria in the sub-game up to the consumer's audit decision (where all consumers are still using prior beliefs), those consumers that choose to audit learn the firm's type and revise their beliefs accordingly, and thus the only equilibria that are identical to classic pooling equilibria are those in which no consumer chooses to audit; in that case, the consumer's posterior beliefs, $b(p)$, will be identical to prior beliefs as no information about the firm s type is learned via price or audit.

Since a single-price equilibrium with price p^{SP} necessarily requires $b(p^{SP}) = \lambda$ (the true proportion of high type firms), Lemma 4 results in four possible cases (using $\widehat{\lambda} = .5\lambda(1-\lambda)X_H^2$). First, if $a > \widehat{\lambda}\gamma_S^2$ and $a > \widehat{\lambda}\gamma_A^2$, then $\delta^A(p^{SP}, \lambda) = 0 = \delta^S(p^{SP}, \lambda)$; that is, no consumers will choose to verify the firm's type by auditing at the single price. Second, if $a \leq \widehat{\lambda}\gamma_S^2$ and $a \leq \widehat{\lambda}\gamma_A^2$, then $\delta^A(p^{SP}, \lambda) = 1 = \delta^S(p^{SP}, \lambda)$; that is, all consumers will choose to verify the firm's type by auditing at the single price. If $\widehat{\lambda}\gamma_A^2 < a \leq \widehat{\lambda}\gamma_S^2$, then $\delta^A(p^{SP}, \lambda) = 0$ and $\delta^S(p^{SP}, \lambda) = 1$; that is, only X-seeker consumers will choose to verify the firm's type at the shared price. Finally, if $\widehat{\lambda}\gamma_S^2 < a \leq \widehat{\lambda}\gamma_A^2$, then $\delta^A(p^{SP}, \lambda) = 1$ and $\delta^S(p^{SP}, \lambda) = 0$; that is, only X-avoider consumers will choose to verify the firm's type at the shared price. The intuition behind this result is fairly straightforward. Refer again to the example of a product with foreign inputs, and suppose that γ_A is very large (and negative), implying that the proportion of foreign inputs has a substantial effect on the utility of X-avoider consumers (less foreign inputs yield higher utility). The loss due to purchasing too much of a high-foreign-input product is therefore very high, and thus X-avoiders would choose to incur a. Similarly, if γ_S is very large, X-seekers would choose to incur a rather than facing the possibility of purchasing the incorrect amount of the product. Intermediate values of γ_S and γ_A result in lower losses from basing the purchasing decision on beliefs alone, and thus in these cases consumers choose not to incur a.

While $b(p) = \lambda$ must be true at the single price in equilibrium, it is important to note that this is not the case for out-of-equilibrium beliefs. Out-of-equilibrium beliefs may not only differ from λ, but may also result in a different combination of consumers (if any) choosing to audit at a given out-of-equilibrium price.[16]

Clearly, consumer demand and subsequent firm profits will depend upon which consumers (if any) choose to audit. Furthermore, when at least one type of consumer chooses to audit at a given price, consumer demand at that price will depend upon the product's type, τ. As a result, individual consumer demand will now be denoted by $q^{\delta^T}(p,\tau,b(p),\theta,T)$, where $\delta^T(p,b(p))$ denotes the audit decision for type T consumers at price p given beliefs $b(p)$. Following the derivations of demand in Sections 2.2 and 3.1, a type T consumer with base taste level θ will make the audit decision $\delta^T(p,b(p)$ at price p, and if the consumer's taste level is given by θ, the consumer's demand for a type τ product will be denoted by:

$$q^{\delta^T}(p,\tau,b(p),\theta,T) = \begin{cases} \theta + \gamma_T b(p) X_H - p & \text{if } \delta^T = 0, \\ \theta + \gamma_T X_\tau - p & \text{if } \delta^T = 1, \end{cases}$$

By taking the expectation of demand over consumer types and integrating (aggregating) this over all possible consumer taste levels, the aggregate expected demand for a type τ product of price p under posterior consumer beliefs $b(p)$, with audit decisions $\delta^A(p,b(p))$ and $\delta^S(p,b(p))$ is given by:

[16] Note that Lemma 4 implies that out-of-equilibrium beliefs can cannot "switch" from having only one type of consumer auditing to having only the other type of consumer auditing; for example, if only X-seekers audit in the equilibrium, then out-of-equilibrium beliefs cannot induce only X-avoiders to audit at out-of-equilibrium prices, as the former requires $\gamma_A < \gamma_S$ while the latter implies $\gamma_A > \gamma_S$.

$$
Q^{\delta_A\delta_S}(p,\tau,b(p)) = \begin{cases} .5+\underline{\theta}+\omega b(p)X_H - p & \text{if } \delta^A = 0 = \delta^S, \\ \alpha(.5+\underline{\theta}+\gamma_A b(p)X_H - p) & \\ \qquad +(1-\alpha)(.5+\underline{\theta}+\gamma_S X_\tau - p) & \text{if } \delta^A = 0 \text{ and } \delta^S = 1, \\ \alpha(.5+\underline{\theta}+\gamma_A X_\tau - p) & \\ \qquad +(1-\alpha)(.5+\underline{\theta}+\gamma_S b(p)X_H - p) & \text{if } \delta^A = 1 \text{ and } \delta^S = 0, \\ .5+\underline{\theta}+\omega X_\tau - p & \text{if } \delta^A = 1 = \delta^S. \end{cases}
$$

Given this, the expected profits for each type at price p given *ex-ante* consumer beliefs $b(p)$ and the consumer audit decisions $\delta^A(p,b(p))$ and $\delta^S(p,b(p))$ are denoted by:

$$
\Pi^{\delta^A\delta^S}(p,H,b(p)) = pQ^{\delta_A\delta_S}(p,H,b(p)) \tag{II.11}
$$

$$
\Pi^{\delta^A\delta^S}(p,L,b(p)) = (p-c)Q^{\delta_A\delta_S}(p,L,b(p)). \tag{II.12}
$$

II.4.3.1 Case 1: No Consumers Audit

This section analyzes the single price equilibria in which both types of consumer will choose not to incur the cost a to determine the firm's type at the single price. This case arises when:

$$
a > \widehat{\lambda}\gamma_S^2 \quad \text{and} \quad a > \widehat{\lambda}\gamma_A^2.
$$

As mentioned previously, this results in the single-price equilibria of this subsection being identical to classic pooling equilibria, and therefore these equilibria are called full pooling equilibria. Since no consumers choose to audit in a full pooling equilibrium, all consumers base their purchasing decisions on their beliefs at the full pooling price p^{pool} (where these beliefs are equal to λ). As no additional information is learned by consumers, any full pooling price would be identical to the pooling prices in the model of asymmetric information without audit and without disclosure discussed in Section 3.3. Since employing the Intuitive Criterion involves comparing profits when beliefs are equal to 0 or 1 (beliefs for which consumers never choose to audit), adding the possibility of audit does not affect whether a pooling equilibrium will satisfy the Intuitive Criterion. Therefore, these equilibria will fail refinement as in Section 3.3. This results in the following proposition.

Proposition 6. *There exists no full pooling equilibrium (a single-price equilibrium in which no consumers choose to audit) that satisfies the Intuitive Criterion.*

Intuitively, this proposition results from audit being too costly from the perspective of the consumer. As a result, the firm is unable to shift the cost of informing consumers onto the consumers themselves. Therefore, the firm type that wishes to be identified must use price signaling in order to reveal its type, and only the separating equilibrium survives refinement.

II.4.3.2 Case 2: All Consumers Audit

This section analyzes the single price equilibria in which both types of consumer will choose to incur the audit cost a to determine the firm's type at the single price. This case arises when:

$$a \leq \widehat{\lambda}\gamma_S^2 \quad \text{and} \quad a \leq \widehat{\lambda}\gamma_A^2.$$

As a result, all consumers choose to audit and the firm faces fully-informed consumers at any single-price equilibrium. To distinguish from other single-price equilibria, these equilibria will be called single-price-full-audit (SPFA) equilibria. Suppose $\omega < 0$; then the firm's low type is strong and always prefers to be thought of as a low type at any given price. Furthermore, since the high type is identified through audit by all consumers in an SPFA equilibrium, $p^{SPFA} = p_H^{full}$ maximizes the profits for the high type over all SPFA equilibria. Given this, suppose (in addition to $\omega < 0$) that there exists an SPFA equilibrium such that $p^{SPFA} \neq p_H^{full}$. In this case it can easily be shown that there will always exist a price, p', between p^{SPFA} and p_H^{full} (an ε deviation) such that if $b(p') = 1$, then the high type would wish to deviate to p' while the low type would not. Thus refinement by the Intuitive Criterion implies $b(p') = 1$, and no such SPFA equilibrium exists; therefore, it must be that if $\omega < 0$, then $p^{SPFA} = p_H^{full}$ in any SPFA equilibrium. If $\omega > 0$ (the low type is the weak type), a similar argument requires $p^{SPFA} = p_L^{full}$. Therefore, the weak type, depending on the sign of ω, must make full-information profits in order for an SPFA equilibrium to exist; this leads to the following lemma.[17]

[17]When $\omega = 0$ there is no incentive for either type to mimic, and therefore no single-price equilibria will exist because one firm would always have incentive to deviate to its full-information price

Lemma 5. *Suppose there exists an SPFA equilibrium price in which both consumer types choose to audit. If $\omega < 0$, then any $p \neq p_H^{full}$ cannot be an SPFA equilibrium price. If $\omega > 0$, then any $p \neq p_L^{full}$ cannot be a SPFA equilibrium price.*

As a result of the above, there can be at most a single SPFA equilibrium. The following proposition provides conditions under which a unique SPFA equilibrium will exist, and is proved in the appendix.

Proposition 7. *Suppose that $a \leq \widehat{\lambda}\gamma_S^2$ and $a \leq \widehat{\lambda}\gamma_A^2$. For $\omega < 0$, if $2c - \omega X_H < z_1$, then there exists a unique single-price-full-audit equilibrium that satisfies the Intuitive Criterion wherein $p^{SPFA} = p_H^{full}$ with supporting beliefs given by $b(p) = \lambda$ if $p = p_H^{full}$ and $b(p) = 1$ otherwise. For $\omega > 0$, if $\max\{c, 2(c - \omega X_H)\} < z_2$, there exists a unique SPFA equilibrium under the Intuitive Criterion in which $p^{SPFA} = p_L^{full}$ with supporting beliefs given by $b(p) = \lambda$ if $p = p_L^{full}$ and $b(p) = 0$ otherwise.*

Note that the conditions for existence in each case of Proposition 9 result in non-empty sets. To see this, first consider the conditions involving the auditing decision ($a \leq \widehat{\lambda}\gamma_S^2, a \leq \widehat{\lambda}\gamma_A^2$). Both of these constraints allow ω to take on any value as there are no other restrictions on α (the proportion of X-avoiders) and a other than $a > 0$ and $\alpha \in (0,1)$. As a result, there are few constraints on the values of the parameters in the condition

$$2c - \omega X_H < z_1 = \sqrt{-2\omega X_H(.5 + \underline{\theta}) - (\omega X_H)^2}. \tag{II.13}$$

To see that the set of parameters meeting these conditions is nonempty, note that $c > 0$ allows c to be sufficiently small to satisfy condition (II.13) since z_1 is independent of c, and is greater than ωX_H when $\omega < 0$ (by A.1′). Thus for any set of values for the other parameters (meeting the assumptions of the model) there will exist a range of values for c, albeit possibly very close to zero, for which this condition will be met and such an equilibrium exists. Note also that a c close to zero is not necessary for the SPFA equilibrium to exist, only sufficient to guarantee existence (a large value for $\underline{\theta}$ can achieve the same effect). For the condition

$$\max\{c, 2(c - \omega X_H)\} < z_2 = \sqrt{2\omega X_H(.5 + \underline{\theta} - c) + (\omega X_H)^2}, \tag{II.14}$$

it is again clear that for any set of parameter values meeting the assumptions of the model, with

$\omega > 0$, there will exist a value for $c > 0$ small enough to meet condition (II.14). Note again that this is not necessary, but sufficient for the condition to be met (a large value for $\underline{\theta}$ can achieve the same effect). Intuitively, both of these conditions ((II.13) and (II.14)) require the distortion in the separating equilibrium to be "large" relative to the difference between the profits for the firm's two types when they are fully identified (a function of c and ωX_H), thereby guaranteeing the loss in profits due to the price distortion in the separating equilibrium will be larger than the loss in profits due to the strong firm type charging the full-information price of the weak firm type. Therefore, these conditions eliminate the incentive to deviate from the SPFA equilibrium and allow beliefs to satisfy the Intuitive Criterion.

II.4.3.3 Case 3: Only X-Avoiders Audit

This case analyzes the single price equilibria in which only the X-avoider consumers choose to incur the cost a to determine the firm's type at the single price. These equilibria will be called single-price-avoider-audit (SPAA) equilibria, and may arise when:

$$a > \widehat{\lambda}\gamma_S^2 \quad \text{and} \quad a \leq \widehat{\lambda}\gamma_A^2.$$

Therefore, X-avoiders will maximize their utility based on the true firm type, while X-seekers will base their purchasing decision upon beliefs. The following proposition results from refinement by the Intuitive Criterion, and is proved in the technical appendix for this paper.

Proposition 8. *There exists no single-price avoider-audit equilibrium (an equilibrium in which both of the firm's types charge the same price and only X-avoiders choose to audit) that satisfies the Intuitive Criterion.*

While this result may appear counterintuitive at first glance, the key to understanding why these equilibria fail refinement lies in the relative costs/benefits of avoider-only audit to each of the firm's types. In this case the cost of audit is small enough for X-avoiders to bear the burden of becoming informed; however, it is not small enough for X-seekers to become informed. When the firm passes the burden of becoming informed on to consumers, only X-avoiders choose to audit, and a single-price equilibrium benefits the low type and hurts the high type. The X-avoiders fully identify the firm's type, and will therefore purchase the same quantities that would be purchased under full information (at a given price). X-seekers, however, are unsure of the firm's type at a single-price equilibrium and must therefore base their purchasing decision on posterior beliefs ($b(p)$) about the firm's type. This causes X-seekers to purchase a larger quantity of the low-type product and a smaller quantity of the high-type product than would be purchased under full information (at a given price). As a result, the high type, regardless of the proportion of X-avoiders versus X-seekers (given by α) and their relative preferences over X (given by γ_A and γ_S), would prefer to be identified. The Intuitive Criterion simply assumes that X-seekers understand that their lack of information negatively affects the high type (X-seeker's "preferred" type) and positively affects the low type, so that beliefs must reflect a greater willingness for the high type to deviate than for the low type to deviate. Although this does not change the fact that for any price p in which $b(p) = \lambda$, X-avoiders have incentive to audit while X-seekers do not, it does result in the high type having incentive to deviate from the SPAA equilibrium price given any beliefs that satisfy the Intuitive Criterion. Therefore all possible beliefs that support SPAA equilibria fail refinement, and the firm is forced to bear the burden of informing the consumer through price separation.

II.4.3.4 Case 4: Only X-Seekers Incur a

This case analyzes the single price equilibria in which only X-seeker consumers choose to incur the cost a to determine the firm's type at the single price. These equilibria will be called single-price-seeker-audit (SPSA) equilibria, and may arise when:

$$a \leq \widehat{\lambda}\gamma_S^2 \quad \text{and} \quad a > \widehat{\lambda}\gamma_A^2.$$

Therefore, X-seekers will maximize their utility based on the true firm type, while X-avoiders will base their purchasing decision upon beliefs. The following proposition results from refinement by the Intuitive Criterion, and has a proof analogous to that of Proposition 10.

Proposition 9. *There exists no single-price seeker-audit equilibrium (an equilibrium in which both of the firm's types charge the same price and only X-seekers choose to audit) that satisfies the Intuitive Criterion.*

As in the previous case, the above result reflects the disparity between the benefits to each of the firm's types from a single-price equilibrium. Here the cost of audit is small enough for X-seekers to bear the burden of becoming informed; however, it is not small enough for X-avoiders to become informed. The firm is unable to pass the burden of becoming informed on to consumers, as the low type is adversely affected and will have sufficient incentive to deviate from the single-price equilibrium given reasonable beliefs (via the Intuitive Criterion). Therefore, the firm bears the cost of informing consumers through price separation.

II.4.3.5 Hybrid Equilibria

Similar to the definition of the partial-pooling equilibria in the asymmetric information case without consumer audit, hybrid equilibria in this case are a broad class of equilibria in which at least one of the firm's types plays a mixed pricing strategy, some information about the firm's type is conveyed to consumers, and at least one of the equilibrium prices has a positive probability of being charged by both types. As any equilibrium requires that both of the firm's types maximize profits given consumers' beliefs and audit decision functions, mixed strategies require that profits at all prices over which a firm type mixes must be equal. Given this, together with the analogous nature between hybrid equilibrium prices charged by both firm types and the single-price equilibria of this section, all hybrid equilibria[18] fail refinement by the Intuitive Criterion; this is proved in the appendix.

Proposition 10. *There exists no hybrid equilibrium that satisfies the Intuitive Criterion.*

II.4.4 Welfare and Policy Implications

The main result of Section 4.3 is that the possibility of consumer audit does not affect the existence of the unique separating equilibria of the model without audit; however, a single-price equilibrium could possibly arise if both consumer types are willing to incur the audit cost to become fully informed. This leads to the natural question of which equilibrium is best from a welfare standpoint when both equilibria are possible, and what, if anything, can be done by a central planner to arrive at the better of the two equilibria. As the value of ω determines which of the firm's types is the strong (high-quality) type, this analysis will be broken into two cases.

[18] There is a degenerate case in which the strong type can mix over its (distorted) separating price and the weak type's full-information price, however, this requires a specific (deterministic) relationship between the parameters (e.g. if $\omega < 0$ it must be that $c = -\frac{\omega X_H(.5+\theta+\omega X_H)}{z_1-\omega X_H}$), and is therefore not considered a general result of the model.

II.4.4.1 Case 1: The Low Type is Strong ($\omega < 0$)

When $\omega < 0$, the price charged by the high type is p_H^{full} in both the unique separating equilibrium and in the unique *SPFA* equilibrium. As a result, the difference in welfare outcomes when the firm is the high type is only the loss in consumer surplus due to the full mass of consumers (equal to 1) incurring the audit cost a. Thus, if CS_H^{sep}, CS_H^{SPFA}, PS_H^{sep}, PS_H^{SPFA}, TS_H^{sep}, and TS_H^{SPFA}, represent the consumer surplus, producer surplus, and total surplus for both types of equilibria when the firm is the high type, the following is true.

$$CS_H^{Sep} - CS_H^{SPFA} = a;$$
$$PS_H^{Sep} - PS_H^{SPFA} = 0;$$
$$TS_H^{Sep} - TS_H^{SPFA} = a.$$

Now consider the low type. If the low type firm does not need to distort in the separating equilibrium, a *SPFA* equilibrium will not exist, therefore the welfare results that follow assume that the low type charges a distorted (upward) price in the separating equilibrium. The difference in each surplus is as follows:

$$\begin{aligned} CS_L^{Sep} - CS_L^{SPFA} &= .5(.5+\underline{\theta} - (p_L^{full} + \frac{z_1 - c}{2}))^2 - (.5(.5+\underline{\theta} - p_H^{full})^2 - a); \\ PS_L^{Sep} - PS_L^{SPFA} &= ((p_L^{full} + \frac{z_1 - c}{2}) - c)(.5+\underline{\theta} - (p_L^{full} + \frac{z_1 - c}{2})) \\ &\quad - (p_H^{full} - c)(.5+\underline{\theta} - p_H^{full}). \end{aligned}$$

Plugging in the values for p_H^{full} and p_L^{full} and simplifying yields:

$$\begin{aligned} CS_L^{Sep} - CS_L^{SPFA} &= a - .25(z_1 - \omega X_H)(.5+\underline{\theta} - \frac{z_1 + \omega X_H}{2}); \\ PS_L^{Sep} - PS_L^{SPFA} &= .5(z_1 - \omega X_H)(c - \frac{z_1 + \omega X_H}{2}); \\ TS_L^{Sep} - TS_L^{SPFA} &= a - .25(z_1 - \omega X_H)(.5+\underline{\theta} - 2c + \frac{z_1 + \omega X_H}{2}). \end{aligned}$$

It can easily be shown that $2(.5+\underline{\theta}) - \omega X_H > z_1$, and thus the difference in consumer surplus is negative whenever the audit cost a is small. The difference in producer surplus is always negative as profits for the low type are required to be greater in the SPFA equilibrium than in the separating equilibrium in order for the SPFA equilibrium to exist (and condition (II.13) guarantees this). As a result, when the firm is the low type and $\omega < 0$, the total surplus for the SPFA equilibrium may be higher or lower than that of the separating equilibrium, depending on the value of a relative to the gain in surplus due to decreasing the price to the SPFA equilibrium price. This makes sense since the firm is a monopoly and the SPFA equilibrium has a lower price (but still above marginal cost) than the separating equilibrium, resulting in higher total surplus if the audit cost is ignored.

II.4.4.2 The High Type is Strong ($\omega > 0$)

When $\omega < 0$, the price charged by the low type is p_L^{full} in both the unique separating equilibrium and in the unique *SPFA* equilibrium. As a result, the difference in welfare outcomes when the firm is the low type is only the loss in consumer surplus due to the full mass of consumers (equal to 1) incurring the audit cost a. Therefore:

$$CS_L^{Sep} - CS_L^{SPFA} = a;$$
$$PS_L^{Sep} - PS_L^{SPFA} = 0;$$
$$TS_L^{Sep} - TS_L^{SPFA} = a.$$

The difference in surplus for when the firm is the high type can be calculated in the same way as for the low type when $\omega < 0$, and simplifies to:

$$CS_H^{Sep} - CS_H^{SPFA} = a + .25(z_2 - \omega X_H)(.5 + \underline{\theta} - c + .5z_2 + 1.5\omega X_H);$$
$$PS_H^{Sep} - PS_H^{SPFA} = .5(z_2 - \omega X_H)(c - \frac{z_2 + \omega X_H}{2});$$
$$TS_H^{Sep} - TS_H^{SPFA} = a + .25(z_2)(.5 + \underline{\theta} + c - .5z_2 + \omega X_H).$$

Note that by condition (II.14), existence of the SPFA equilibrium requires $-.5z_2 < \omega X_H$, and thus it is not possible for the SPFA equilibrium to be more socially efficient than the separating equilibrium. This makes intuitive sense as the SPFA equilibrium will have a higher price, and thus the total surplus with a monopolistic firm having marginal costs of zero (the high type) must be lower under the higher price, despite the greater profits for the high type (required for the existence of the SPFA equilibrium).

II.4.4.3 Results and Policy Implications

If the above differences in welfare (where the firm type is known) are defined as the *ex-post* welfare comparisons, the *ex-ante* expected welfare comparisons can be calculated using λ (the probability of the firm being the high type) and $(1-\lambda)$. This results in the following proposition:

Proposition 11. *If both the unique separating equilibrium and the unique SPFA equilibrium exist:*

i) Ex-post welfare will be at least as high at the unique SPFA equilibrium as the unique separating equilibrium if and only if consumers prefer the low type product on average ($\omega < 0$), the firm is the low type, and $a \leq .25(z_1 - \omega X_H)(.5 + \underline{\theta} - 2c + \frac{z_1 + \omega X_H}{2})$, otherwise welfare will be higher at the unique separating equilibrium.

ii) The ex ante expected welfare will be at least as high at the unique SPFA equilibrium as the unique separating equilibrium if and only if consumers prefer the low type product on average ($\omega < 0$) and $a < (1-\lambda).25(z_1 - \omega X_H)(.5 + \underline{\theta} - 2c + \frac{z_1 + \omega X_H}{2})$, otherwise welfare will be higher at the unique separating equilibrium.

Intuitively, this is because when $\omega < 0$, consumers (on average) prefer the product that costs more to produce, and thus separation via price requires upward distortion. As the SPFA equilibrium has a lower price than that of the separating equilibrium, the SPFA equilibrium will be socially preferred if the gain in surplus is greater than the loss due to the audit cost. When $\omega > 0$, consumers prefer the product that is cheaper to produce, and thus separation via price is achieved through a downward distortion. The SPFA equilibrium price will be above the price in the separating equilibrium, and although the firm may have higher profits in the SPFA equilibrium, consumers are worse off and the total surplus is decreased due to the decreased quantity sold in the SPFA equilibrium.

Now consider possible action by a central planner. If for some reason the central planner knows the firm to be the low type, and ω is known to be negative, there are conditions such that the central planner would prefer the SPFA equilibrium.[19] If the central planner does not yet know the firm's type, but does know that ω is negative, there are still conditions such that the central planner

[19] A central planner may have access to information about a firm's product via mandatory product registration, exogenous government audit, etc.

would prefer the SPFA equilibrium to the unique separating equilibrium (*ex-ante*). When $\omega > 0$, the separating equilibrium is always preferred, regardless of the central planners information about the firm's type. Though a central planner is able to determine which equilibrium is socially preferred, the question remains as to how a central planner could lead the market to result in the preferred equilibrium. Suppose the SPFA equilibrium is both possible and socially preferable. As shown above, it must be that $\omega < 0$, and the SPFA equilibrium can only exist at p_H^{full}. The problem for the central planner is that even if the audit cost a is "small enough," the existence of the SPFA depends upon consumer beliefs; if consumers believe that both firm types will charge p_H^{full}, then the SPFA will result, otherwise the separating equilibrium will result. A seemingly simple solution would be for the central planner to fix the price of the good at p_H^{full}. Provided that consumers are aware of the fixed price, and can adjust their beliefs accordingly (so that $b(p_H^{full}) = \lambda$), fixing the price at p_H^{full} would lead to the same total surplus and profits as the SPFA equilibrium. The case in which the SPFA equilibrium is possible but not preferable is not as simple. If the central planner wishes to guarantee the separating equilibrium is reached, he or she must somehow convince consumers that the strong type will never charge the same price as the weak type. The difficulty of this endeavor is compounded by the fact that the strong type could charge more or less than the weak type in order to separate (depending on the sign of ω). If the central authority knows the firm's type, it may be possible to fix the price at what would have been the separating equilibrium price for the type, however, this could entail fixing a price higher than the SPFA equilibrium price that could have resulted.

Though the central planner may not be willing (or able) to guarantee a separating equilibrium by fixing the price, another possible action would be to adjust the audit cost a. When the separating equilibrium is preferable, an increase in a could eliminate the possibility of the SPFA equilibrium altogether, as the incentive for consumers to audit can be eliminated. Though one might believe that the government is both willing and able to obfuscate information (and possibly adept at doing so), reducing the availability of product information (increasing a) is a difficult policy recommendation to make. What can be said, however, is that when central authorities enact laws and/or programs that seek to provide more information to consumers about products they purchase, social welfare may benefit from such programs being selective in the products they target as per the above. To see this, note that if the audit cost a can be affected by the central planner, when $\omega < 0$ and $z_1 >$

$2c - \omega X_H$ or when $\omega > 0$ and $z_2 > \max\{c, 2(c - \omega X_H)\}$, an SPFA equilibrium can be made possible where it was not possible without intervention. As a result, if a central planner determines that a separating equilibrium is best for a given product (e.g. when the good would be priced lower in a separating equilibrium than in a SPFA equilibrium), the planner could actually harm welfare by making information about the product more easily accessible. Furthermore, if the SPFA equilibrium could exist under lower audit costs, the central planner could increase welfare by lowering a and inducing a SPFA equilibrium. Therefore, this model predicts that if the central planner is selective in the products about which it seeks to make information more accessible, social welfare could be increased.

II.5 Conclusion

The model presented here, wherein a monopolistic firm of two possible types faces a heterogeneous consumer base, is primarily designed to analyze the interaction between price signaling and consumer verification. The price signaling model is first analyzed without the possibility of verification, and is then analyzed with the possibility of disclosure for the sake of completeness. Although the model is constructed differently, the price signaling results are similar to previous results by Daughety and Reinganum (2008a), wherein a unique Perfect Bayesian equilibrium exists in which the higher-quality firm distorts its price to separate, and if the firm with higher production costs produces the high quality product the distortion is upward, while if the lower cost firm produces the high quality product the distortion is downward. The possibility of consumer audit is then added to the model, and this also results in a unique separating equilibrium; however, under certain conditions it is possible that a unique single-price equilibrium will also satisfy the Intuitive Criterion, wherein all consumers choose to pay the audit cost, and both firm types charge the full information price of the weak firm type. As these equilibria exist simultaneously, social welfare at these equilibria are compared, and it is shown that if the strong (high-quality) type is the firm type with higher production costs, and the audit cost is small enough, the single-price equilibrium could result in higher social welfare via higher firm profits and increased consumer surplus. Consequently, implications for a central planner facing the possibility of two equilibria are discussed, and as the results depend upon the availability of product information, it is suggested that in order to maximize social welfare, efforts to make product information more accessible to consumers should be selective in the products they target.

Though the results of the paper are designed to be as complete as possible, there are extensions that could be made to the model. An intuitive extension would be to add the possibility of disclosure to the model of signaling and consumer verification. Although the timing of the events in the model is reasonable (the consumer knows the price before making the audit decision), alternative timing arrangements could also be considered. Finally, as this model is designed around the effects of a single product characteristic produced by a monopolist, it may also be interesting to contrast the results of a similar model with multiple (possibly interdependent) characteristics or multiple competing firms.

II.6 Appendix

II.6.1 Proof of Proposition 1

By Lemma 2, $p_H^{sep} = p_H^{full}$. Then in order for the high type to have no incentive to deviate to the low type's price (p_L^{sep}), it must be that $\Pi(p_H^{full}, H, 1) \geq \Pi(p_L^{sep}, H, 0)$. Solving this inequality for p_L^{sep} yields the following condition:

$$p_L^{sep} \leq \frac{.5+\underline{\theta}-z_1}{2} \quad \text{or} \quad \frac{.5+\underline{\theta}+z_1}{2} \leq p_L^{sep}. \tag{II.15}$$

Where $z_1 = \sqrt{-2\omega X_H(.5+\underline{\theta})-(\omega X_H)^2}$. As $-\omega X_H > 0$, the z_1 term will be real by assumption A.1$'$, and since $c > 0$, it must be the case that $\frac{.5+\underline{\theta}-z_1}{2} < \frac{.5+\underline{\theta}+c}{2} = p_L^{full}$. This leaves two cases to analyze. First, suppose that $\frac{.5+\underline{\theta}+z_1}{2} \leq p_L^{full}$, or equivalently, $z_1 \leq c$. In this case, $p_L^{sep} = p_L^{full}$ will meet condition (II.15), and since $\Pi(p_L^{sep}, H, 0) > \Pi(p, H, 0)$ for all $p > p_L^{sep}$, if beliefs are given by $b(p) = 1$ if $p < p_L^{sep}$ and $b(p) = 0$ otherwise, the full information prices will satisfy the incentive compatibility constraint (II.3). Since p_L^{full} maximizes the low type's profits over all beliefs and all prices when $b(p_L^{full}) = 0$, the incentive compatibility constraint (II.4) is also met; moreover, since the high type has no incentive to deviate to p_L^{full} when $b(p_L^{full}) = 0$, refinement by the Intuitive Criterion eliminates all alternative equilibria in which $p_L^{sep} \neq p_L^{full}$. Therefore, if $z_1 \leq c$, the unique separating equilibrium satisfying the Intuitive Criterion will be $p_H^{sep} = p_H^{full}$ and $p_L^{sep} = p_L^{full}$, with supporting beliefs given by $b(p) = 1$ if $p < p_L^{sep}$ and $b(p) = 0$ otherwise.

For the second case, suppose $z_1 > c$. Then $\frac{.5+\underline{\theta}-z_1}{2} < p_L^{full} < \frac{.5+\underline{\theta}+z_1}{2}$, and the low type must distort from its full information price to separate. Comparing the profits at the cutoffs for p_L^{sep} given by condition (II.15) it can be easily shown that $0 < \Pi\left(\frac{.5+\underline{\theta}-z_1}{2}, L, 0\right) < \Pi\left(\frac{.5+\underline{\theta}+z_1}{2}, L, 0\right)$. Thus by the strict concavity of the profit function for a fixed b, the price $p_L^{full} + \frac{z_1-c}{2} = \frac{.5+\underline{\theta}+z_1}{2}$ maximizes $\Pi(p, L, 0)$ over all p satisfying condition (II.15). As p_L^{sep} must satisfy condition (II.15), and since the high type will have no incentive to deviate to $p' > p_L^{sep}$ when $b(p') = 0$, any beliefs that support an equilibrium price $p_L^{sep} \neq p_L^{full} + \frac{z_1-c}{2}$ will fail to satisfy the Intuitive Criterion (as this will require $b(p_L^{full} + \frac{z_1-c}{2}) = 0$). Again consider the beliefs given by $b(p) = 1$ if $p < p_L^{sep}$ and $b(p) = 0$ otherwise. These beliefs, together with $p_L^{sep} = p_L^{full} + \frac{z_1-c}{2}$ and $p_H^{sep} = p_H^{full}$ will meet the incentive compatibility constraint (II.3), since $\Pi(p, H, 0) < \Pi(p_L^{sep}, H, 0)$ for all $p > p_L^{sep}$. To show that these beliefs and equilibrium prices will meet the incentive compatibility constraint (II.4), note that it is sufficient to show $\Pi\left(p_L^{full} + \frac{z_1-c}{2}, L, 0\right) \geq (\max_{p \geq c}\ \Pi(p, L, 1))$, which can be shown to be the case, and a proof is provided in the technical appendix. Therefore, if $c < z_1$, the unique separating equilibrium satisfying the Intuitive Criterion will be $p_H^{sep} = p_H^{full}$ and $p_L^{sep} = p_L^{full} + \frac{z_1-c}{2}$, with supporting beliefs given by $b(p) = 1$ if $p < p_L^{sep}$ and $b(p) = 0$ otherwise.

II.6.2 Proof of Proposition 4

Suppose that p^{pool} is the price in a pooling equilibrium. As consumer beliefs must be correct at the pooling price, it must be that $b(p^{pool}) = \lambda$, and profits will be given by $\Pi(p^{pool}, H, \lambda) = p^{pool}(\frac{1}{2} + \underline{\theta} + \lambda\omega X_H - p^{pool})$ and $\Pi(p^{pool}, L, \lambda) = (p^{pool} - c)(\frac{1}{2} + \underline{\theta} + \lambda\omega X_H - p^{pool})$. There are now two cases to analyze; the $\omega < 0$ case is shown here, with the $\omega > 0$ being analogous. [20] Given that $\omega < 0$, it must be that the high type is earning profits at least as large as full information profits, otherwise the high type would deviate to its full information price regardless of beliefs.[21]. Therefore, it must be that $c < p^{pool} < \frac{1}{2} + \underline{\theta} + \lambda\omega X_h$. Solving for the larger of the two prices such

[20]When $\omega = 0$ firms always have incentive to charge full information prices

[21] Note that if the pooling price occurs at the full information price of the high type, then the pooling profits for the high type are necessarily higher than the high type's full information profits

that $\Pi(p^{pool}, H, \lambda) = \Pi(p'_H, H, 0)$ yields:

$$p'_H = \frac{\frac{1}{2} + \underline{\theta} + \sqrt{\left(\frac{1}{2} + \underline{\theta}\right)^2 - 4p^{pool}\left(\frac{1}{2} + R + \lambda\omega X_H - p^{pool}\right)}}{2},$$

which can easily be shown to be real as the expression under the square root is a concave quadratic (in p^{pool}) that is positive at its minimum; moreover, it can easily be shown that $\frac{\partial\Pi(p,H,0)}{\partial p}|_{p'_H} < 0$. Therefore, by the strict concavity of the profit functions for a fixed b, if $b(p) = 0$ for any price $p > p'_H$, the resulting profits for the high type at p would be less than the high type's pooling profits at p^{pool}.

Solving for the larger price of the two prices such that $\Pi(p^{pool}, L, \lambda) = \Pi(p'_L, L, 0)$ yields,

$$p'_L = \frac{\frac{1}{2} + \underline{\theta} + c + \sqrt{\left(\frac{1}{2} + \underline{\theta} + c\right)^2 - 4\left(p^{pool}\left(\frac{1}{2} + \underline{\theta} + \lambda\omega X_H + c - p^{pool}\right) - c\lambda\omega X_H\right)}}{2},$$

which can easily be shown to be real as the expression under the square root is a concave quadratic (in p^{pool}) that is positive at its minimum; moreover, it can easily be shown that $\frac{\partial\Pi(p,L,0)}{\partial p}|_{p'_L} < 0$. Therefore, for any price $p < p'_L$, if $b(p) = 0$ the profits for the low type would be greater at p than at the pooling price. What remains is to prove that $p'_H < p'_L$, which can be shown, and a proof is provided in the technical appendix. Therefore there will be a price p between p'_H and p'_L such that if $b(p) = 0$, the profit for the high type will be less than the pooling equilibrium profit, but the profit for the low type will be greater than the pooling equilibrium profit. Thus by the Intuitive Criterion, the beliefs at this price must be 0, and the pooling equilibrium fails refinement. The $\omega > 0$ case is similar, and therefore refinement by the Intuitive Criterion eliminates all pooling equilibria of the model.

To see that there can be no partial-pooling equilibria, suppose that p^{p-pool} is a pooling price that is part of a partial-pooling equilibrium. As there is a positive probability of either type choosing p^{p-pool}, let $\phi \in (0,1)$ denote the true probability that the firm is the high type given the price p^{p-pool} is observed by consumers. As the above proof that all pooling equilibria fail refinement by the Intuitive Criterion was derived using an arbitrary λ (which in the demand/profit equations above is interpreted as the true probability of the firm being the high type given the observed pooling price), the same proof can be used to show that the partial-pooling price p^{p-pool} will fail refinement as well, simply by replacing λ with ϕ in the analysis. As a result, there will always be at least one firm with incentive to deviate from p^{p-pool}, and refinement by the Intuitive Criterion will eliminate all partial-pooling equilibria.

II.6.3 Proof of Proposition 7

Here the case for $\omega < 0$ is shown, with the $\omega > 0$ case being similar and left to the reader. Since consumers must have correct beliefs in a Perfect Bayesian equilibrium, by Lemma (4) it must be that $a \leq \widehat{\lambda}\gamma_S^2$ and $a \leq \widehat{\lambda}\gamma_A^2$ in order for both types of consumer to audit in a single-price equilibrium. By Lemma (15) the only possible SPFA equilibrium price is $p^{SPFA} = p_H^{full}$. The incentive compatibility constraints are then given by:

$$\Pi^{11}(p_H^{full}, H, \lambda) \geq \max_{p \geq 0} \ \Pi^{\delta_A \delta_S}(p, H, b(p)) \tag{II.16}$$

$$\Pi^{11}(p_H^{full}, L, \lambda) \geq \max_{p \geq c} \ \Pi^{\delta_A \delta_S}(p, L, b(p)) \tag{II.17}$$

Note that since the firm's type is known by all consumers in an SPFA equilibrium, the high type earns full information profits. This implies that the prices at which the high type is indifferent between charging $p^{SPFA} = p_H^{full}$ and charging a different price at which $b(p) = 0$ (mimicking the low type) are the same as in condition (II.15). Thus if $b(p') = 0$ for any p' such that $p' \leq \frac{.5+\theta-z_1}{2}$ or $p' \geq \frac{.5+\theta+z_1}{2}$, the high type would have no incentive to deviate to p' from p^{SPFA}.

If $z_1 \leq c$, then $\frac{.5+\theta+z_1}{2} \leq p_L^{full}$,and the high type has no incentive to deviate to p_L^{full} when $b(p_L^{full}) = 0$. Since p_L^{full} maximizes the profits for the low type over all prices where the low type is identified, and the firm is identified in an SPFA equilibrium, refinement of the SPFA equilibrium by the Intuitive Criterion will require $b(p_L^{full}) = 0$ and the SPFA equilibrium will fail. Thus in any SPFA equilibrium it must be that $z_1 > c$.

If $z_1 > c$, then $\frac{.5+\theta-z_1}{2} < p_L^{full} < \frac{.5+\theta+z_1}{2}$, and it is easily shown that $\Pi^{00}(\frac{.5+\theta-z_1}{2}, L, 0) < \Pi^{00}(\frac{.5+\theta+z_1}{2}, L, 0)$. Using $p_L^{sep} = \frac{.5+\theta+z_1}{2} = p_L^{full} + \frac{z_1-c}{2}$, it must be that the SPFA equilibrium profits for the low type are at least as large as those received by the low type when charging p_L^{sep} with

$b(p_L^{sep}) = 0$, or equivalently:

$$\Pi^{11}(p_H^{full}, L, \lambda) \geq \Pi^{00}(p_L^{full} + \frac{z_1 - c}{2}, L, 0); \tag{II.18}$$

otherwise, the Intuitive Criterion would require $b(p_L^{sep}) = 0$ and the low type would have incentive to deviate from p_H^{full}. It is shown in the proof of Proposition 1 that $\Pi(p_L^{sep}, L, 0) \geq \max_{p \geq c} \Pi(p, L, 1)$, which implies $\Pi^{00}(p_L^{sep}, L, 0) > \max_{p \geq c} \Pi^{00}(p, L, 1)$. Therefore, if $z_1 > c$ and condition (II.18) is met, the SPFA equilibrium wherein $p^{SPFA} = p_H^{full}$ with supporting beliefs given by $b(p) = \lambda$ if $p = p_H^{full}$ and $b(p) = 0$ otherwise, will satisfy the incentive compatibility constraint (II.16) and (II.17). Moreover, the strict concavity of the low type's profit function at a fixed b guarantees the low type would have no incentive to deviate to any p for which $b(p) = 0$ that does not satisfy condition (II.15), and while the low type may have incentive to deviate to some price p that does satisfy condition (II.15) if $b(p) = 0$, the high type would also have incentive to deviate to any such p. Thus the SPFA equilibrium wherein $p^{SPFA} = p_H^{full}$ with supporting beliefs given by $b(p) = \lambda$ if $p = p_H^{full}$ and $b(p) = 0$ otherwise will satisfy the Intuitive Criterion. As condition (II.18) reduces to $z_1 > 2c - \omega X_H$ the necessary and sufficient conditions to guarantee the existence of an SPFA equilibrium satisfying the Intuitive Criterion in which $p^{SPFA} = p_H^{full}$ with supporting beliefs given above are $a \leq \widehat{\lambda}\gamma_S^2$, $a \leq \widehat{\lambda}\gamma_A^2$, and $z_1 > \max\{c, 2c - \omega X_H\} = 2c - \omega X_H$. These conditions are shown to result in a non-empty set in the main text.

II.6.4 Proof of Proposition 8

Consider a potential SPAA equilibrium price p^{SPAA}. As only X-avoiders audit in the SPAA equilibrium, this implies that $\gamma_S^2 < \gamma_A^2$, and the audit decisions at out-of-equilibrium beliefs, (δ_A, δ_S), can only take the form $(0,0)$, $(1,0)$, or $(1,1)$, as $(0,1)$ would require $\gamma_S^2 > \gamma_A^2$. Given beliefs $b(p)$ and the audit decisions δ_A and δ_S, the profits for the high type when charging a price other than p^{SPAA} are given by equation (II.11), and the profits for the low type when charging a price other than p^{SPAA} are given by equation (II.12). If $\omega < 0$, this implies that the lowest possible demand that the low type could face at a given price p (over all possible beliefs and audit choices) is given by $Q^{00}(p, L, 1)$, yielding profits $\Pi^{00}(p, L, 1)$. As a result, it must be that $\Pi^{10}(p^{SPAA}, L, \lambda) \geq \max_p \Pi^{00}(p, L, 1)$. Subsequently, since $\max_p \Pi^{00}(p,L,1) \geq \Pi^{00}(p_H^{full},L,1)$, the low type would have no incentive to deviate from p^{SPAA} to p_H^{full}. Therefore, it must be that $\Pi^{10}(p^{SPAA},H,\lambda) \geq \Pi^{00}(p,H,1)$, otherwise the Intuitive Criterion would require $b(p_H^{full}) = 1$ and the high type would deviate from the SPAA price. Finally, (from the proof of Prop 1), since the high type is receiving profits at least as large as those the separating equilibrium, the high type will have no incentive to deviate to p_L^{sep}. As a result, the low type must be earning at least as much profit in the SPAA equilibrium as in the separating equilibrium, as otherwise the Intuitive Criterion would require $b(p_L^{sep}) = 0$, and the low type would deviate from the SPAA price. Now consider if $\omega > 0$. In this case, the lowest possible demand the low type could receive over all possible beliefs and audit decisions is given by $Q^{00}(p,L,0)$, yielding profits $\Pi^{00}(p,L,0)$. Thus the lowest possible profits the low type can receive in the SPAA equilibrium are given by $\max_p \Pi^{00}(p,L,0)$, which are exactly the profits the low type would receive in the separating equilibrium. Moreover, since this implies that the low type has no incentive to deviate to p_H^{sep}, it must be that the high type is making SPAA profits at least as large as in the separating equilibrium, as otherwise the Intuitive Criterion would require $b(p_H^{sep}) = 1$ and the high type would deviate from the SPAA price. Therefore, regardless of the value of ω, it must be that both firm types are earning profits at least as large as those of the corresponding separating equilibrium. As a result, for a SPAA equilibrium to exist, it must be the case that $c < p^{SPAA} < .5 + \underline{\theta} + \alpha\gamma_A X_H + (1-\alpha)\lambda\gamma_S X_H$ so that profits for both firm types are positive (as separating equilibrium profits are positive via assumption A.1$'$)

Having established that SPAA profits are positive for both types, the proof is split into two cases (note that there are no longer restrictions on ω). For the first case suppose that $(1-\alpha)(1-\lambda)\gamma_S \leq -\alpha\gamma_A$, or equivalently, $\Pi^{10}(p,L,\lambda) \geq \Pi^{00}(p,L,1)$ for all prices $p > c$. As this implies that the SPAA profits for the low type at any price $p \geq c$ are at least as large as the profits the low type could earn if believed to be the high type at the same price, it must be that $\Pi^{10}(p^{SPAA},L,\lambda) \geq \Pi^{00}(p^{SPAA},L,1)$. Since avoider audit implies that $\Pi^{10}(p,H,\lambda) < \Pi^{00}(p,H,1)$ for all $p > 0$, there will exist an ε price deviation (upward or downward) such that the low type would have no incentive to deviate, but the high type would have incentive to deviate (a rigorous derivation of this deviation is provided in the technical appendix). As a result, the Intuitive Criterion will require $b(p^{SPAA}+\varepsilon) = 1$ and the high type will deviate from the SPAA price.

Now consider the second case in which $(1-\alpha)(1-\lambda)\gamma_S > -\alpha\gamma_A$, or equivalently, $\Pi^{10}(p,L,\lambda) <$

$\Pi^{00}(p,L,1)$ for all prices p such that $p \geq c$. Solving for the larger of the two prices such that $\Pi^{10}(p^{SPAA},L,\lambda) = \Pi^{00}(p,L,1)$ yields

$$p_L' = .5\Bigg(.5+\underline{\theta}+c+\omega X_H + \sqrt{(.5+\underline{\theta}+c+\omega X_H)^2+4((p^{SPAA})^2 - p^{SPAA}(.5+\underline{\theta}+c+(1-\alpha)\gamma_S\lambda X_H) - c(\omega-(1-\alpha)\gamma_S\lambda)X)}\Bigg),$$

which can easily be shown to be real as the expression under the square root is a concave quadratic (in p^{SPAA}) that is positive at its minimum under the assumption of this case. Moreover, it can easily be shown that $\frac{\partial\Pi^{00}(p,L,1)}{\partial p}|_{p_L'} < 0$. Therefore by the strict concavity of the profit functions for a fixed b, if $b(p) = 1$ for any price $p > p_L'$, the resulting profits for the low type at p would be less than the low type's SPAA profits. Solving for the larger of the two prices such that $\Pi^{10}(p^{SPAA},H,\lambda) = \Pi^{00}(p,H,1)$ yields

$$p_H' = .5\Bigg(.5+\underline{\theta}+\omega X_H + \sqrt{(.5+\underline{\theta}+\omega X_H)^2+4((p^{SPAA})^2 - p^{SPAA}(.5+\underline{\theta}+\omega X_H-(1-\alpha)(1-\lambda)\gamma_S X_H))}\Bigg),$$

which can easily be shown to be real as the expression under the square root is a concave quadratic (in p^{SPAA}) that is positive at its minimum under the assumption of this case. Moreover, it can easily be shown that $\frac{\partial\Pi^{00}(p,H,1)}{\partial p}|_{p_H'} < 0$. Therefore by the strict concavity of the profit functions at a fixed b, if $b(p) = 1$ for any price $p < p_H'$, the resulting profits for the high type at p would be greater than the high type's SPAA profits. What remains to be shown to prove failure of the equilibrium is that $p_L' < p_H'$, which can be shown to be the case, and a proof is provided in the technical appendix to this paper. Therefore, there will always exist a price p between p_L' and p_H' such that if $b(p) = 1$ the high type has incentive to deviate but the low type does not, and any beliefs that support an SPAA equilibria will fail to satisfy the Intuitive Criterion. Furthermore, since the two cases exhaust all possibilities for SPAA equilibria, refinement by the Intuitive Criterion eliminates all such equilibria.

II.6.5 Proof of Proposition 10

To see why such equilibria cannot exist under consumer audit, assume that p^{hybrid} is an equilibrium price that is part of a hybrid equilibrium of the model wherein both types charge p^{hybrid}

with positive probability. Given this, define $\phi \in (0,1)$ as the true probability that the firm is the high type given that the price p^{hybrid} is observed. Therefore, since consumer beliefs must be correct in equilibrium, it must be that $b(p^{hybrid}) = \phi$. Given the consumer audit decision functions, there are now four possibilities to consider. Depending on the parameters of the model, p^{hybrid} could be a full pooling price (wherein no consumers audit), an SPAA price, an SPSA price, or an SPFA price. Regardless of the which of the four possibilities characterizes p^{hybrid}, it is clear that this price is analogous to the single-price equilibria of this section, with the differences being i) that the true probability of the firm being the high type is no longer required to be λ, and ii) there are other prices over which one or both of the firm types are mixing. However, p^{hybrid}, together with ϕ, must still meet the profit maximization condition and survive refinement by the Intuitive Criterion in order for the hybrid equilibrium to survive refinement. Since Propositions 6-9 are proved using an arbitrary value for λ (interpreted in the single-price equilibria as the probability of the firm being the high type given that the single price is observed) these same propositions will apply to p^{hybrid}, simply by replacing λ with the analogous ϕ. Therefore, it cannot be that p^{hybrid} is a true pooling price, an SPAA price or an SPSA price, as any beliefs that support p^{hybrid} would fail refinement by the Intuitive Criterion (by the proofs of Propositions 6,8 and 9). This leaves the possibility that p^{hybrid} is an SPFA equilibrium; however, this would require that p^{hybrid} be equal to the full information price of the weak type (as the weak type is fully identified at p^{hybrid}). To see why this cannot result in a hybrid equilibrium, consider the case where $\omega < 0$ (with the $\omega > 0$ case being analogous). In order for the equilibrium to be a hybrid equilibrium in this case, one or both of the firm's types must be mixing between $p^{hybrid} = p_H^{full}$ and another price. First, suppose that the high type is mixing between p^{hybrid} and another price $p' \neq p^{hybrid}$. Since the high type is earning full information profits $\left(\Pi^{11}(p_H^{full}, H, \phi)\right)$ at p^{hybrid} it must also be earning full information profits at p'; however, it cannot be the case that p' is charged only by the high type, as this implies that the high type is identified and by the strict concavity of the high type's full information profit function $\Pi^{00}(p', H, 1) < \Pi^{00}(p_H^{full}, H, \phi)$. Also, it cannot be the case that both firm types charge p', as this would require p' to be an SPFA price as well (for reasons described above), again implying $\Pi^{11}(p', H, 1) < \Pi^{00}(p_H^{full}, H, 1)$. Thus the high type could not be playing a mixed strategy in equilibrium. Now suppose the low type is mixing between p^{hybrid} and some price $p' \neq p^{hybrid}$. Again, p' cannot be a full pooling price, an SPAA price, or an SPSA price, and as shown above, the high type would never choose to mix between p^{hybrid} and p' if it where an SPFA price. Thus it must be that only the low type charges p', and thus $b(p') = 0$ with no consumers auditing at p'. From the proof of Proposition 1, p' cannot be between $\frac{.5+\underline{\theta}-z1}{2}$ and $p_L^{sep} = \frac{.5+\underline{\theta}+z1}{2}$ as this would cause the high type to deviate to p' (as the high type is earning full information profits at p^{hybrid} and could earn larger profits via mimicry by charing p' in this price range). Additionally, from the proof of Proposition 1 and by the strict concavity of the low type's full information profit function, since $\frac{.5+\underline{\theta}-z1}{2} < p_L^{full}$ and $\Pi^{00}(\frac{.5+\underline{\theta}-z1}{2}, L, 0) < \Pi^{00}(p_L^{sep}, L, 0) \leq \Pi^{11}(p_H^{full}, L, \phi)$, it must be that $\Pi^{00}(p, L, 0) < \Pi^{11}(p_H^{full}, L, \phi)$ for all $p \leq \frac{.5+\underline{\theta}-z1}{2}$; moreover, $\Pi^{00}(p, L, 0) < \Pi^{00}(p_L^{sep}, L, 0)$ for all $p > p_L^{sep}$ as well. Finally, from Proposition 7, it must be that $\Pi^{11}(p_H^{full}, L, \phi) \geq \Pi^{00}(p_L^{sep}, L, 0)$ in order for $p^{hybrid} = p_H^{full}$ to meet the conditions of an SPFA equilibrium price. Therefore, the only possible way for a hybrid equilibrium to exist occurs in the degenerate case when $\Pi^{11}(p_H^{full}, L, \phi) = \Pi^{00}(p_L^{sep}, L, 0)$, or equivalently, when $\Pi(p_H^{full}, L, 0) = \Pi(p_L^{sep}, L, 0)$ in the case without consumer audit, so that the low type is exactly indifferent between being fully identified at the high type's full information price, and being fully identified at the separating price. Note that this requires a deterministic relationship between the parameters, in this case $c = -\frac{\omega X_H(.5+\underline{\theta}+\omega X_H)}{z_1-\omega X_H}$, and thus this case does not occur in general. Therefore, refinement by the Intuitive Criterion eliminates all (general) hybrid equilibria in the model with consumer audit.

II.7 Technical Appendix

II.7.1 Proof of $\max_{p>c} \Pi(p, L, 1) \leq \Pi\left(p_L^{full} + \frac{z_1-c}{2}, L, 0\right)$ when $c < z_1$ (For Propositions 1 and 9)

To see that this is the case, assume to the contrary that $c - 2\omega X_H > 2z_1$, or equivalently after squaring both sides and rearranging terms,

$$c(c+4\omega X_H) + 8\omega X_H(\frac{1}{2} + \underline{\theta} + \omega X_H - c) > 0.$$

If $c + 4\omega X_H \leq 0$, then $c(c+4\omega X_H) + 8\omega X(\frac{1}{2} + \underline{\theta} + \omega X_H - c) \leq 0$, resulting in the desired contradiction. Otherwise, if $c + 4\omega X_H > 0$, note that A.1$'$ implies $-\omega X_H < z_1$, and since $c < z_1$:

$$
\begin{aligned}
&c(c+4\omega X_H)+8\omega X_H(\frac{1}{2}+\underline{\theta}+\omega X_H-c)\\
&< z_1\,(z_1+4\omega X_H)+8\omega X_H(\frac{1}{2}+\underline{\theta}+\omega X-c)\\
&= -2\omega X_H(\frac{1}{2}+\underline{\theta})-(\omega X_H)^2+4\omega X_H z_1+8\omega X_H(\frac{1}{2}+\underline{\theta}+\omega X_H-c)\\
&= \omega X_H\left(6(\frac{1}{2}+\underline{\theta}+\omega X_H-c)+(\omega X_H-2c+4z_1)\right)\\
&< 0.
\end{aligned}
$$

Thus $c+4\omega X_H>0$ results in a contradiction as well. Thus it must be that $c-2\omega X_H\leq 2z_1$, and $\max_{p>c}\Pi(p,L,1)\leq\Pi\left(p_L^{full}+\frac{z_1-c}{2},L,0\right)$ when $c<z_1$.

II.7.2 Proof that $p'_H < P'_L$ for Proposition 4

Suppose to the contrary that $p'_H\geq p'_L$, or equivalently (after simplifying, squaring both sides, and rearranging) that:

$$
\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H+\sqrt{(\frac{1}{2}+c+\underline{\theta})^2-4(-p^{*2}-\lambda c\omega X_H+p^{pool}(\frac{1}{2}+c+\underline{\theta}+\lambda\omega X_H))}\leq 0
$$

Then since $\omega<0$ and $p^{pool}<\frac{1}{2}+\underline{\theta}+\lambda\omega X_H$,

$$
\begin{aligned}
&\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H+\sqrt{(\frac{1}{2}+\underline{\theta}+c)^2-4(-p^{*2}-\lambda c\omega X_H+p^{pool}(\frac{1}{2}+c+\underline{\theta}+\lambda\omega X_H))}\\
&=\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H\\
&\qquad+\sqrt{(\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H)^2+2p^{pool}\lambda\omega X_H-4(.5+\underline{\theta}+\lambda\omega X_H)\lambda\omega X_H}\\
&>\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H\\
&\qquad+\sqrt{(\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H)^2+2(.5+\underline{\theta}+\lambda\omega X_H)\lambda\omega X_H-4(.5+\underline{\theta}+\lambda\omega X_H)\lambda\omega X_H}\\
&=\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H+\sqrt{(\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H)^2-2(.5+\underline{\theta}+\lambda\omega X_H)\lambda\omega X_H}\\
&>\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H+\sqrt{(\frac{1}{2}+\underline{\theta}+c-2p^{pool}+2\lambda\omega X_H)^2}\\
&\geq 0,
\end{aligned}
$$

a contradiction. Thus $p'_H<p'_L$ for all values of $\underline{\theta}$, X_H, and c when $\omega<0$.

II.7.3 Derivation of the ε price deviation for Proposition 10

There are two sub cases to consider. First, suppose that $p^{SPAA} \leq \arg\max\{\Pi^{10}(p,H,\lambda)\}$, that is, the SPAA equilibrium price is no larger than the price that would maximize the low type's profits in a SPAA equilibrium. Now consider the profits for both types at the price $p^{SPAA} - \varepsilon_1$ when $b(p^{SPAA} - \varepsilon_1) = 1$, and ε_1 is defined by:

$$\begin{aligned} 0 < \varepsilon_1 < \min \Big\{ & p^{SPAA} - c, .5(-(.5 + \underline{\theta} + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_H - 2p^{SPAA}) \\ & + \sqrt{(.5 + \underline{\theta} + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_H - 2p^{SPAA})^2 + 4(1-\alpha)(1-\lambda)\gamma_S p^{SPAA}} \,) \Big\}. \end{aligned}$$

Note that the latter expression in this minimum can easily be shown to be positive under the assumptions of the model. As $\Pi^{10}(p,L,\lambda)$ will be strictly increasing[22] in p at p^{SPAA}, and $\Pi^{10}(p,L,\lambda) \geq \Pi^{00}(p,L,1)$ for all $p > c$, it must be that $\Pi^{10}(p^{SPAA},L,\lambda) > \Pi^{00}(p^{SPAA} - \varepsilon_1, L, 1)$ so that the low type would have no incentive to deviate from p^{SPAA} if beliefs $b(p^{SPAA} - \varepsilon_1) = 1$. The profits for the high type at $p^{SPAA} - \varepsilon_1$ are given by $\Pi^{00}(p^{SPAA} - \varepsilon_1, H, 1)$, and these profits are greater than the high type's profits in the SPAA equilibrium when:

$$\begin{aligned} (p^{SPAA} - \varepsilon)(.5 + \underline{\theta} + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_H - p^{SPAA} + \varepsilon) \\ > (p^{SPAA})(.5 + \underline{\theta} + \alpha\gamma_A X_H + (1-\alpha)\gamma_S \lambda X_H). \end{aligned}$$

This results in a quadratic inequality in ε_1, with solution given by:

$$\begin{aligned} .5\Big(& -(.5 + \underline{\theta} + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_H - 2p^{SPAA}) \\ & - \sqrt{(.5 + \underline{\theta} + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_H - 2p^{SPAA})^2 + 4(1-\alpha)(1-\lambda)\gamma_S p^{SPAA}} \Big) \\ < \varepsilon_1 < .5\Big(& -(.5 + \underline{\theta} + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_H - 2p^{SPAA}) \\ & + \sqrt{(.5 + \underline{\theta} + (\alpha\gamma_A + (1-\alpha)\gamma_S)X_H - 2p^{SPAA})^2 + 4(1-\alpha)(1-\lambda)\gamma_S p^{SPAA}} \Big), \end{aligned}$$

where the lower bound is always negative and the upper bound is always positive. As a result, the

[22] If $p^{SPAA} = \arg\max\{\Pi^{10}(p,H,\lambda)$, then $\Pi^{10}(p,L,1\lambda)$ will not be strictly increasing at p^{SPAA}; however, $\Pi^{10}(p,L,\lambda)$ will be strictly less for all $p < p^{SPAA}$.

high type will earn larger profits by deviating to $p^{SPAA}-\varepsilon_1$ when $b(p^{SPAA}-\varepsilon_1)=1$. Thus there will always be a price $p<p^{SPAA}$ for which $b(p)=1$ will provide incentive for the high type to deviate, but no incentive for the low type to deviate. Therefore there are no SPAA equilibria in this subcase that satisfy the Intuitive Criterion.

For the second case suppose that $p^{SPAA}>\arg\max\{\Pi(p,X_H,1,0;\lambda)$. A similar argument to that of the previous subcase shows that if $b(p^{SPAA}+\varepsilon_2)=1$, where ε_2 is defined:

$$0<\varepsilon_2<\min\Bigg\{.5+\underline{\theta}+(\alpha\gamma_A+(1-\alpha)\gamma_S\lambda)X_H-p^{SPAA},.5\Bigg((.5+\underline{\theta}+(\alpha\gamma_A+(1-\alpha)\gamma_S)X_H-2p^{SPAA})$$
$$+\sqrt{(.5+\underline{\theta}+(\alpha\gamma_A+(1-\alpha)\gamma_S)X_H-2p^{SPAA})^2+4(1-\alpha)(1-\lambda)\gamma_S p^{SPAA}}\Bigg)\Bigg\},$$

then the high type would have incentive to deviate to $p^{SPAA}+\varepsilon$, while the low type would not. Therefore any SPAA equilibrium in this subcase will not satisfy the Intuitive Criterion. As the two subcases above exhaust the possibilities of the larger case, there are is no SPAA equilibrium in which $\Pi^{10}(p,L,\lambda)\geq\Pi^{00}(p,L,1)$ (ie. $(1-\alpha)(1-\lambda)\gamma_S\geq-\alpha\gamma_A$) that satisfy the Intuitive Criterion.

II.7.4 Proof that $p'_L<p'_H$ for Proposition 10

To show that $p'_L<p'_H$, or equivalently (simplifying, squaring both sides, and rearranging)

$$2p^{SPAA}(\alpha\gamma_A X_H)+c(.5+\underline{\theta}-\omega X_H+2(1-\alpha)\gamma_S\lambda X_H-2p^{SPAA})$$
$$-c\sqrt{(.5+\underline{\theta}+\omega X_H)^2+4((p^{SPAA})^2-p^{SPAA}(.5+\underline{\theta}+\omega X_H-(1-\alpha)(1-\lambda)\gamma_S X_H))}<0.$$

Then

$$
\begin{aligned}
&2p^{SPAA}(\alpha\gamma_A X_H)+c(.5+\underline{\theta}-\omega X_H+2(1-\alpha)\gamma_S\lambda X_H-2p^{SPAA})\\
&\quad -c\sqrt{(.5+\underline{\theta}+\omega X_H)^2+4((p^{SPAA})^2-p^{SPAA}(.5+\underline{\theta}+\omega X_H-(1-\alpha)(1-\lambda)\gamma_S X_H))}\\
=&2p^{SPAA}(\alpha\gamma_A X_H)+c(.5+\underline{\theta}-\omega X_H+2(1-\alpha)\gamma_S\lambda X_H-2p^{SPAA})\\
&\quad -c\sqrt{(.5+\underline{\theta}+\omega X_H-2p^{SPAA})^2+4p^{SPAA}(1-\alpha)(1-\lambda)\gamma_S X_H}\\
<&2p^{SPAA}(\alpha\gamma_A X_H)+c(.5+\underline{\theta}-\omega X_H+2(1-\alpha)\gamma_S\lambda X_H-2p^{SPAA})-c\sqrt{(.5+\underline{\theta}+\omega X_H-2p^{SPAA})^2}\\
=&2p^{SPAA}(\alpha\gamma_A X_H)+c(.5+\underline{\theta}-\omega X_H+2(1-\alpha)\gamma_S\lambda X_H-2p^{SPAA})-c|(.5+\underline{\theta}+\omega X_H-2p^{SPAA})|\\
=&2p^{SPAA}(\alpha\gamma_A X_H)+c(.5+\underline{\theta}+\omega X_H-2p^{SPAA})-c|(.5+\underline{\theta}+\omega X_H-2p^{SPAA})|+c(-2\omega X_H+2(1-\alpha)\gamma_S\lambda X_H)\\
=&2p^{SPAA}(\alpha\gamma_A X_H)+c(.5+\underline{\theta}+\omega X_H-2p^{SPAA})-c|(.5+\underline{\theta}+\omega X_H-2p^{SPAA})|+c(-2\alpha\gamma_A-2(1-\alpha)(1-\lambda)\gamma_S)X_H\\
<&\,0
\end{aligned}
$$

since $(1-\alpha)(1-\lambda)\gamma_S > -\alpha\gamma_A$. Thus it must be that $p'_L < p'_H$. Therefore there will always exist a price p between p'_L and p'_H such that if $b(p)=1$ the high type has incentive to deviate but the low does not, and any SPAA equilibria in this case will fail to satisfy the Intuitive Criterion.

CHAPTER III

ASSESSING THE VALUE OF A PROXY FOR CONSUMER HEALTH CONSCIENTIOUSNESS IN THE ESTIMATION OF DEMAND

III.1 Introduction

Nutrition is currently an active area of policy debate because of its potential to reduce obesity rates as well as other associated health problems. Disentangling the effects heterogeneity in healthy vs unhealthy eating habits is not only an important step toward understanding how consumer health can be improved, but is also important in determining a firm's optimal strategy across markets in which consumer preferences over a product's "healthiness" differ. To this end, we seek to assess the value of a proxy variable for consumer health conscientiousness in estimation of demand for products wherein a brand's healthiness may be of concern. Naturally, a consumer's determination of a product's healthiness is a matter of perception, which could be argued to be a function of any number of observable and unobservable product characteristics together with the consumers' level of concern for their health (consumer health conscientiousness). One could attempt to perform an analysis that controls for heterogeneity in the perception of a product's healthiness. However, a complicated analysis of perception is unnecessary given our stated goal (and would almost necessarily be confined to an experimental setting). We proceed by investigating the significance of the coefficients on the interactions of a variable measuring health conscientiousness with observable product characteristics in the estimation of demand across a category wherein a brand's "healthiness" is important to some consumers.

The ideal dataset for such a study would be a panel of individual (household) purchases, complete with demographic information including a variable serving as a measure of health conscientiousness. Given such a dataset, the analysis would be a straightforward application of discrete choice, such as a random coefficients logit model. However, to our knowledge no such panel currently exists. Therefore, we apply the discrete choice model presented in Nevo (2001) to a simulated population of households combining data from both the Current Population Survey (henceforth

CPS) and the Behavioral Risk Factor Surveillance System (henceforth BRFSS) in order to associate each household with a dummy variable for smoking. Given the well known and very serious health implications of smoking, the result is a simulated panel of households for each city that includes an easily justifiable proxy variable for consumer health conscientiousness. This dataset is then used to investigate whether or not the coefficients on the interaction of the smoking dummy and health-related product characteristics are statistically significant in the estimation of demand in the ready-to-eat cereal market; a category chosen for the substantial variation in the characteristics associated with healthy eating, such as sugar, fat, and fiber content, and for the variety of brands that target health-conscious consumers.

This investigation is of interest for two reasons. First, if consumer health conscientiousness is an important component in the decisions of consumers and firms, models that do not adequately control for this could produce biased estimates. Thus, the significance of the interactions in this paper should be of interest to both marketers and economists. Obviously, knowing that a variable is important is somewhat moot if the variable is not available in the data. To this end, our analysis also provides necessary evidence for firms considering the collection of health-related demographic variables, as there are significant costs in collecting, storing, and protecting consumer data. The final reason that the results of this paper are important is that if the fairly basic interactions considered in this model have significant effects on demand, it stands to reason that a more complex model of consumer perception, product "healthiness" and product characteristics is justified and potentially valuable to marketing academics and practitioners.

III.1.1 Outline

Section 2 of this paper describes the consumer demand model. Section 3 provides a description of the data used in the estimation. Section 4 outlines the estimation procedure, and Section 5 provides the results of the structural estimation. Section 6 concludes the paper.

III.1.2 Related Literature

To our knowledge, there are no papers that use proxy variables for consumer health conscientiousness in the estimation of demand for a product. However, there is a substantial literature in the field of marketing consumer behavior that relates to our goal. Andrews, Burton, and Netemeyer (2000), Kozup, Creyer, and Burton (2003), Raghunathan, Naylor, and Hoyer (2006), and Downs, Loewenstein, and Wisdom (2009) are all examples of research directed at how consumers choose healthy versus unhealthy products. These papers investigate the effects of nutrition claims, information disclosure, and consumer perception on such choices. Though interesting and related, these papers are the result of experimental work and yield no information with respect to our stated goal other than suggesting that health conscientiousness may be valuable in the estimation of demand.[1]

As mentioned in the previous section, there is currently no individual-level panel dataset complete with purchases and demographic information about consumer health conscientiousness. Therefore, we use city-level scanner data together with demographic information from other sources to create and analyze a model of individual demand at the aggregate level. The prominent paper by Berry, Levinsohn, and Pakes (1995), henceforth BLP, develops the random-coefficients logit model for aggregate data that is used in our analysis. This method allows the use of instrumental variables to control for the endogeneity of prices, and is applied to the US automobile market (at the aggregate level) to derive estimates for elasticities that allow for consumer heterogeneity in a structural model without the unreasonable substitution patterns resulting from a basic multinomial logit analysis. A key step in the process of estimation is the inversion of the market shares via simulation with a technique presented in Berry (1994), alternatives to which are discussed in the conclusion of this paper.

The model most similar to what we apply to our data is that of Nevo (2001), wherein he applies the method of BLP to panel data from the IRI Marketing data set. A nice feature of this model is that the panel structure allows the inclusion of brand-specific dummy variables, allowing for the assumption of the exogeneity of product characteristics (required in BLP) to be relaxed. Nevo uses this method to investigate the strategic behavior/market power of firms in the ready-to-eat cereal

[1] The results from the consumer behavior literature may not be directly applicable to the model presented here, but could be quite useful in the construction of a model that controls for consumer perception in the estimation of demand.

industry. Though similar in construction, there are several differences between Nevo's methodology and what we present here. First, Nevo is not interested in consumer health conscientiousness. More importantly, since Nevo does not control for health conscientiousness, any residual effects of health conscientiousness not picked up by his demographic variables (age and income) are left in the error term. As a result, if health conscientiousness is considered by firms in the pricing decision, it must be assumed that health conscientiousness is uncorrelated across the markets within the different Census regions (South, Northeast, Midwest, West) for Nevo's price instruments to be valid. Another difference between Nevo's model and our model is the determination of the size of each market in the sample. Nevo simply sets the number of people in the market to the population of the corresponding city. However, the stores in the IRI dataset are far from a census of the stores in each city, and thus Nevo's estimate of market size is a sizable overestimate. Moreover, the degree of overestimation varies in magnitude from city to city, as the number (and size) of stores from which data is collected in each city is not proportional to population. Alternatively, we seek to estimate the number of people "served" by each store in the dataset, then aggregate this number over the stores in each city to construct our market size estimates.

III.2 Consumer Demand

As mentioned in the previous section, there does not currently exist a panel dataset with individual-level purchases together with a proxy variable for health conscientiousness. Therefore, to answer the question at hand we adapt the model of consumer demand described in Nevo (2001), applicable to aggregate-level datasets, wherein a consumer is assumed to either purchase a single unit of the good[2] or take the outside option. Suppose that for each of $m = 1, ..., M$ markets there are $i = 1, ..., \mathcal{I}_m$ consumers, and $j = 1, ..., J_m$ brands, together with a K-dimensional row vector of observable product characteristics x_j (other than price). Given this, if p_{jm} is the price of brand j in market m and y_i is consumer i's income, the indirect utility that consumer i receives from consuming a single unit of

[2] The calculation of observed market shares is made via an assumption on the maximum amount of cereal a person can consume in a single sitting (Nevo uses a maximum of 1 serving per day, per consumer). Thus the unit demand assumption can be justified as reflecting a per-sitting decision, or can simply be viewed as a reasonable approximation to true consumer demand. It is worth noting that there are recent papers that relax the unit demand assumption, such as Nevo, Rubinfield, and McCabe (2005) and Gentzkow (2007). However, we believe that our goal of a simple analysis is best served with the model at hand.

brand j is given by[3]

$$u_{ijm} = \alpha_i(y_i - p_{jm}) + x_j\beta_i + \xi_j + \Delta\xi_{jm} + \varepsilon_{ijm}, \tag{III.1}$$

wherein α_i is the individual-specific marginal utility of income, β_i is the K-dimensional vector of individual-specific coefficients for product characteristics, ξ_j is the (unobserved) mean contribution to utility of brand j across all consumers in all markets, $\Delta\xi_{jm}$ is the market-specific deviation from ξ_j, and ε_{ijm} is an individual-specific mean-zero stochastic deviation from ξ_j. Given that α_i and β_i determine how consumer i values brand j, it is natural to think that these coefficients will not only depend on the product's characteristics, but on the characteristics of the consumer as well. To that end, suppose D_i is a d-dimensional vector of demographic variables of consumer i. Since a marketer (or an econometrician) cannot observe all characteristics of the consumer, let v_i be a $K+1$-dimensional vector capturing the unobserved characteristics that interact with product characteristics. Assuming that the unobserved characteristics v_i are distributed multivariate normal and are independent of D_i,

$$\binom{\alpha_i}{\beta_i} = \binom{\alpha}{\beta} + \Pi D_i + \Sigma v_i, \qquad v_i \sim N(0, I_{K+1}), \tag{III.2}$$

wherein Π is a $(K+1)\times d$-dimensional matrix of coefficients on the interaction between consumer demographic variables and product characteristics, Σ is a $(K+1)\times(K+1)$-dimensional scaling matrix of parameters determining the interaction of the unobserved consumer characteristics in v_i with the product's characteristics, and α and β are the components of α_i and β_i that are identical across consumers. Since α, β, and ξ_j are invariant across all consumers, and $\Delta\xi_{jm}$ is invariant across all consumers in market m, the portion of consumer i's valuation of brand j that depends upon the characteristics of brand j can be broken into the sum of a mean valuation of the characteristics of brand j across all consumers in the market and a heterogeneous (individual specific) component. Defining $x_{jm} = \binom{p_{jm}}{x_j}$, a vector of product characteristics of brand j in market m, utility can be written as follows:

[3] Note that this indirect utility function is the result of a quasilinear utility function, wherein there are no wealth effects (a reasonable assumption in the ready-to-eat cereal category). If this assumption is considered too strong for a given product, utility can be constructed similarly to that of BLP (using Cobb-Douglas utility) so that wealth effects are included in the model.

$$u_{ijm} = \alpha_i y_i + \delta_{jm} + x'_{jm}(\Pi D_i + \Sigma v_i) + \varepsilon_{ijm},. \tag{III.3}$$

where the income term $\alpha_i y_i$ will be shown to have no effect, $\delta_{jm} = x_j\beta - \alpha p_{jm} + \xi_j + \Delta\xi_{jm}$ is the mean utility, and the remainder of the expression is a mean-zero individual-specific heteroskedastic deviation from the mean utility. Note that this specification yields two distinct sets of parameters, those that enter utility in a linear fashion (α, β, ξ_{jm}) and those that enter utility in a non-linear way (Π, Σ); this difference in parameter types is exploited to simplify the estimation procedure in Section 3.2.

Based on the utility for each brand j, consumers will choose to purchase one unit of the brand that provides the greatest utility, or choose to purchase none of the brands. To facilitate the latter possibility, it is assumed that there is an outside good with indirect utility given by u_{i0m}. The components of the utility for the outside good are not identified individually, and are estimated only via the coefficient on the intercept term in the x_j matrix.

III.3 Data and Estimation

III.3.1 The Data

Given the model described above, the following data is required to proceed with the estimation: i) market shares and prices of a variety of cereal brands in several different markets, ii) characteristics of each cereal brand in our sample, and iii) the distribution of relevant consumer characteristics in each of the markets in our sample, in particular, a characteristic that serves as a proxy for consumer health conscientiousness. The prices and market shares (as well as our instrumental variables to handle price endogeneity) all come from the IRI marketing database (described below). Consumer characteristics are constructed by combining data from the annual March Current Population Survey with data from the Behavioral Risk Factor Surveillance System. Finally, the cereal characteristics were obtained by visiting grocery stores and looking at the labels of each cereal in our data set.

Table III.1: Cereal brands used in the analysis.

1) GENERAL MILLS CHEERIOS	13) KELLOGGS FROSTED FLAKES
2) GENERAL MILLS CINNAMON TOAST CRUNCH	14) KELLOGGS FROSTED MINI WHEATS
3) GENERAL MILLS HONEY NUT CHEERIOS	15) KELLOGGS RAISIN BRAN
4) GENERAL MILLS KIX	16) KELLOGGS RICE KRISPIES
5) GENERAL MILLS LUCKY CHARMS	17) KELLOGGS SPECIAL K
6) GENERAL MILLS TOTAL	18) POST GRAPE NUTS
7) GENERAL MILLS TRIX	19) POST HONEY BUNCHES OF OATS
8) GENERAL MILLS WHEATIES	20) POST RAISIN BRAN
9) KELLOGGS CORN FLAKES	21) POST SHREDDED WHEAT
10) KELLOGGS CORN POPS	22) QUAKER CAP N CRUNCH
11) KELLOGGS CRISPIX	23) QUAKER LIFE
12) KELLOGGS FROOT LOOPS	

III.3.1.1 IRI Marketing Data

Information Resources, Inc. have recently made a large set of scanner data available for academic research. These data cover 31 categories in 50 different markets from 2001 and 2006. Although the full data set contains both grocery stores and drug stores, we limit our analysis to the market shares of cereal sold in grocery stores. [4] The weekly store-level information is aggregated to the city-quarter level, less store-quarters for which there are missing weeks as well as cities for which we do not have sufficient consumer data. This leaves us with a total of 28 cities with observations for each of 24 quarters. We define each city-quarter as a market, and thus we have 672 total markets in the sample, with observations for each market-brand of the number of servings sold and revenue.[5] Price is the average price paid per serving (for each brand-market-quarter). We estimate demand across the 23 universally available cereal brands listed in Table III.1, leaving us with a total of 15,456 observations.

The dependent variable in our structural model is market share. Thus, we need an operational definition of market share that is a close approximation to the concept assumed in the underlying demand model. This is accomplished by assuming that the total market size is some (maximum) number of servings of cereal per person per day for each person in the market. In order to implement this idea, an estimation of the number of people served by each store is required. In Nevo (2001), the size of each market is estimated using the total population of the city (market) in question to calculate the maximum number of cereal servings possible per quarter. As the IRI data is not a

[4] Our presumption is that cereal purchase patterns found in drug stores may not be representative of those found in the rest of the market.

[5] Serving size was obtained from cereal packaging labels.

census of all stores in the market, this estimation of market size will generally overestimate the number of people served by the stores in the sample, and the degree of this overestimation may vary substantially across markets. Moreover, the small fraction of the population of a given city that actually shops at the included stores could also vary substantially over time. As a result, we chose to estimate the market size based on the number of individuals "served" by each store, that is, the total number of people in households wherein at least one individual from the household made at least one purchase at the given store in the given quarter. Once the number of people served by each store is estimated, we can then aggregate the number of people served by individual stores to get the total number of people served by each market.

The basic idea behind our approach is to use household level data included in the IRI data set for multiple stores in two US cities to estimate the conditional per capita consumption of 20 different categories of goods purchased at grocery stores using local demographic information. We then use the conditional averages and demographic data from the IRI scanner data markets to forecast per capita consumption for each store in the full sample. This leaves us with a forecast of per capita consumption for each of 20 product categories together with the observed total consumption in those categories for every store/quarter in the dataset. Finally, we pick the number of people that minimizes the sum of squared differences between actual volume sold and forecasted volume. This gives us an estimate of number of people served in each store/quarter.

We begin by assuming that per capita consumption depends on price of a category (averaged over the quarter and across all brands), and two demographic variables (age and income), and apply what is essentially a representative agent approach.

Suppose that the quantity sold in category j, at store s, during quarter t is determined the following way:

$$Y_{jst} = N_{st} \times g_j(x_{st}).$$

or on a per capita basis:

$$y_{jst} \equiv Y_{jst}/N_{st} = g_j(x_{st}).$$

Suppose now that the function $g_j(.)$ is such that the equations indexed by j above can be written as in matrix form as

$$\ln(y_{st}) = \gamma \ln(p_{st}) + G x_{st}.$$

where y_{st} is a $J \times 1$ vector with the j^{th} element equal to the volume sold per capita in category j at store s during quarter t, γ is a $J \times J$ diagonal matrix of price elasticities[6], p_{st} is a vector of the log of the average price per unit in each category, and where G is a matrix of coefficients that captures the effect of demographic variables x_{st} on the average quantity sold per quarter in each category. These demographic variables are the log of income, the average age, and the percent of the population that are children.

Prior to completing this stage of the estimation, the data are aggregated to the store-quarter level. The volumes are averaged across people-weeks so that the unit of measure is physical units per person per week. We then use the consumer panel data to estimate the system of equations indicated above using seemingly unrelated regressions, resulting in parameter estimates $\hat{\gamma}$ and $\hat{G}$.

Next, using the (aggregate level) IRI demographic data linked to the individual stores, together with the parameters estimated in the procedure outlined above, we predict the per capita consumption in each category via

$$\ln(\hat{y}_{jst}) = \hat{\gamma} \ln(p_{jst}) + \hat{G} x_{st}$$

As Y_{jst} is observed in the scanner data (and n_{st} is the variable we wish to estimate), we assume the model above can be applied to other stores[7] yielding

$$Y_{jst} = N_{st} \hat{y}_{jst} + \varepsilon_{jst},$$

where ε_{jst} is the forecast error. Given this, we can use the generated regressor $\hat{y}$ to run an OLS regression using the $J = 20$ observations (for each combination of s and t in the data) and obtain an estimate of N_{st} for each market.

There are, of course, potential problems in this estimation of market size. Forcing the regres-

[6] For simplicity we assume that the cross-category price elasticities are zero, thus this matrix is diagonal.

[7] The two US cities from which we estimate the model (Eau Clair, WI, and Springfield, IL) were specifically chosen by IRI as cities that represent the "average" person in the US. While there are most certainly differences between the populations of these cities and other cities in our dataset, we feel this should still provide a reasonable approximation of the number of people served by a store in other cities.

sion line through the origin removes the guarantee of mean zero errors in the OLS stage of the regression, and could also result in endogeneity between the generated regressors (predicted per-capita consumption) and the OLS errors. Figure III.1 provides a scatterplot of the OLS residuals versus the predicted total sales volumes for 10 randomly-selected stores. As expected, given that the regression line passes through the origin and that predicted total volumes must be positive, the residuals for very small predicted total sales volumes tend to be positive. However, Figure III.1 still provides evidence that the errors are mean zero away from the origin. The lack of endogeneity between the explanatory variable and the errors in the OLS stage (also potentially created by the lack of a constant term) can be seen in Figure III.2, plotting the predicted per-capita consumption for each category against the observed total volumes for 10 randomly selected stores. This graph provides evidence that the OLS errors are still orthogonal to the generated regressors, and that the estimation technique leaving the constant out of the regression is valid. Though it is clear from Figure III.1 that the residuals for the individual categories can be quite large relative to the predicted total volumes, it should be noted that the estimates of the number of people served by each store are consistent across quarters, while also tracking well with the observed total volumes of the categories for stores that are decreasing in total sales volumes (and subsequently the number of people the store serves). Moreover, an in-sample pseudo R-squared calculation using the IRI household panel data (wherein we have the true value of N_{jst}) yields a value of 94%, indicating that the above method does have significant explanatory power for predicting the number of people served by a given store. We contend that this procedure provides a much better picture of the size of the market than simply setting the number of people served by each market equal to the population of the city.

The final step in estimating market size is to define the maximum number of servings of cereal per person per quarter. Nevo (2001) makes the assumption that each individual consumes a maximum of one serving per day, though the data shows that some people, on average, consume significantly more than one serving of cereal per day. Analysis of the IRI Household Panel Data, reflected in Figure III.3, indicates that the median household consumes 5.3 servings of cereal per person per week and 10% of households consume more than 7 servings per person per week (based on a serving size of 28 mg). Though Nevo's estimation of market size is unlikely to be affected by this (given that the population of a city is a significant over-estimate of the people who enter the stores in the data set), to insure an accurate representation of the market (and to assure that the

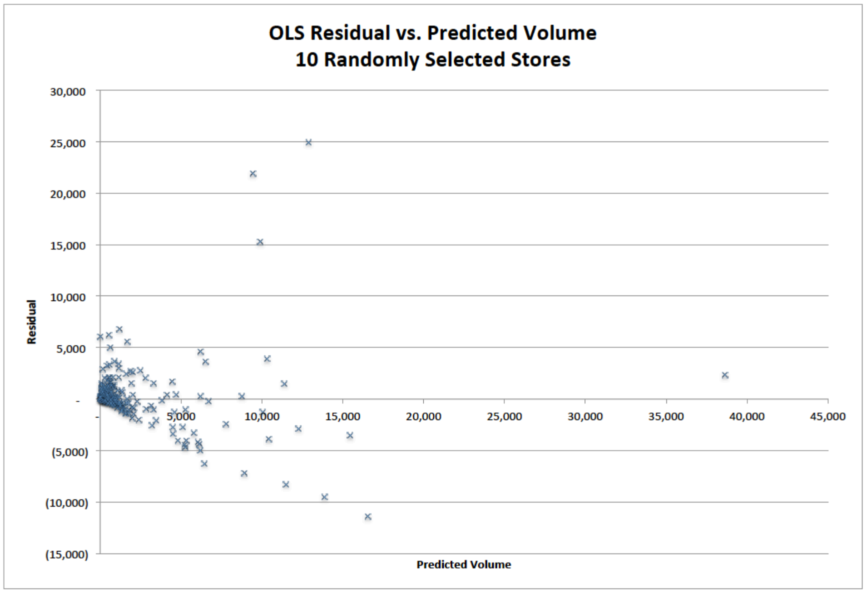

Figure III.1: Residuals versus fitted values.

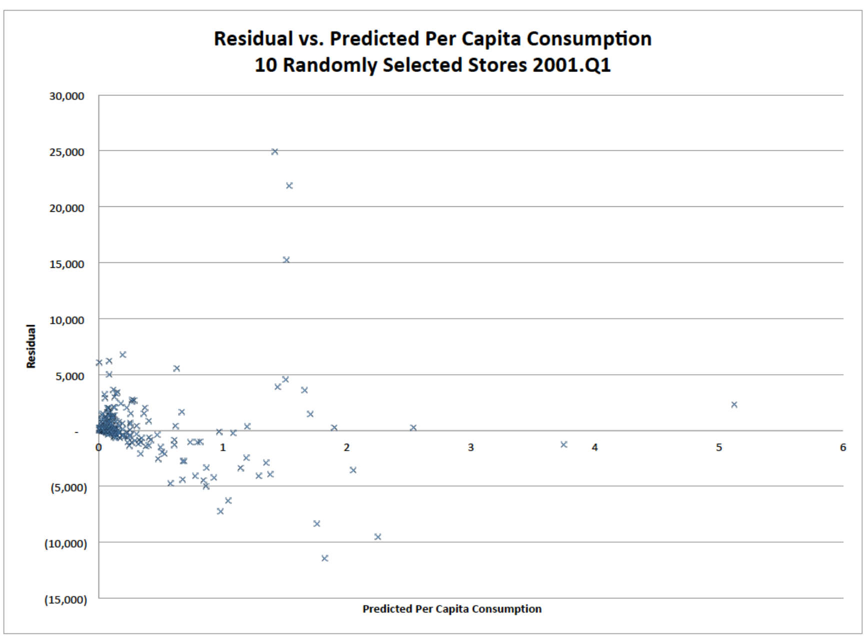

Figure III.2: Predicted per-capita consumption versus total volume.

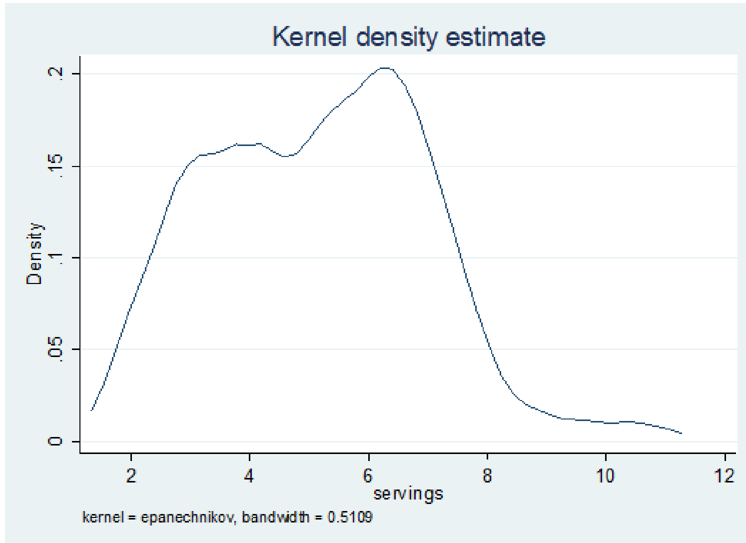

Figure III.3: Kernel density estimate of consumed weekly servings per person.

outside share is nonnegative) we take the highest per capita total cereal consumption found in the data set and use that as the basis for market size computation (slightly less than 3 servings per day). Given this, we can calculate the market size for each individual store (and for each market, which is the sum across all stores in a city/quarter).

III.3.1.2 Consumer Data

We simulate draws from the population of consumers using the March Current Population Survey for the years corresponding to the cereal market share data. For our primary estimates, we sample (with replacement) 40 individuals from the CPS for each market-year. The consumer variables we consider are income (total household income divided by the number of individuals in the house), income squared, age, and the dummy variable *child*, set equal to one if the individual is 16 years old or younger. Income is deflated by the CPI-U specific to that market. For the unobservable consumer characteristics, we simulate draws from an iid standard normal distribution.

Our proxy variable for health conscientiousness comes from data collected by the CDC via the Behavioral Risk Factor Surveillance System (BRFSS). This annual phone survey is conducted across the United States and includes a variety of demographic information including questions about tobacco use, the consumption of fruits and vegetables, amount of physical exercise, etc. Fig-

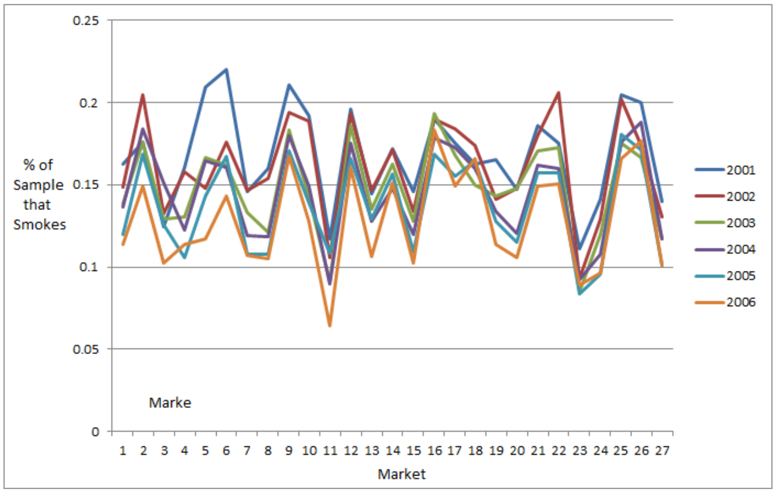

Figure III.4: Annual BRFSS smoking rates for each market

ure III.4 shows the smoking rates for the annual samples of the BRFSS for each of the 28 markets in our data, demonstrating the relative consistency of the survey data (and the slow decline of smoking over time). The problem, however, is that this information is limited to the adult that answers the phone, and does not include information about others in the household. In order to estimate the full structural model presented here, it is necessary to have a distribution of consumer characteristics (including that of children), from which to simulate the market shares. Thus, the limitations of the BRFSS prevent us from using this data alone. In order to construct a distribution of consumer characteristics, we need to predict tobacco use in the CPS dataset using the BRFSS data set.

A simple approach would be to use the BRFSS survey data to predict smoking behavior across all adults in the CPS data set. However, this would give no proxy variable for the health conscientiousness of children. This could be handled by assigning a dummy variable that takes the value 1 if an adult within the household smokes. However, note that tobacco use is very likely correlated within households, and thus predicting smoking across all adults across a market will result in an upward bias in the number of households with an adult smoker.

To facilitate the construction of a variable that reflects some measure of tobacco use across a market using the BRFSS data, we first estimate a probit model to predict the probability of smoking by income, age, education level, and whether or not there are children in the home (the demographic variables included in the BRFSS). After this, we draw a random adult from each household in the

Table III.2: Number of observations for each income/education level

	No School	Grades 1- 8	Grades 9 - 11	HS Grad or GED	1 to 3 yrs College	>= 4 Yrs of College
<$10k	28	1,639	2,814	5,065	3,356	1,812
$10k to $15k	12	1,443	2,615	5,856	3,516	1,727
15 to 20	32	1,449	3,023	8,393	5,211	2,816
20 to 25	19	1,119	2,852	10,764	7,779	4,759
25 to 35	22	872	2,409	13,447	12,084	9,153
35 to 50	13	363	1,631	13,975	15,882	18,238
50 to 70	7	160	912	11,395	16,392	27,106
>75k	17	147	721	9,884	18,520	62,315

CPS data set, and using the results of the aforementioned probit model we forecast the probability that each randomly selected person smokes. These probabilities are then used in a single Bernoulli trial for each randomly selected (adult) individual from each household to determine a prediction of whether or not the individual smokes or does not smoke. The resulting value of the dummy variable is then extended to each individual in the household. It is important to note that since we are using many years worth of the BRFSS data to estimate the probit, the parameter estimates for the probit model are constant for each city (market) across all quarters in the dataset.

Though psuedo R-squared calculations for the probit model in each individual market yield very low numbers (between 7% and 10%), these results are not surprising given the many contributing factors to individual smoking behavior (beyond age, income and education level). Moreover, if these demographic variables easily explained smoking behavior, then these same demographics would be able to control for health conscientiousness (given our assumption that smoking is a good proxy for health conscientiousness), and so the low explanatory power of these variables in predicting smoking behavior suggests the potential value of information about smoking behavior in the estimation of demand. On the other hand, our technique does do well in aggregate. Given income and education levels, Table III.2 provides the total number of observations for each income/education combination, while Table III.3 provides the difference between the aggregate predicted rates and the true aggregate rates from the BRFSS. As shown by these results, the model provides predictions that track well with aggregate smoking rates for income/education blocks with many observations.

Table III.3: Difference between predicted and actual smoking rates by income/educaton level.

	No School	Grades 1- 8	Grades 9 - 11	HS Grad or GED	1 to 3 yrs College	>= 4 Yrs of College
<$10k	28	1,639	2,814	5,065	3,356	1,812
$10k to $15k	12	1,443	2,615	5,856	3,516	1,727
15 to 20	32	1,449	3,023	8,393	5,211	2,816
20 to 25	19	1,119	2,852	10,764	7,779	4,759
25 to 35	22	872	2,409	13,447	12,084	9,153
35 to 50	13	363	1,631	13,975	15,882	18,238
50 to 70	7	160	912	11,395	16,392	27,106
>75k	17	147	721	9,884	18,520	62,315

Therefore, as the individual dummy variable will reflect both the explained variation in the city-level probit model and the across-city variation in smoking rates, we contend that this variable is a reasonable proxy for health conscientiousness (provided that health conscientiousness is significantly correlated across individuals within a household).

III.3.1.3 Instruments

We instrument prices with an average of that brand's price in all other markets in the region for each of 24 quarters. [8] For example, the price of Cheerios in Chicago during Q1 in 2001 is instrumented by the average price of Cheerios in the Midwest excluding Chicago for each of the 24 quarters in the data set. Note that these instruments should be correlated with market-level unobservables that effect endogenous prices, and thus are reasonable instruments.

III.3.2 Estimation

Given that each consumer is assumed to purchase a single unit of one brand, or purchase nothing at all, we can define the set of variables for which a consumer will choose to purchase good j; if $x_{\cdot m}$ is the matrix of characteristics, $p_{\cdot m}$ is the vector of prices, and $\delta_{\cdot m}$ is the vector of mean utilities (the same for all consumers in the market), for all brands in market m, this set is:

[8] The IRI markets are assigned to one of four census regions: West, Midwest, South, or Northeast.

$$C_{jm}(x_{\cdot,m}, p_{\cdot,m}, \delta_{\cdot,m}; \Pi, \Sigma) = \{(D_i, v_i, \varepsilon_i jm) | u_{ikm} \leq u_{ijm} \forall k \neq j\}.$$

Assuming that D_i has a known distribution $F(D)$, ε_{ijm} has distribution $F(\varepsilon)$, and the distribution of v_i is $F(v)$, we can use the assumed independence of these variables to derive the market share for brand j in market m (assuming that consumers are never indifferent between products):

$$s_{jm} = \int_{C_{jm}} dF(D)dF(v)dF(\varepsilon). \tag{III.4}$$

As previously mentioned, $F(v)$ is a multivariate normal distribution, and we assume that the values for ε_{ijm} are i.i.d. and that $F(\varepsilon)$ is a type 1 extreme value distribution. As we "observe" the market share of each brand in our data set, the above market share equation, which is dependent upon the values of the parameters, suggests a comparison between the observed market shares and those predicted via the consumer model. To facilitate easier calculation of s_{jm}, it could be assumed that consumers are identical save for the value of ε_{ijm}, and thus the values of all parameters that enter utility non-linearly are zero, $\alpha_i = \alpha$, $\beta_i = \beta$, and the only remaining error is the additively separable error ε_{ijm}. Though restrictive, this assumption results in the model reducing to the multinomial logit model, in which case equation (III.4) becomes

$$s_{jm} = \frac{\exp(x_{jm}\beta - \alpha p_j m + \xi_{jm})}{1 + \sum_{r=1}^{r=J_m} \exp(x_{rm}\beta - \alpha p_r m + \xi_{rm})},$$

where income drops out due to its presence in the utility of all brands. Though computationally easy to work with, the multinomial logit model results in unreasonable substitution patterns and downward biased own-price elasticities as discussed in the literature review of this paper. For a basis of comparison we do report results for the simplified (multinomial logit) model, but we then estimate the full model using the method described in Section 2. To begin, we construct a set of consumers for each market, complete with the vector of demographics D_i, by randomly selecting a fixed number (*ncons*) of "consumers" from the CPS data discussed in Section 3.1. Following Nevo (2001) and BLP, we use this set of consumers to approximate equation (III.4) using the smooth estimator

$$s_{jm} \approx \frac{1}{ncons} \sum_{i=1}^{ncons} \frac{\exp(\delta_{jm} + x'_{jm}(\Pi D_i + \Sigma v_i))}{1 + \sum_{r=1}^{r=J_m} \exp(\delta_{rm} + x'_{rm}(\Pi D_i + \Sigma v_i))}. \tag{III.5}$$

As this estimation of market shares is a function of the parameters, we could move forward with the estimation by simply trying to minimize the difference between observed shares from the data and the estimated shares; however, there are two problems with this approach. First, the number of parameters over which the minimization occurs would be very costly in the computational sense.[9] Second, and more importantly, price in this model is endogenously chosen by the firm, and thus may be chosen in response to variables remaining in the modeled error (in this case the difference between the observed shares and the shares given by equation (III.5)). Thus it is reasonable to assume that price is correlated with the error term, and we are left with an endogeneity problem. Both of these issues are resolved by applying the method put forth by BLP.

To implement this approach using the addition of brand dummy variables as in Nevo (2001), note that Berry (1994) defines the structural error term to be the effect on utility of the unobserved product characteristics, or $\xi_j + \Delta\xi_{jm}$. As brand dummy variables will capture all effects of unobserved product characteristics that are identical across consumers (ξ_j), this implies that the only remaining structural error[10] in the model is due to $\Delta\xi_{jm}$. Recall that $\delta_{jm} = x_j\beta - \alpha p_j + \xi_j + \Delta\xi_{jm}$, and thus $\Delta\xi_{jm} = \delta_{ijm} - (x_j\beta - \alpha p_j + \xi_j)$. Since product characteristics for a given brand are constant across all markets in the data set, brand dummy variables will also capture the effect of observed product characteristics (save for price), thus $\Delta\xi_{jm} = \delta_{jm} - (B_j - \alpha p_j)$, where B_j denotes the value of coefficient on the brand dummy variable for brand j. Therefore, $\Delta\xi_{jm}$ can be written as a function of the model's parameters (as δ_{jm} is a function of these parameters). Denoting the parameters as θ, and using a set of instruments for each brand/market, Z_{jm}, such that $E[Z'_{jm} \cdot \Delta\xi_{jm}(\theta^*)] = 0$ for all j and all m, where θ^* denotes the structural error at the true parameter values, we can then construct a GMM estimator as follows:[11]

[9] The MPEC method put forth in Dubé, et al. (2012) allows direct maximization using a different objective subject to the "constraints" given by the equality of observed versus simulated shares and moment conditions. We have not yet implemented this approach.

[10] Note that ε is not considered as part of the structural error as the i.i.d. assumption on ε would imply this to be superfluous in the moment condition expression.

[11] Note that our instruments, the average price of a brand over the region less the price within the given market, should be correlated with any market-level shocks to the price that would otherwise be in the error term. Thus the identification assumption is reasonable.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \xi(\theta)' Z \phi^{-1} Z' \xi(\theta), \tag{III.6}$$

wherein ϕ^{-1} is a weighting matrix consisting of a consistent estimator of $E[Z'\xi\xi'Z]$.

Now consider δ_{jm}, the mean utility for brand j in market m, required to calculate the value of the GMM objective. To solve for the value of δ_{jm} for each brand in each market we numerically invert the corresponding shares via the contraction mapping

$$\delta_{\cdot m}^{a+1} = \delta_{\cdot m}^{a} - \ln(S_{\cdot m}) - \ln(s_{\cdot m}), \tag{III.7}$$

for $m = 1, ..., M$, $a = 1, 2, ...$, and wherein $S_{\cdot m}$ denotes the observed shares. This equation is iterated until $||\delta_{\cdot m}^{a+1} - \delta_{\cdot m}^{a}||$ is below a predetermined tolerance level, yielding an estimate for the vector $\delta_{\cdot m}$ for each market. As a result, for any given value for the model parameters, $\delta_{\cdot m}$ can be estimated as a function of those parameters. Also note that since α and B_j enter the GMM objective in a linear fashion, the FOC's w.r.t. these parameters allow us to solve for the linear parameters in terms of the nonlinear parameters.[12] Therefore, since the only parameters required to estimate $\delta_{\cdot m}$ for all m are those that enter consumer utility in a non-linear fashion (that is, α, β, and B, where B is the vector of B_j's, are not required in order to get an estimate for mean utility), once the value of $\delta_{\cdot m}$ has been estimated for all markets, this estimate can be used to get an estimate for the parameters α and B. The estimates for $\delta_{\cdot m}$, α, and B can then be used to calculate the value of the GMM objective.

To summarize, for a given set of starting values for the nonlinear parameters Π and Σ, we can then estimate $\delta_{\cdot m}$ using the aforementioned contraction mapping. With the estimate of $\delta_{\cdot m}$ we can estimate the values of α and B, and with all of these estimates we can calculate the value of the GMM objective given by equation (III.6). We then search over values of Π and Σ to minimize this objective, which is done using the optimization program Knitro with an analytical gradient.

There are now three details left to discuss. First, we must estimate $E[Z'\xi\xi'Z]$. We follow Nevo's primary suggestion by starting with $\hat{\phi} = Z'Z$, searching for a "first-stage" estimate of the parameters, $\hat{\theta}_1$, and then using $\hat{\phi} = \sum_{m=1}^{M} \sum_{j=1}^{J_m} (\Delta\xi_{jm}(\hat{\theta}_1)^2 Z'_{jm} Z_{jm}$ to find a (re-weighted) second-stage estimate $\hat{\theta}_2$. Second, the variance covariance matrix for the estimates above is calculated, correcting for

[12] Using matrix calculus, with X_B denoting a matrix of prices and brand dummy variables, it can be shown that $[\alpha B.]' = (X_L' Z \Phi^{-1} Z' X_L)^{-1} X_L' Z \Phi^{-1} Z' \delta_{--}$.

the estimation error due to the contraction mapping. Finally, once an estimate of α, $B.$, Π, and Σ is found, we follow Nevo in exploiting the fact that the brand dummy variable coefficients B_j are composed of the mean effects of the observed product characteristics $x_j\beta$ and the mean effects of the unobserved product characteristics ξ_j so that $B_j = x_j\beta + \xi_j$. Thus by assuming that the conditional expectation of ξ_j given the observed product characteristics is zero ($E[\xi|x_j] = 0$), and defining V_B to be the variance covariance matrix of the $\hat{B}_j's$, generalized least squares can be used to obtain

$$\hat{\beta} = (X'V_B^{-1}X)^{-1}X'V_B^{-1}B,$$

wherein $X = [x_1\ x_2\ ...\ x_J]$ is the matrix of product characteristics of all brands in the sample (i.e. the observations are the brands), and $\xi. = B - X\hat{\beta}$.

III.4 Results

As a first step in investigating the interaction of health conscientiousness and product characteristics, we estimate the model as a multinomial logit as described in Section 3.2. This reduces the problem to a regression of $\ln(S_{jm}) - \ln(S_{0m})$ (wherein S_{0m} is the share of the outside good in market m) on price and product characteristics, with added interactions of product characteristics and consumer variables that are averaged over markets. These regressions are repeated for various combinations of these variables using two stage least squares regressions with the aforementioned price instruments, together with dummy variables for each brand and quarter. The final regression (vi) also includes a dummy variable for each market. The results of these regressions can be found in Table III.4, with heteroskedastically robust standard errors reported in parentheses. Note that throughout these results there is a positive and significant coefficient on the interaction of the tobacco proxy variable and sugar content, while the interactions of this proxy with fat content and fiber content are consistently insignificant. Though these regressions do support the notion that a proxy variable for consumer health conscientiousness is important in the estimation of demand, we must be careful in the interpretation of what this means given the aggregated (city-level) structure of this simplified model. For example, in regression (v) the coefficient on the interaction of the proportion of children in a market with the fiber content of a product is positive and significant (a seemingly counter-intuitive result). These results demonstrate the difficulty in estimating such effects using city-level consumer charactersitics, as the proportion of children is highly correlated with household size and age, and the resulting coefficient on this interaction could be explained via non-linearity of demand for fiber with respect to age combined with other variables "soaking up" the intuitive effects of the proportion of children in a city. While these logit results do provide evidence of the potential value of our smoking dummy variable in the estimation of demand, they also provide strong motivation for continuing to the full model, wherein the included consumer heterogeneity allows these variables to differ across consumers in the same market.

The next step is to estimate the full random-coefficients logit model using the smoking proxy variable. This was performed with a simulation draw of 40 consumers, resulting in the estimates presented in Table III.5 below,[13] where the first column presents the minimum-distance estimates, the second column is the estimates of the deviations from these values due to the unobserved consumer characteristics, and the final five columns are estimates of the coefficients on the interaction terms (with standard errors in parentheses).

The model results in estimates of price coefficients (and others) that are in line with previous results, including a negative coefficient on the interaction of child and fiber. While the coefficient on the interaction of tobacco and sugar is nearly, but not quite significant at the 10% level, it does suggest that tobacco use may be linked with increased demand for sugary cereals (especially in leu of the previously provided multinomial logit results). Moreover, the coefficients on the interaction terms of the smoking dummy and the other health-related product characteristics are also quite close to being significant at the 10% level, and in-line with what a researcher might intuitively expect.[14] Given these results with our simulated explanatory variables, it stands to reason that further analysis with survey data and/or panel data containing stated smoking behavior could sharpen these results significantly. In any case, marketers and economists should take note that measures of health conscientiousness could have significant interactions with health-related product characteristics in the determination of demand, and that the collection of these (possibly very relevant) consumer characteristics should be considered.

[13]The simulation was run 10 times for pseudo-random starting values generated from a normal distribution with mean zero and standard deviation 25, all of which converged to the results presented here. The (tight) inner loop tolerance for the contraction mapping was 10^{-14}.

[14]Higher fat content is typically associated with cereals containing nuts, and thus the negative coefficient on fat would indicate that people who smoke are less likely to purchase such cereals.

Table III.4: Multinomial logit results.

VARIABLES	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Price	-4.755***	-5.026***	-4.709***	-4.994***	-4.973***	-6.308***
	(0.212)	(0.212)	(0.212)	(0.212)	(0.212)	(0.212)
Log of Median Income	0.755***	0.663***	0.819***	0.707***	0.736***	-1.069
	(0.0306)	(0.0308)	(0.0639)	(0.0639)	(0.0664)	(1.107)
Log of Median Age	-1.405***	-1.148***	-1.409***	-1.150***	-1.209***	-3.427
	(0.117)	(0.116)	(0.116)	(0.116)	(0.274)	(2.664)
Median HH Size	0.175***	0.153***	0.204***	0.174***	0.189***	-0.119
	(0.00904)	(0.00907)	(0.0183)	(0.0183)	(0.0206)	(0.830)
Prop of Children	-6.991***	-6.453***	-6.987***	-6.451***	-7.248***	-10.83
	(0.263)	(0.260)	(0.263)	(0.260)	(0.607)	(6.973)
Sugar Content	0.00101	-0.000604	-0.0331***	-0.0563***	0.125	-0.0649***
	(0.00230)	(0.00228)	(0.00744)	(0.00855)	(0.101)	(0.00582)
Fiber Content	0.300***	0.298***	0.374***	0.374***	-0.0556	0.367***
	(0.0172)	(0.0170)	(0.0231)	(0.0248)	(0.220)	(0.0166)
Fat Content	4.350***	4.254***	4.477***	4.413***	3.914***	3.943***
	(0.212)	(0.210)	(0.214)	(0.215)	(0.691)	(0.151)
Tobacco Use	-	-1.772***	-	-2.364***	-2.334***	-9.012***
		(0.0945)		(0.213)	(0.215)	(2.403)
Tobacco x Sugar	-	-	-	0.101***	0.111***	0.106***
				(0.0201)	(0.0202)	(0.0145)
Tobacco x Fiber	-	-	-	-0.00845	-0.0364	-0.0211
				(0.0453)	(0.0455)	(0.0323)
Tobacco x Fat	-	-	-	-0.166	-0.189	-0.135
				(0.128)	(0.130)	(0.0848)
Age x Sugar	-	-	-	-	-0.0351	-
					(0.0249)	
Age x Fiber	-	-	-	-	0.0763	-
					(0.0542)	
Age x Fat	-	-	-	-	0.106	-
					(0.163)	
Income x Sugar	-	-	-0.0476***	-0.0437***	-0.0332***	-0.0431***
			(0.00597)	(0.00599)	(0.00641)	(0.00433)
Income x Fiber	-	-	0.0307**	0.0295**	0.000164	0.0250**
			(0.0135)	(0.0136)	(0.0145)	(0.00973)
Income x Fat	-	-	0.221***	0.215***	0.191***	0.218***
			(0.0355)	(0.0353)	(0.0383)	(0.0260)
Children x Sugar	-	-	-	-	-0.279***	-
					(0.0567)	
Chidren X Fiber	-	-	-	-	0.800***	-
					(0.125)	
Children x Fat	-	-	-	-	0.616*	-
					(0.361)	
HH Size X Sugar	-	-	0.00211	-0.0168***	0.00764***	0.00343***
			(0.00162)	(0.00361)	(0.00189)	(0.00109)
HH Size x Fiber	-	-	-0.0167***	-0.0168***	-0.0304***	-0.0165***
			(0.00359)	(0.00361)	(0.00428)	(0.00241)
HH Size X Fat	-	-	0.00439	0.00200	-0.00566	0.000947
			(0.0102)	(0.0102)	(0.0119)	(0.00647)
Constant	-4.479***	-4.925***	-4.563***	-4.875***	-4.527***	6.511
	(0.577)	(0.575)	(0.578)	(0.578)	(1.160)	(8.365)
Observations	15,456	15,456	15,456	15,456	15,456	15,456
1st stage R-Sq	0.9499	0.9449	0.9449	0.945	0.945	0.9561
1st stage F-Test	382.27	380.05	381.691	378.801	380.052	232.745
R-Sq	0.759	0.765	0.761	0.767	0.768	0.889

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table III.5: Results of the full model.

Variable			Demographic Variables: Income	Income Sq	Age	Child	Smoking
Constant	-8.398	0.010	-9.389		-0.001		-307.312
	(1.376)	(1.685)	(3.004)		(0.097)		(203.347)
Price	-26.218	0.103	31.133	-5.387		87.428	
	(1.685)	(5.957)	(8.987)	(2.916)		(11.372)	
Sugar	1.780	-0.008	0.621		-0.001		50.478
	(0.203)	(0.169)	(0.173)		(0.012)		(31.809)
Fiber	-0.053	-0.006				-30.407	-71.006
	(0.010)	(0.342)				(6.790)	(44.379)
Fat	0.733	-0.134					-82.343
	(0.020)	(1.365)					(52.080)

III.5 Conclusion

The goal of this paper is to investigate whether or not the interactions of a proxy variable for consumer health conscientiousness and easily-observed health-related quantitative product characteristics are useful in the estimation of demand for products where healthiness may be a concern to some consumers. We have shown that there is some evidence to suggest that smoking behavior, as a proxy for health conscientiousness, may be an important explanatory variable in the determination of demand for health-related product characteristics of ready-to-eat cereals. Replication of the above estimation procedures using different proxy variables and/or different food categories, or using a true household panel complete with stated smoking behavior, may provide more evidence for the importance of considering consumer health conscientiousness in the estimation of demand. These results may not only provide an impetus to pursue the question with more complex models, they may also provide guidance in how such models could and/or should be constructed. The use of health rankings of products, surveys of consumer perceptions about products, and other such data sources could potentially be justified by our analysis.

It should also be noted that while we have chosen to use the random coefficients logit with the BLP-style contraction mapping to analyze our aggregate-level dataset, there are alternatives to this method. The MPEC method put forth in Dubé, Fox, and Su (2012) uses the properties of sparse matrices to directly minimize the objective using the share equation as an extra constraint, and the SQUAREM technique suggested by Reynarts, Varadhan, and Nash (2012) is also an alternative to the BLP-stlye contraction mapping. Finally, the Bayesian technique presented in Jiang, et al. (2009) provides an additional avenue with which aggregate level data could be used to estimate the model of this paper. Though replication of the above results should be possible with the above methods, we believe the next step in the analysis is to collect individual-level data rather then rely on (inherently noisy) aggregate-level estimation techniques.

CHAPTER IV

ESTIMATING THE COMPONENTS OF HOUSEHOLD DEMAND USING REVEALED PREFERENCE DATA

IV.1 Introduction

One of the primary goals of marketing is to determine how a consumer base feels about the brands and/or relevant product characteristics in a product category. The ability to answer such questions is complicated by the significant amount of consumer heterogeneity present within most markets, which must be handled in such a way as to draw inference about "average" consumer preferences while still allowing for the recovery of actionable information about how preferences differ across the consumer base. This last point is especially important given the increasing amount of data available to marketing practitioners about consumer characteristics and purchase history, and the ever-improving ability to reach individual consumers with targeted marketing.

An often overlooked problem in modeling consumer preferences originates from the form of a common type of dataset, the household-level panel. It is widely accepted that the decisions of a household are affected by the interactions of the individuals within the household, and that a single set of preferences is insufficient to control for the underlying amalgamation of the (possibly) several sets of preferences that leads to observed purchasing behavior. Though there are many methods for handling heterogeneity across households (e.g. Random coefficients models, finite mixture models, and continuous mixture models[1]), almost every model in the extant literature on the estimation of consumer preferences either treats the household as having a single set of preferences or ignores households altogether and (essentially) assumes that individuals act independently. These approaches are somewhat defensible in much of the econometrics literature, as the parameters yielding individual consumer preferences are typically considered nuisance parameters in the estimation of the aggregate-level parameters used in market analysis and/or policy. However, given the aforementioned goal, such models are inadequate for the purposes of marketing.

[1] For a thorough discussion of the model available for handling heterogeneity across households, and their applicability to marketing problems, see Allenby and Rossi (1998).

In contrast, this paper presents a hierarchical Bayesian finite mixture model that allows the individual purchases in a household panel to be segmented into similar groups. The model estimates the parameters of the distribution of each segment as well as the probabilities that a given household's purchases belong to each segment, yielding valuable information about the "important" segments within a household demographic. Thus, the model presented in this paper is novel in that it retrieves both aggregate-level preferences and the individual-level information useful for targeted marketing, while also allowing for household purchases to be the result of multiple sets of preferences. Additionally, the Markov chain Monte Carlo (henceforth MCMC) algorithm for estimating the model may be run in parallel across households, allowing for application to the large datasets that are increasingly common in practice. For the purposes of demonstrating the function of the algorithm, the model is applied to a simulated household panel dataset, with convergence providing good results in the estimation of the true parameters used in the generation of the data.

IV.1.1 Outline

Section 2 of this paper describes the consumer demand model. Section 3 provides a description of the simulated data on which the model is applied. Section 4 demonstrates the potential problems with ignoring within-household heterogeneity. Section 5 provides the algorithm for the MCMC routine. Section 6 presents the results of the model when applied to the simulated dataset and discusses application to real household panel data. Section 7 concludes the paper.

IV.1.2 Related Literature

One potential method for handling the within-household heterogeneity discussed above is the technique found in Nevo (2001), applying the popular model of Berry et al. (1995) to household-level scanner data from U.S. grocery stores. Nevo's model specifies the utility for individuals as a function of the interactions between product characteristics and the characteristics of the individual (including age). However, as the data used in the Nevo's analysis is limited to aggregate-level shares of each brand (i.e. purchases are not linked to a specific person or household), estimation is achieved through simulation of the city-level markets. This is accomplished by taking a random sample of individuals from the current population survey (census data) of the cities in question and using the given values of the utility parameters to calculate the simulated shares of each brand in each market. Estimation of the parameters proceeds via a method for selecting the parameter values that minimize the distance between the observed shares and the simulated shares in each market. While useful in many applications, this method has several drawbacks that make it unsuitable for handling the problem highlighted here. First, Nevo's model essentially assumes that each individual makes their own purchases independently, ignoring the interactions between individuals in each household that result in the observed shares. Though my model does not explicitly estimate these interactions, any such effects will be picked up in the simulation of the household-level preference distributions. Second, Nevo's method is designed for aggregate-level data, and while household-level purchase data could be aggregated to create such a dataset, this results in the loss of potentially useful information (such as the vector of prices faced by individual households). Obviously, this loss of information can result in poor estimation of the individual vs. product interaction parameters. While these are nuisance parameters in Nevo's analysis, the fact that these parameters are the the unique points of interest to marketers make this technique unsuitable for the problem at hand.

Another method that could be used to address the problem at hand is given in Dubé (2004). In this paper, Dubé develops a model of household demand that is applied to household-level puchase data within the carbonated soft-drinks category. In order to handle the issue of multiple discreteness (i.e. multiple brands being purchased for a single household on a given shopping trip), Dubé defines the concept of a consumption occasion. Each purchase made by the household is assumed to maximize the utility for a specific consumption occasion, and the preferences for each product at a specific occasion are drawn from a household-specific distribution (a distribution of heterogeneity within each household's purchases). This distribution allows each household's set of purchases to be split over multiple preference draws (from the aforementioned distribution), and the estimated parameters of the this distribution can then be used to explain the observed multiple discreteness in the soft drink data. While my model adopts the notion of each purchase being linked to a "consumption occasion," a reasonable approach to handling variety-seeking in a household panel, there is an important flaw in Dubè's model with respect to the goal of this paper.

While Dubé does explicitly model within-household heterogeneity, the model assumes the distribution of the vector of utilities for a given household is distributed multivariate normal with a diagonal variance-covariance matrix. As a result, the unimodal structure of these distributions prevents the model from picking up differences in household preferences that are multi-modal. To see this, consider an example in the ice cream category in which there are two types of occasions for which a household purchases, one in which the adults in the household consume the ice cream and strongly prefer brand *a* over brand *b*, and another where children are the consumers and strongly prefer brand *b* over brand *a*. Modeling the distribution of this household's utilities over the two products with a bivariate normal distribution will provide poor estimates of the household's true preferences, as the unimodal distribution will (incorrectly) force the "average" of these preferences to be a more likely draw than either of the true sets of preferences associated with observed purchases. This problem is made worse by the use of a diagonal variance-covariance matrix in the within-household heterogeneity distribution, resulting in parameter estimates that yield no information about how the utility of brand *a* covaries with the utility of brand *b*. In this situation, Dubé's model will provide the marketer little to no evidence of the underlying preferences of such households, a serious problem when such preferences are prevalent in the category.

Rather than using a single distribution to model within-household heterogeneity, the model presented here applies what is known as a finite mixture model. In such models, the preferences of an individual (yielding the vector of probabilities of choosing any one product) are assumed to be described by one of finitely many "mixing distributions." For the previous example, one could imagine a model of the preferences over ice-cream flavors where adults are modeled with one distribution and children are modeled with another. The usefulness of such models comes from the ability to estimate the parameters of each of the mixing distributions without knowing, ex-ante, the distribution from which the preferences of each individual arise. One method for evaluating such models, known as the latent class model (henceforth LCM), is presented in Kamakura and Russell (1989). As in McFadden (1974), Kamakura and Russel specify an individual's utility to be a (linear) function of product characteristics with a random (Type I extreme value) error component, but with the parameters in the utility function of each consumer coming from one of finitely many "segments." It is assumed that the probability of each household belonging to a given segment is fixed, and estimation of the model proceeds via maximum likelihood using the likelihood of the observed purchase histories of the individuals. While Kamakura and Russell's model is an interesting, and very popular, method for segmenting a sample of households into homogeneous groups, it should be noted that this method alone does not address the aforementioned problem of within-household heterogeneity. The LCM assigns each household to a single segment, and therefore treats each household as having a single set of preferences.

The paper that is closest in structure to what is present in my model is that of Varki and Chintagunta (2004), wherein an LCM is augmented by the possibility that the purchases of some proportion of households are the result of a mixture of segments; that is, while some percentage of households are assigned to a single segment, the purchases of the rest of the households are a mixture wherein each purchase of the household results from the preferences of an individual segment. Varki and Chintangunta's choice model therefore allows for some degree of within-household segmentation to be picked up in the estimation, including the percentage of households in which purchases are the result of a mixture of segments and the parameters of the Dirichlet distribution that determines the probabilities of membership to each segment. Though their model is a step in right direction, there are still important reasons that Varki and Chintagunta's model does not adequately solve the problem. First, their model would have significant problems handling large datasets in which households purchase a wide variety of brands. This is because (as in most LCMs) the preferences for each individual at each choice occasion are a single draw from a fixed distribution that is the same for all members of the segment. As a result, if there is a significant number of households purchasing several brands from a large set of possible brands, the distribution of product purchases assigned to each segment for each household must reflect the fixed preferences of that segment. While this is not an issue if the number of segments is large relative to the number of brands being considered, or in cases where only a single brand is purchased fairly consistently (as in Varki and Chinatgunta's datasets of observed peanut butter and ketchup purchases), their model is ill-suited for datasets such as ready-to-eat cereal and soft-drinks.

Another important drawback of Varki and Chintagunta's approach is the treatment of the latent probability that a single purchase belongs to a given segment (for those households with purchases reflecting a mixture of segments). Households with different demographic characteristics will differ substantially with respect to the segments that combine to yield their observed purchases. For example, a household consisting of a young unmarried male will not likely have similar probabilities of assignment across segments as a family with several children. In Varki and Chintagunta's model, these probabilities are all drawn from the same distribution for all households (a single Dirichlet distribution). While this Dirichlet distribution will be able to pick up some heterogeneity in the assignment probabilities across households, the resulting parameters estimates will provide little to no information about how these households vary across the segments. That is, if the Dirichlet parameters are relatively small, as to allow for vastly differing probability vectors across households, the resulting distribution will be very "flat" and will yield little information about the probability of segment membership. However, if the estimated Dirichlet parameters are large enough to be somewhat informative about segment membership they will severely restrict the ability of the model to pick up heterogeneity in the assignment probabilities across household types, leading to dubious results.

In defense of Varki and Chintagunta's analysis, this is not likely an issue in the peanut butter and ketchup categories to which they apply their model, as preferences may not vary as widely across household types. Again, however, Varki and Chintagunta's model is unsuitable for categories such as ready-to-eat cereal and soft drinks, wherein the assignment to different segments will differ substantially across household types. A final problem with their model is that the estimation proceeds by integrating over the aforementioned Dirichlet density (via simulated maximum likelihood). Though the results include an estimated aggregate probability vector of all purchases belonging to a given segment, the mixture probabilities of households are not individually identified. Therefore, though some households are modeled as a mixture across the segments, there is little information about segment membership of these households recovered in the estimation. So Varki and Chintagunta's model does not provide much in the way of useful information for targeted marketing.

By comparison, the model presented in this paper uses Bayesian techniques to fit a finite mixture choice model, a technique presented in Diebolt and Robert (1994). I adapt the model of Allenby et al. (1998), a variant of Diebolt and Robert's model that (like Diebolt and Robert's and the typical LCM) assigns each household to a single segment to allow the individual choices of each household to be the result of a mixture of segments (defined as "consumption occasion types)." This allows my model to easily handle multiple discreteness/variety-seeking in complex markets such as ready-to-eat cereal and soft drinks. Moreover, unlike the models above, my model allows the probability vector of choice assignments to be drawn from a different Dirichlet distribution for each pre-specified household demographic. The benefit of applying this method over the traditional

LCM or hierarchical Bayesian mixture model is that within-household heterogeneity is modeled in such a way as to estimate the distribution of preferences within each consumption occasion type, while providing useful information about the how the choices of specific household (demographic) types are distributed across the consumption occasion types. As a result, the model provides both aggregate-level information about preferences and the more granular information about individual households that is valuable in modern marketing, all while being logically and computationally suitable for larger household panel datasets.

IV.2 The Model

The assumptions about household behavior used to derive the model are straightforward. As previously mentioned, I follow Dubé (2004) in using the notion of a "consumption occasion." A consumption occasion, the specific point at which the product is consumed, can differ greatly across individuals, or even across the same individual using a product for different purposes. For example, ready-to-eat cereal can be consumed for breakfast or for a late-afternoon snack, with preferences quite possibly being very different between these occasion types, even for the same individual. Thus, a structure built upon the idea of a consumption occasion can be very flexible in handling heterogeneity in purchase decisions, both across individuals and within the purchases of a single individual. To link consumption occasions to the observed purchase choices of a household within the context of a discrete choice model, it is assumed that each observed choice results from the maximization of the household's utility for a specific consumption occasion, so that each choice corresponds to a single consumption occasion, and viceversa. These consumption occasions are then assumed to belong to one of finitely many consumption occasion types, with the type of the consumption occasion determining the preferences for the specific choice. These consumption occasion types have similar, but not identical, preferences across households, and the probability that a choice belongs to any one of the consumption occasion types is fixed for each individual household. Unlike the model of Varki and Chintagunta, each household's vector of probabilities is not restricted to be drawn from the same distribution, but these probability vectors are assumed to be similar across demographically similar households. The choice model used to describe this structure is then constructed using the following notation:

i) The product is denoted by $j \in \{1,2,...,J\}$

ii) The household is denoted by $h \in \{1,2,...,H\}$

iii) The household demographic type is denoted by $\tau \in \{1,2,...,\Upsilon\}$

iv) The consumption occasion is denoted by $t \in \{1,2,..,T\}$

v) The consumption occasion type is denoted by $k \in \{1,2,...,K\}$

Given this, the utility for household h for a unit of product j at a consumption occasion of type k is given by

$$U_{hk} = \gamma_{hkj} + \varepsilon_{hkjt}$$

where γ_{hkj} is the mean utility of household h for product type j at consumption occasions of type k, and ε_{hkjt} is a stochastic deviation from that mean utility that is different for each single consumption occasion of type k of the household. As all households are assumed to have similar, but not identical, preferences for a given consumption occasion type, the γ_{hkj} for each consumption occasion type of household h is a function of product characteristics defined as

$$\gamma_{hjk} = x_{hj}\beta_k + \xi_{hjk}$$

where x_{hj} is a household-specific row vector of product characteristics, β_k is a column vector of coefficients that are the same across all households at consumption occasions of type k, and ξ_{hjk} is household-product specific error for that occasion type. Thus, the term $x_{hj}\beta_k$ is the mean utility (given the value of the product characteristics) across all households at consumption occasions of type k and ξ_{hjk} is the deviation from that mean utility for household h.

Estimation proceeds via a hierarchical Bayesian mixture model, the algorithm for which is described in section 6 below, by assuming that the ε's are distributed iid type I extreme value and that the vector of ξ_{hjk}'s are distributed multivariate normal with variance-covariance matrix Σ_k. Note that, given the resulting logit choice probabilities, the γ value for one of the products must be fixed for the model to be identified, and thus there are $J-1$ elements in the γ_h vector for each household/occasion type. Therefore, $x_{hj}\beta_k$ actually yields the difference in mean utility between the given product and the first (fixed) product. The values for the γ_h vectors are drawn using the purchase choices of individual households, while identification of the β_k vectors and the Σ_k matrices is achieved using the observed values of the γ_h vectors across all households with a positive number of choices from the corresponding occasion type k.

As previously mentioned, the vector of probabilities for assignment into the consumption occasion types is the same across demographically similar households[2], and this vector is denoted $pvec_\tau$ for each household type τ. These probabilities are identified across households of the same type that differ in their values of γ_{hk}. Estimation of this vector is important in this analysis, as it these probabilities that will allow marketers to identify the segments that are "important" to a specific demographic.

The intuition behind the above structure is as follows. While a household's preferences for a specific consumption occasion type are fixed, these preferences are a function of many variables, some of which are unobservable. To model the unobserved portion of occasion type/household choices, the iid "error" is added to preferences, as in any discrete choice model. Given the coefficients on utility of individual households over the occasion types, these coefficients (within an occasion type) will be related across households. For example, households purchasing for a "healthy breakfast" will tend to have similar preferences over characteristics of ready-to-eat cereals. These preferences will differ slightly[3] due to heterogeneity in individual tastes, and thus the second vector of errors is added to reflect a distribution of tastes within each occasion type. This distribution is assumed to be multivariate normal, as it makes sense that differences in tastes for one product over another will be correlated across households. Note that estimation of these distributions of household coefficients over an occasion type will provide useful information about the relative importance of various product characteristics within the occasion type, how the coefficients "move" together across households, the relationship of products purchased for the given occasion, and could be used for counterfactual analysis and/or new product introductions. It is also important to note that I make no attempt to identify the actual context of these occasion types, that is, I do not attempt to determine whether the occasion is "healthy breakfast" or "late-afternoon cereal snack." However, given that

[2] For the sake of practical application to identify the preferences of demographic segments, the set of household types are assumed to be pre-specified by the marketer. The model could be extended to make the *pvec* vector draws a function of household demographics.

[3] I use "slightly" to describe these differences, as it is expected that the differences within an occasion type will differ, but to a lesser degree than the preferences across occasion types.

the probabilities of each household belonging to a specific occasion type will also be identified by the model, the occasion types with the largest probabilities (within a targeted demographic) can be further analyzed to discover the structure of the the market for that occasion type[4]. Thus the market structure for "important" occasion types can be analyzed without specifically labeling the occasions.

IV.3 The Simulated Data

To demonstrate the general convergence properties of the model, I simulate a household-level panel dataset and estimate the known value of the given parameters. I simulate 5 brands/products in a category in which there are frequent purchases and significant variety-seeking within individual household purchases, even within occasion types. The simulate product data uses brand intercepts together with two characteristics for each product. Therefore, the design matrix for each product consists of 4 dummy variables and two generated characteristics. The first generated characteristic is assumed to be price, and is simulated accordingly to reflect similarity across households. This was done by randomly drawing a vector of mean prices from a uniform distribution and using these mean prices to create random draws of prices for each product. To realistically reflect price differences across markets, each household is assigned a "price inflator" of 1, 1.1, 1.2, or 1.3, which is used (via multiplication) to change the mean price vector for that household across all choice occasions. The mean and standard deviation of prices across households for each product are shown in table IV.1. The second characteristic is simply a vector of standard uniform draws, and could be interpreted as a measure of promotion, advertising in the area, etc. Given the logit form of the choice probabilities, the design matrix used in the simulation is the matrix of differences between each product and the first product. That is, if x_1 denotes the characteristics of the first product, the characteristics used in the simulation are given by $x_j^d = x_j - x_1$ for all $j \neq 1$.

The household data is simulated using the data generating mechanism described in the previous section, wherein:

i) Some number of purchase choices are observed for each household.

ii) The choices of each household are randomly distributed across a finite number of consumption

[4] It should be noted that given the preferences over the products within an occasion type, it may be possible for a practitioner to draw some conclusions about the context of each occasion type.

Table IV.1: Mean and variance of simulated prices.

	μ_p	σ_p
Product 1	1.3892	0.06
Product 2	1.2931	0.06
Product 3	1.9845	0.06
Product 4	1.7875	0.06
Product 5	1.8337	0.06

Table IV.2: Values of the β vectors used to simulate the data.

	Occasion Type 1	Occasion Type 2
β_1	-0.1434	-0.0031
β_2	-0.1392	-0.2976
β_3	0.4247	-0.5856
β_4	1.1136	-0.9191
β_5	-1.3186	-1.4431
β_6	1.7035	0.7797

Table IV.3: Values of the Σ matrices (and corresponding correlation matrices) used to simulate the data.

Σ_1	ξ_1	ξ_2	ξ_3	ξ_4	Σ_2	ξ_1	ξ_2	ξ_3	ξ_4
ξ_1	0.1175	-0.1317	0.0453	0.0281	ξ_1	0.1477	0.0569	-0.0103	-0.0610
ξ_2		0.4996	-0.0220	-0.0915	ξ_2		0.1630	0.0451	-0.0156
ξ_3			0.1412	0.0847	ξ_3			0.0741	-0.0096
ξ_4				0.1264	ξ_4				0.0764
ρ	ξ_1	ξ_2	ξ_3	ξ_4	ρ	ξ_1	ξ_2	ξ_3	ξ_4
ξ_1	1.0000	-0.5436	0.3518	0.2305	ξ_1	1.0000	0.3524	-0.0981	-0.5746
ξ_2		1.0000	-0.0829	-0.3642	ξ_2		1.0000	0.4103	-0.1395
ξ_3			1.0000	0.6338	ξ_3			1.0000	-0.1280
ξ_4				1.0000	ξ_4				1.0000

occasion types according to a fixed probability vector for each household type.

iii) The choice probabilities for each household at each consumption occasion type are calculated using a random draw from the fixed distribution of preferences (across households) for each consumption occasion type.

iv) The choice for each consumption occasion is randomly assigned using the resulting choice probabilities of the corresponding occasion from step iii) above.

To proceed with the simulation, the number of households and the number of choices are both fixed at 300, representative of a long household panel of a relatively frequently purchased product. Two occasion types are used in the simulation of the choices, and the vector of probabilities of assignment to each consumption occasion type for each household type is calculated from normal-

Table IV.4: Simulated purchase quantities of 4 randomly selected households.

H15	Occasion Type 1	Occasion Type 2	H54	Occasion Type 1	Occasion Type 2
Product 1	21	12	Product 1	9	17
Product 2	22	13	Product 2	8	44
Product 3	19	8	Product 3	31	58
Product 4	54	12	Product 4	47	41
Product 5	131	8	Product 5	41	4

H43	Occasion Type 1	Occasion Type 2	H80	Occasion Type 1	Occasion Type 2
Product 1	7	6	Product 1	13	42
Product 2	50	17	Product 2	7	90
Product 3	13	52	Product 3	27	48
Product 4	52	14	Product 4	24	37
Product 5	83	6	Product 5	8	4

ized vectors of 2 random standard uniform draws. The values of the β vectors for each occasion type are randomly drawn as iid vectors of Uniform(-2,2) draws, and are shown in table IV.2. The values for the Σ matrices are random draws from an inverted Wishart distribution with parameters $\nu = 8$ and $V = 2 * \mathbb{I}_6$, the values of which (together with the corresponding correlation matrices) are shown in table IV.3, yielding a decent amount of correlation in the heterogeneity of tastes across households within the same occasion type. The γ vectors for each consumption occasion type for each household are simulated by multiplying a vector of iid standard normal draws by the transpose of the Cholesky root of the inverse of the corresponding Σ matrix and adding the corresponding β vector[5]. Using the drawn probability of assignment for each household, the household choice occasions are randomly assigned over the two segments using random binomial draws. Using the choice probabilities resulting from the draws of the γ vectors, the choices for each consumption occasion are random draws from a multinomial distribution. The cumulative choices for each consumption occasion type for 4 randomly selected households are presented in table IV.4, demonstrating the presence of significant heterogeneity across households even within a given consumption occasion type.

[5] This is a common way to draw from a multivariate normal with mean β and variance-covariance matrix Σ.

IV.4 Demonstration of the Problem

To demonstrate the seriousness of ignoring within-household heterogeneity, I analyze the simulated dataset described above using a simplified logit choice model. Suppose that a practitioner chooses to model the simulated data with the (single) choice utility for each household defined by

$$u_{hj} = x_j^d \beta + \varepsilon_{hj}$$

where β is similar across all households and ε is distributed iid type I extreme value.Given this structure, if y_t denotes a single choice and $y_{tj} = 1$ indicates that product j is chosen, then the choice probabilities of household h are given by

$$\Pr(y_{jt} = 1 | X_h^d, \beta) = \frac{\exp(x_{hj}^d \beta)}{1 + \sum_{i=2}^N \exp(x_{hi}^d \beta)},$$

where the 1 in the denominator represents the choice utility of product 1 begin fixed at 0 for identification of the model, and X_h^d is the matrix composed of the differenced vectors of product characteristics. Note that each choice probability is essentially (via the modeling assumption) the population proportion of choice occasions over the "infinite" population of choice occasions in which which household h chooses product j. Taking the natural log of this expression yields

$$\frac{\log(\Pr(y_j | X_h^d, \beta))}{\log(\Pr(y_1 | X_h^d, \beta))} = x_{hj} \beta.$$

Therefore, given the large number of choice occasions over the relatively small number of products so that the observed shares of each product for each household are positive, the ratio of observed shares can be calculated as above and the logit model reduces to regressing the ratio of log observed shares on the observed (differenced) data.

Running this regression yields the results presented in table IV.5, wherein the regression could be taken as having a reasonably believable result given an R-squared of nearly 60% and the flat/balanced plot of the residuals versus the fitted values in IV.1. Intuitively speaking, modeling the household with a single set of preferences would seem to estimate the single β vector as an "average" of the true β vectors, and in the case of the (differenced) brand intercept for product 4 (β_3), this is

Table IV.5: Results of the logit regression on simulated purchase shares.

Residuals:

Min	1Q	Median	3Q	Max
-2.00512	-0.29461	-0.02265	0.28144	2.08193

Coefficients:

	Estimate	Std. Error	t-Value	Pr(>|t|)	
β_1	-0.03806	0.0282	-1.349	0.177465	
β_2	-0.11333	0.06262	-1.81	0.070583	.
β_3	-0.00678	0.07309	-0.093	0.926112	
β_4	0.19078	0.05547	3.44	0.000603	***
β_5	-1.16434	0.15466	-7.528	1.01E-13	***
β_6	1.11347	0.03279	33.96	2.00E-16	***

Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 0.4854 on 1194 DF
Multiple R-squared: 0.5848 Adjusted R-squared: 0.5827
F-statistic: 280.3 on 6 , 1194 DF p-value: < 2.2E-16

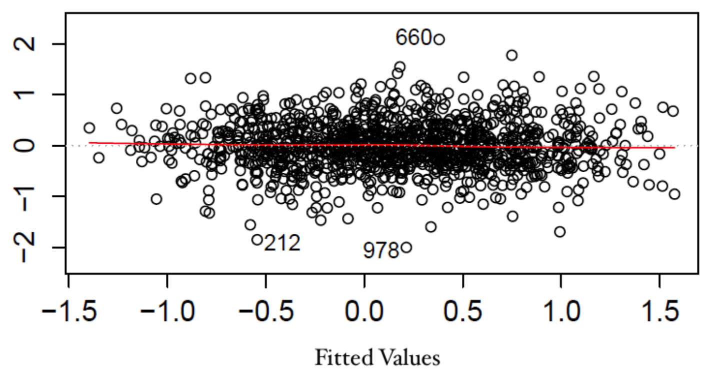

Figure IV.1: Residual plot for the logit regression on observed purchase shares.

true. The estimated value of β_3 is essentially zero, which is roughly half-way between the values of the two true occasion type means in table IV.2. Though an average can often be quite useful, this estimate could yield very poor results if used in the calculation of elasticities for counterfactual analysis on pricing or promotion. To see this, note that the true values of the coefficients imply that most households either significantly prefer product 3 to product 1 or significantly prefer product 1 to product 3, and thus changes in the price or promotions of either product will not induce as great a change in the observed shares as the estimated coefficients would indicate. Even more concerning is the unmodeled correlation that, if picked up by the single estimated β vector, can result in different values than one might expect. For example, the price coefficients (β_5) in the simulated data are

-1.32 and -1.44, respectively, while the estimated coefficient in the simplified logit analysis is -1.16, far from an "average" of the two modes of the occasion types. Therefore, even if a marketer desires only an "average" effect over the true β vectors of household purchases, ignoring within-household heterogeneity for the sake of computational simplicity could yield very misleading estimates. My model is specifically designed to be applied by practitioners to avoid such problems.

IV.5 MCMC Algorithm

Given the structure presented in section 2, the Markov chain Monte Carlo estimation algorithm proceeds as follows:

1. Choose the parameters of the prior distributions for the Σ_k matrices and the β_k vectors. The Σ_k matrices are assumed to be drawn from an inverted Wishart distribution with parameters ν and V_0, and the β_k vectors are assumed to be normally distributed with mean $\overline{\beta}$ and variance A^{-1}.
2. Pick arbitrary starting values of γ_{hk} for each of the k segments for each household h.
3. Begin estimation at the household level, h (this step allows parallelization across households).
    i) Data Augmentation step: Assign each cereal choice, y_t, purchased by household h of demographic type τ to a given occasion type according to the vector of probabilities $pvec_\tau$, augmented by the probability that the choice y_t maximizes the utility of the household given the current value of γ_{hk} for each occasion type. That is, if π_{khi} denotes the probability that the choice y_t of household h with demographic type τ is assigned to occasion type k, and $pvec_{\tau k}$ denotes the element of $pvec_\tau$ that corresponds to occasion type k, and $\Pr(y_t|\gamma_h k)$ denotes the probability that choice y_t would result from the vector of utilities γ_{hk}, then:

    $$\pi_{khi} = \frac{pvec_{\tau k}\Pr(y_t|\gamma_{hk})}{1+\sum_{i=2}^{K} pvec_{\tau i}\Pr(y_t|\gamma_{hi})}$$

    where, given that the ε's are distributed type I extreme value, the choice probabilities $\Pr(y_t|\gamma_{h\cdot})$ are of the logit form. Thus, if $y_{tj} = 1$ corresponds to choosing product j at choice occasion t,

$$\Pr(y_{tj}=1|\gamma_{hk})=\frac{\exp(\gamma_{hkj})}{1+\sum_{i=2}^{J}\exp(\gamma_{hki})}$$

where γ_{hk1} is fixed at 0 for all households. Note that due to the simplicity of these calculations, they are computationally inexpensive and allow the chain to move relatively quickly over each choice of the household.

ii) Given the choices assigned to each of the k occasion types of household h, draw a new value of γ_{hk} using a Metropolis Hastings Gaussian random-walk. To implement this step, a new γ_{hk} is drawn and is either accepted or rejected. Dropping the subscripts, define γ_{old} as the current value of γ_{hk}, and draw a new value via

$$\gamma_{new}=\gamma_{old}+\varepsilon \quad \text{where} \quad \varepsilon \sim N(0,s^2\mathbb{I})$$

with s defined as the step size of the chain. Denoting the set of observed choices of household h assigned to occasion type k as Y_k, the random walk proceeds by calculating

$$\alpha=\min\left\{1,\frac{\Pr(Y_k|\gamma_{new})p(\gamma_{new}|\beta_k,\Sigma_k)}{\Pr(Y_k|\gamma_{old})p(\gamma_{old}|\beta_k,\Sigma_k)}\right\}$$

where $\Pr(Y_k|\gamma_\cdot)$ can be easily calculated given the previously stated logit probabilities using the multinomial distribution,and $p(\gamma_\cdot|\beta_k,\Sigma_k)$ is simply the density of the multivariate normal distribution with parameters β_k and Σ_k evaluated at $\gamma_\cdot$. Given that these are common distributions, most software programs have prewritten functions to calculate these values[6]. Once α is calculated, the new value of γ_{hk} is accepted with probability α, and rejected (i.e. leaving the current value in place) with probability $(1-\alpha)$. Note that the choice of step size, s, is important to how a Metropolis Hastings random-walk functions and must be chosen in such a way as to allow the chain to navigate the space without accepting new values too frequently[7].

[6] The simulations in this paper are run using the R software package, with the function dmultinom being used to calculate the probability of Y_k and lndMvn (from the bayesm package) being used to evaluate the multivariate normal density at $\gamma_\cdot$

[7] The simulations in this paper dynamically adjust the step size s over the first several thousand iterations to achieve acceptance rates around .23, the value suggested by Rossi and Allenby (2003).

4. After drawing the choice assignments and γ vectors for each household, the value of $pvec_{\tau}$ is updated for each household demographic type. This is accomplished via a draw from a Dirichlet distribution with parameters $(a_k + n_k)$ where a_k is the initial value of the Dirichlet parameters and n_k is the number of choices assigned to occasion type k at the current draw. A draw from the corresponding Dirichlet distribution is taken for each household demographic type, which allows the probabilities to differ substantially between household types.

5. Begin estimation at the aggregate, occasion-type, level k (this step allows parallelization across occasion types).

    i) Assuming that the ξ_{hjk}'s are distributed multivariate normal with covariance matrix Σ_k, use the given values of γ_{hk} for the households assigned to occasion type k to draw new values for the β_k's and Σ_k's using the seemingly-unrelated-regression (henceforth SUR) sampler proposed by McCulloch, Polson, and Rossi (2000) for the multinomial probit model. This is implemented by first drawing β_k given the current value of Σ_k and the observed matrix of γ_{hk} vectors, using

    $$\beta_k \sim N(\tilde{\beta}, \tilde{V}),$$

    with

    $$\tilde{V} = (X_{st}' X_{st} + A)^{-1} \quad , \quad \tilde{\beta} = \tilde{V}(X_{st}' \gamma_{st} + A\overline{\beta})$$

    where X_{st} is the stacked matrix of the product characteristics for each household and γ_{st} is the vector composed of the stacked γ_{hk} vectors of the household, with both multiplied by the transpose of the Cholesky root of the inverse of Σ_k. The draw of β is then completed by the common procedure of multiplying a vector of standard normal draws by the Cholesky root of $\tilde{V}$ and adding $\tilde{\beta}$.

    ii) The draw of Σ_k is then completed using an inverted Wishart distribution with parameters $(\nu + n)$ and $(V_0 + S)^{-1}$, where n is the number of households with a positive number of choices assigned to occasion type k, and S is the matrix composed of the sum of the crossproduct of the row of the residual vectors for each household ($S =$

$\sum_{i=1}^{n}(\gamma_{hk} - X_h\beta_k)(\gamma_{hk} - X_h\beta_k)'$). Most popular software programs can provide draws from an inverted Wishart distribution once these parameters are calculated.

6. Repeat steps 2-4 for a sufficient number of draws for the Markov chain to burn-in and converge.

IV.6 Applying the Model to Data

IV.6.1 Application to Simulated Data

The model was applied to the simulated data using the typical "flat" prior distribution parameters of $\overline{\beta} = 0$, $A = .1 * \mathbb{I}_6$, and $\nu = 8$, with the exception being the prior parameter V_0. This parameter for the inverted Wishart distribution is typically chosen to be $\nu * \mathbb{I}_6$ in the SUR regression for the multinomial probit. However, for the purposes of the model presented here, V_0 was set to $\mathbb{I}_6$. The reason for this change is in the modeling assumption that preferences within an occasion type are somewhat similar across households. Larger values for V_0 allow the values on the diagonal of the draws of the Σ matrices (and the subsequent deviations from mean utility across households) to become very large relative to the fixed size of the within-household type I extreme value errors, removing the necessity of similarity across households required by the model (a problem not present in the multinomial probit model). Therefore, the prior was set to a smaller value that still provides ample freedom for the sampler to navigate the space, and is still "diffuse" relative to the expected size of the ξ error terms. The algorithm was run for 100,000 draws using a thinning parameter of 100 and a burn-in of 15,000 draws.

A summary of the posterior draws of the elements of the β vectors and Σ matrices are presented in tables IV.6 and IV.7, respectively, together with the upper and lower bounds for 95% Bayesian credible intervals and the corresponding true values. The algorithm does very well in navigating to the true values for both occasion types with all of the 95% intervals of the elements of the β vectors including their corresponding true values, and all but one of the 95% intervals of the elements of the Σ matrices including their corresponding true values (the single element of the Σ matrix of occasion type 2 that is not included in the interval is still within the range of the post-burn-in draw sequence).

The trace plots for all of these draw sequences are presented in the appendix for the interested reader.

Table IV.6: Summary of the posterior draws for the elements of the β vectors of each occasion type, together with the true values of each element (an "*" indicates that the true value falls within the given 95% Bayesian credible interval).

Occasion Type 1	β_1	β_2	β_3	β_4	β_5	β_6
95% L.Bd. :	-0.2734	-0.5745	-0.0970	0.7563	-2.2728	1.4570
Median :	-0.1348	-0.2866	0.2644	1.0186	-1.5043	1.6191
Mean :	-0.1377	-0.2813	0.2579	1.0237	-1.5060	1.6202
95% U.Bd. :	-0.0102	0.0439	0.5704	1.3157	-0.8193	1.8060
Stand. Dev.:	0.0694	0.1605	0.1699	0.1425	0.3744	0.0923
True :	-.1434*	-.1392*	.4247*	1.1135*	-1.3138*	1.7035*
Occasion Type 2	β_1	β_2	β_3	β_4	β_5	β_6
95% L.Bd. :	-0.1080	-0.4042	-0.7081	-0.9602	-1.8202	0.5478
Median :	-0.0048	-0.1458	-0.4123	-0.7087	-1.2041	0.6863
Mean :	-0.0058	-0.1460	-0.4114	-0.7093	-1.1965	0.6850
95% U.Bd. :	0.0990	0.1071	-0.1246	-0.4515	-0.6004	0.8221
Stand. Dev.:	0.0531	0.1286	0.1456	0.1303	0.3056	0.0679
True :	-.0031*	-.2976*	-.5855*	-.9191*	-1.4431*	.7797*

Table IV.7: Summary of the posterior draws for the elements of the Σ matrix of each occasion type, together with the true values of each element (an "*" indicates that the true value falls within the given 95% Bayesian credible interval).

Occasion Type 1	$\Sigma_{1,1}$	$\Sigma_{1,2}$	$\Sigma_{2,2}$	$\Sigma_{1,3}$	$\Sigma_{2,3}$	$\Sigma_{3,3}$	$\Sigma_{1,4}$	$\Sigma_{2,4}$	$\Sigma_{3,4}$	$\Sigma_{4,4}$
95% L.Bd. :	0.1003	-0.2038	0.2322	-0.0169	-0.0960	0.0999	-0.0245	-0.1732	0.0545	0.1216
Median :	0.1620	-0.1167	0.3686	0.0309	-0.0232	0.1501	0.0271	-0.0789	0.1017	0.1945
Mean :	0.1657	-0.1191	0.3780	0.0326	-0.0224	0.1544	0.0284	-0.0796	0.1079	0.2019
95% U.Bd. :	0.2458	-0.0489	0.5772	0.0858	0.0450	0.2315	0.0933	0.0031	0.1849	0.3156
Stand. Dev.:	0.0390	0.0389	0.0884	0.0264	0.0363	0.0340	0.0303	0.0459	0.0343	0.0487
True :	0.1175*	-0.1317*	0.4996*	0.0453*	0.0220*	0.1412*	0.0281*	-0.0915*	0.0847*	0.1264*
Occasion Type 2	$\Sigma_{1,1}$	$\Sigma_{1,2}$	$\Sigma_{2,2}$	$\Sigma_{1,3}$	$\Sigma_{2,3}$	$\Sigma_{3,3}$	$\Sigma_{1,4}$	$\Sigma_{2,4}$	$\Sigma_{3,4}$	$\Sigma_{4,4}$
95% L.Bd. :	0.0986	0.0288	0.1034	-0.0164	-0.0007	0.0532	-0.1084	-0.1174	-0.0515	0.0845
Median :	0.1432	0.0654	0.1694	0.0115	0.0325	0.0794	-0.0450	-0.0406	-0.0061	0.1715
Mean :	0.1446	0.0664	0.1746	0.0123	0.0340	0.0816	-0.0467	-0.0415	-0.0044	0.1810
95% U.Bd. :	0.2017	0.1098	0.2662	0.0430	0.0749	0.1210	0.0091	0.0228	0.0542	0.3306
Stand. Dev.:	0.0263	0.0208	0.0400	0.0153	0.0190	0.0176	0.0289	0.0351	0.0252	0.0636
True :	0.1477*	0.0547*	0.1630*	-0.0103*	0.0451*	0.0741*	-0.0610*	-0.0156*	-0.0096*	0.0764

A summary of the posterior draws of the values of *pvec* for the 4 aforementioned randomly selected households are presented in table IV.8, together with the bounds of the corresponding 95% Bayesian credible intervals and the corresponding true values. Though more diffuse than the draws of parameters estimated over all households (as expected), the model also does well in picking up these values, with the exception being the draws for simulated household number 80 for which the 95% credible intervals do not include the true values (though the true values are within the range of the post-burn-in draw sequence). Note that, coincidentally, simulated household number 80 is one of only 3 households of the 300 simulated households in which the credible intervals for the draws of the elements of *pvec* do not include the true values.

Table IV.8: Summary of the posterior draws for the elements of *pvec* for four randomly selected households, together with the true values of each element (an "*" indicates that the true value falls within the given 95% Bayesian credible interval).

H15	Occasion Type 1	Occasion Type 2	H54	Occasion Type 1	Occasion Type 2
95% L.Bd. :	0.5586	0.0162	95% L.Bd. :	0.1967	0.1257
Median :	0.7935	0.2065	Median :	0.4388	0.5612
Mean :	0.7902	0.2098	Mean :	0.4681	0.5319
95% U.Bd. :	0.9839	0.4414	95% U.Bd. :	0.8743	0.8033
Stand. Dev. :	0.1153	0.1153	Stand. Dev. :	0.1676	0.1676
True :	0.8067*	0.1933*	True :	0.4942*	0.5058*

H43	Occasion Type 1	Occasion Type 2	H80	Occasion Type 1	Occasion Type 2
95% L.Bd. :	0.4834	0.0063	95% L.Bd. :	0.0023	0.7688
Median :	0.7792	0.2208	Median :	0.0530	0.9471
Mean :	0.7708	0.2292	Mean :	0.0705	0.9295
95% U.Bd. :	0.9934	0.5166	95% U.Bd. :	0.2312	0.9977
Stand. Dev. :	0.1500	0.1500	Stand. Dev. :	0.0618	0.0618
True :	0.6343*	0.3657*	True :	0.2766	0.7234

Given the structure of the model, one may think that the algorithm would have difficulty identifying the parameters in datasets for which there are strong differences in preferences between occasion types (i.e. large differences in the values of the β vectors). The intuition behind this being that the β vectors could "soak up" the variation in purchases, leaving very little heterogeneity for identification of the Σ matrices. To investigate this, the results of another simulation, run on three occasion types, are presented in tables IV.10, IV.11, and IV.12 in the appendix. The β vectors used in this simulation are presented in table IV.9, and are clearly representative of very strong preferences for different products across the 3 occasion types.

As shown in the appendix, the results of 100,000 draws with a burn-in of 15,000 draws are still very good. All of the 95% credible intervals of the elements of the β vectors of the three occasion types include their corresponding true values, and while the Σ results are slightly weaker given the strong effects from the β vectors, the sampler still consistently navigates to the true values. All but 6 of the 95% credible intervals of the elements of the Σ matrices for the 3 occasion types include their

Table IV.9: Values of the β vectors used to simulate the 3 occasion type data.

	Occasion Type 1	Occasion Type 2	Occasion Type 3
β_1	-1.5000	0.2500	2.0000
β_2	-1.5000	0.2500	-0.5000
β_3	1.7500	-2.0000	-0.5000
β_4	1.7500	-2.0000	2.0000
β_5	-1.0000	-0.7500	-0.7500
β_6	0.7500	0.5000	0.7500

corresponding true values, with all 30 of the true values being within the range of the post burn-in draw sequences. Finally, the *pvec* draws for the 4 randomly selected households are also quite good, with all intervals for these elements including their corresponding true values. Moreover, only 1 element of the *pvec* vector for two household types (out of the 90 total simulated elements across all household types) is outside of the corresponding credible interval, indicating that the model again does an impressive job of picking up the underlying true values. If, however, a practitioner is concerned that the model may not pick up the small covariance parameters of the distributions across occasion types, the model can easily be changed to include only a single Σ distribution across all occasion types, with the assumption being that correlation in household's tastes across products are relatively similar across occasions (a strong, but not unrealistic assumption).

IV.6.2 Application to a Household Panel

As shown above, the model will perform quite well on a variety of household panel datasets in which the practitioner believes the data-generating process to be in line with the assumptions laid out in section 2. The remaining issue is choosing the number of segments with which to model the given data. For this purpose, computation of the model likelihood can be used for comparison of models with a different number of specified occasion types. Though more complex than a basic LCM or finite mixture model, the logit choice probabilities do allow direct calculation of the conditional likelihood of the observed data given the value of the parameters. The first step to calculate the conditional likelihood is to calculate the vector of probabilities that a single consumption occasion for a given household results in the selection of each product. Given the design matrix X_h^d, the values of the γ_{hk} vectors for household h (henceforth denoted γ_h) , and the value of *pvec* for household h, the probability that product j is chosen at choice occasion t is then

$$\Pr(y_{tj}=1|X_h^d,\gamma_h,pvec)=\prod_{i=1}^{K}\Pr(y_{tj}=1|k=i,X_h^d,\gamma_{hi})\Pr(k=i|pvec).$$

Once these probabilities have been calculated (using the previously provided formula for logit choice), the distribution of the observed purchase quantities for household h, conditional on γ_h and $pvec$, is multinomial (via the modeling assumption that each unit results from an independent choice). Therefore, the probability of the observed choice data for household h conditional on γ_h and $pvec$, denoted $\Pr(\text{data}_h|X_h^d,\gamma_h,pvec)$, can be easily calculated for each household using built-in functions for multinomial probabilities. As this probability is conditional on parameter values, the conditional likelihood of these parameter values are also included in the full conditional likelihood calculation for each household. After calculating $p(\gamma_{hk}|\beta_k,\Sigma_k)$ for each occasion type k (via the evaluation of a multivariate normal density discussed in section 4), the conditional likelihood of the observed data for each household h given the parameter values can easily be calculated as

$$\ell(\text{data}_h|X_h^d,\gamma_h,\beta_h,\Sigma_h,pvec)=\Pr(\text{data}_h|X_h^d,\gamma_h,pvec)\prod_{i=1}^{K}p(\gamma_{hi}|\beta_i,\Sigma_i)$$

where $\beta_h=\{\beta_1,\beta_2,...,\beta_K\}$ and $\Sigma_h=\{\Sigma_1,\Sigma_2,...,\Sigma_K\}$. Given this likelihood for each household, the (log) conditional likelihood for the observed choice data given the parameter values is then

$$\ell(\text{Data}|X_h^d,\gamma_h,\beta_h,\Sigma_h,pvec)=\sum_{i=1}^{H}\log(\ell(\text{data}_h|X_h^d,\gamma_h,\beta_h,\Sigma_h,pvec)).$$

Despite the nested calculations, this likelihood is fairly inexpensive computationally as almost all of the above calculations are required at some point during the sampling process of each iteration. To compare two given models (with different numbers of segments), a practitioner may either compare the sequences of log likelihoods given above (stored at each iteration) as a gauge of the fit of each model, or use the method of Newton and Raftery (1994) to calculate the log marginal density for each model (a built-in function which is included in R and other popular software programs) as a comparison of fit.

Another issue that should be addressed is the possibility of "label switching" over the occasion types. Given that each occasion type of the model has the same prior, it is possible (as in all mixture models) that the occasion types could switch "labels." That is, two (or more) occasion types switch the preferences with which they are identified and continue in the draw sequence. This can typically be identified via the trace plots of the posterior draws, as the parameter values of two or more occasions will appear to "switch places" through the sequence of draws. Though label switching is not of great concern in my model due to the mitigating effects of Metropolis Hastings (preventing large changes in the γ values of each occasion type between iterations) and the connection of households through *pvec* (preventing large changes in *pvec* between iterations[8]), there are methods for handling label switching that I will discuss for the sake of being thorough in my presentation of a mixture-driven model.

There are two popular ways to handle the label switching problem: post-processing the draws and constraining the segments. As mentioned above, the iteration at which a switch of label occurs can be (roughly) identified via the posterior draw sequence. Post-processing consists of marking these events in the chain and relabeling the segments according to the "new" labels. Though theoretically simple, post processing can be quite challenging in practice as developing a strategy for deciding exactly when a label switch has occurred is not a trivial exercise. The latter option, constraining the segments, prevents label switching from occurring in the first place and avoids the challenges of the post-processing method. In this latter method, some of the parameters are constrained across segments, preventing the draws of these parameters from switching between those segments. While this may seem to defeat the purpose of the mixture model, these constraints need not be heavy handed. For example, if one is attempting to model health-related segments of a category, a simple ordering constraint over the value of the coefficient on fat content (across the segments of the model) may alleviate any observed label switching in a model without constraints. Moreover, it is typical that practitioners will have some information about the category they seek to analyze, and parameter constraints can be an effective way of incorporating this information into the model. Implementation can be as simple as redrawing parameters that do not meet the constraints, making such constraints a quick and easy way to handle label switching when it does occur.

[8] The Gibbs sampler typically used in Bayesian mixture models can allow large changes between iterations, leading to label switching in the draw sequence. My model uses a hybrid sampler, including both Gibbs and Metropolis Hastings steps, and (together with the modeling assumptions) is much less likely to struggle with label switching.

IV.7 Conclusion

The goal of this paper was to develop a method, applicable to large household panel datasets, for investigating preferences over products in a category in such a way as to allow household purchases to be the result of multiple sets of preferences while also allowing for recovery of household-level information useful for targeted marketing. The presented hierarchical Bayesian finite mixture model is both logically and computational attractive, allows for application to large complex datasets such as ready-to-eat cereal and soft drinks, and allows for useful household-level information about preferences over different types of purchases to be recovered where other (previously presented) models do not. Results applying the model to multiple simulated datasets demonstrates the ability of the Markov chain Monte Carlo algorithm to pick up the true underlying parameters, yielding a model that is both logically and computationally attractive for estimating the components of household demand.

IV.8 Appendix

Occasion Type 1	β_1	β_2	β_3	β_4	β_5	β_6
95% L.Bd. :	-2.0455	-1.6168	0.9906	0.5853	-1.3918	0.4779
Median :	-1.1017	-0.9205	1.3737	1.1607	-0.2120	0.6741
Mean :	-1.1270	-0.9604	1.3961	1.1886	-0.2397	0.6826
95% U.Bd. :	-0.4612	-0.5666	1.9304	1.9414	0.8075	0.9365
Stand. Dev. :	0.3959	0.2557	0.2481	0.3583	0.5719	0.1168
True :	-1.5000*	-1.5000*	1.7500*	1.7500*	-1.0000*	0.7500*
Occasion Type 2	β_1	β_2	β_3	β_4	β_5	β_6
95% L.Bd. :	0.2417	0.0510	-3.4094	-2.1685	-1.9927	0.3672
Median :	0.6130	0.1668	-2.1129	-1.2204	-1.3321	0.5275
Mean :	0.6099	0.1658	-2.1786	-1.2319	-1.3215	0.5273
95% U.Bd. :	0.9623	0.2724	-1.3473	-0.3536	-0.6026	0.6801
Stand. Dev. :	0.1847	0.0585	0.5421	0.4826	0.3475	0.0762
True :	0.2500*	0.2500*	-2.0000*	-2.0000*	-0.7500*	0.5000*
Occasion Type 3	β_1	β_2	β_3	β_4	β_5	β_6
95% L.Bd. :	1.4332	-0.7482	-1.0886	1.4271	-1.9952	0.4867
Median :	2.0276	-0.2611	-0.1670	2.1156	-0.8863	0.7529
Mean :	2.0665	-0.2762	-0.1838	2.1337	-0.8755	0.7583
95% U.Bd. :	2.8237	0.0763	0.7841	2.9955	0.1715	1.0267
Stand. Dev. :	0.3650	0.2359	0.4736	0.4138	0.5404	0.1302
True :	2.0000*	-0.5000*	-0.5000*	2.0000*	-0.7500*	0.7500*

Table IV.10: Summary of the posterior draws for the elements of the β vectors of each occasion type in the 3 occasion type dataset, together with the true values of each element (an "*" indicates that the true value falls within the given 95% Bayesian credibility interval).

Occasion Type 1	$\Sigma_{1,1}$	$\Sigma_{1,2}$	$\Sigma_{2,2}$	$\Sigma_{1,3}$	$\Sigma_{2,3}$	$\Sigma_{3,3}$	$\Sigma_{1,4}$	$\Sigma_{2,4}$	$\Sigma_{3,4}$	$\Sigma_{4,4}$
95% L.Bd. :	0.1156	-0.3109	0.1432	-0.1580	-0.0351	0.0924	-0.1683	-0.1444	-0.0341	0.1202
Median :	0.2753	-0.0794	0.3064	-0.0361	0.0560	0.1392	-0.0301	-0.0234	0.0090	0.1926
Mean :	0.3159	-0.0861	0.3296	-0.0382	0.0596	0.1426	-0.0325	-0.0222	0.0123	0.1978
95% U.Bd. :	0.7182	0.1101	0.6193	0.0535	0.1639	0.2147	0.0865	0.1001	0.0724	0.2990
Stand. Dev.:	0.1566	0.1034	0.1288	0.0517	0.0491	0.0319	0.0632	0.0618	0.0268	0.0470
True :	0.8591	-0.3687	0.3069*	-0.1959	0.1079*	0.1456*	0.0544*	-0.0851*	0.0026*	0.1666*
Occasion Type 2	$\Sigma_{1,1}$	$\Sigma_{1,2}$	$\Sigma_{2,2}$	$\Sigma_{1,3}$	$\Sigma_{2,3}$	$\Sigma_{3,3}$	$\Sigma_{1,4}$	$\Sigma_{2,4}$	$\Sigma_{3,4}$	$\Sigma_{4,4}$
95% L.Bd. :	0.0520	0.0096	0.2101	-0.1068	-0.3232	0.1399	-0.2046	-0.4628	-0.0456	0.4665
Median :	0.0828	0.0460	0.2802	-0.0282	-0.1505	0.3424	-0.0781	-0.2030	0.2984	1.1137
Mean :	0.0849	0.0470	0.2842	-0.0300	-0.1542	0.3900	-0.0829	-0.2153	0.3151	1.1803
95% U.Bd. :	0.1309	0.0873	0.3799	0.0365	-0.0118	0.8764	0.0198	-0.0302	0.7352	2.3128
Stand. Dev.:	0.0199	0.0206	0.0450	0.0368	0.0788	0.1944	0.0572	0.1100	0.2000	0.4828
True :	0.1057*	0.0566*	0.2599*	0.0245*	-0.0341*	0.1114	-0.1282*	-0.2447*	0.1408*	1.7405*
Occasion Type 3	$\Sigma_{1,1}$	$\Sigma_{1,2}$	$\Sigma_{2,2}$	$\Sigma_{1,3}$	$\Sigma_{2,3}$	$\Sigma_{3,3}$	$\Sigma_{1,4}$	$\Sigma_{2,4}$	$\Sigma_{3,4}$	$\Sigma_{4,4}$
95% L.Bd. :	0.0668	-0.0822	0.0963	-0.0291	-0.1712	0.0942	0.0000	-0.1108	-0.0299	0.1134
Median :	0.1310	0.0098	0.2135	0.0564	0.0111	0.2346	0.0616	0.0183	0.0742	0.2076
Mean :	0.1425	0.0109	0.2356	0.0668	0.0141	0.2555	0.0773	0.0228	0.0902	0.2274
95% U.Bd. :	0.2946	0.1226	0.5037	0.2119	0.1955	0.5745	0.2539	0.1840	0.2760	0.4759
Stand. Dev.:	0.0569	0.0523	0.1032	0.0664	0.0895	0.1278	0.0641	0.0732	0.0831	0.0910
True :	0.0616	-0.0244*	0.1987*	-0.0206*	0.0071*	0.1807*	0.0193*	-0.1244	0.0344*	0.2475*

Table IV.11: Summary of the posterior draws for the elements of the Σ matrix of each occasion type in the 3 occasion type dataset, together with the true values of each element (an "*" indicates that the true value falls within the given 95% Bayesian credibility interval).

H15	Occasion Type 1	Occasion Type 2	Occasion Type 3	H54	Occasion Type 1	Occasion Type 2	Occasion Type 3
95% L.Bd. :	0.4461	0.1976	0.1327	95% L.Bd. :	0.1295	0.6840	0.0272
Median :	0.5172	0.2707	0.2094	Median :	0.1704	0.7378	0.0890
Mean :	0.5175	0.2705	0.2120	Mean :	0.1708	0.7385	0.0907
95% U.Bd. :	0.5912	0.3399	0.3049	95% U.Bd. :	0.2151	0.7979	0.1648
Stand. Dev. :	0.0367	0.0372	0.0427	Stand. Dev. :	0.0226	0.0290	0.0359
True :	0.4571*	0.2888*	0.2541*	True :	0.1428*	0.7526*	0.1046*

H43	Occasion Type 1	Occasion Type 2	Occasion Type 3	H80	Occasion Type 1	Occasion Type 2	Occasion Type 3
95% L.Bd. :	0.0054	0.2846	0.4570	95% L.Bd. :	0.5101	0.0064	0.2682
Median :	0.0448	0.4013	0.5586	Median :	0.5855	0.0678	0.3463
Mean :	0.0438	0.3976	0.5586	Mean :	0.5842	0.0694	0.3465
95% U.Bd. :	0.0779	0.4935	0.6682	95% U.Bd. :	0.6459	0.1493	0.4208
Stand. Dev. :	0.0196	0.0559	0.0589	Stand. Dev. :	0.0348	0.0349	0.0404
True :	0.0596*	0.4332*	0.5071*	True :	0.5575*	0.1045*	0.3380*

Table IV.12: Summary of the posterior draws for the elements of *pvec* for four randomly selected households in the 3 occasion type dataset, together with the true values of each element (an ”*” indicates that the true value falls within the given 95% Bayesian credibility interval).

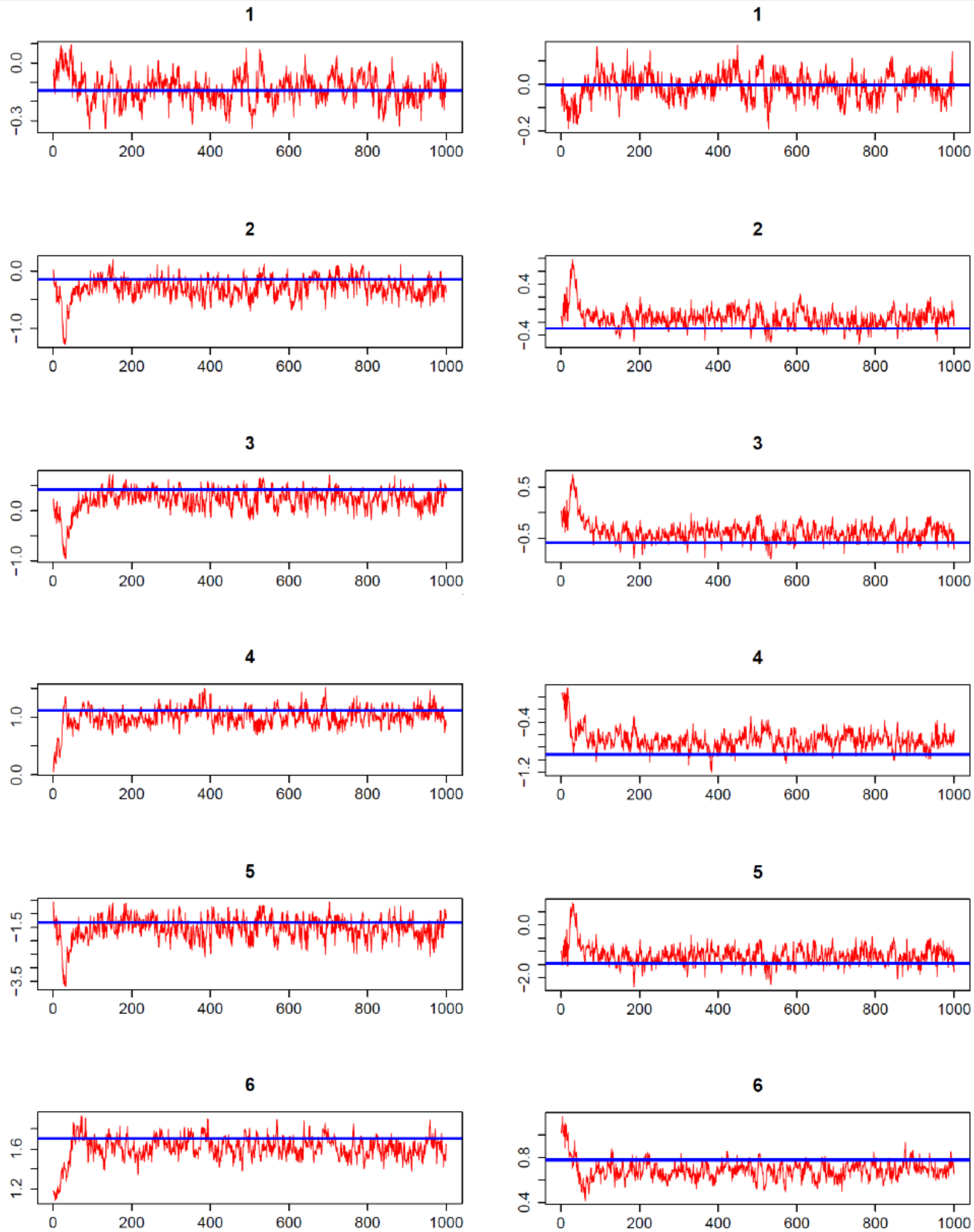

Table IV.13: Trace plots for the posterior draws of the elements of the β vector for both occasion types (type 1 on the left).

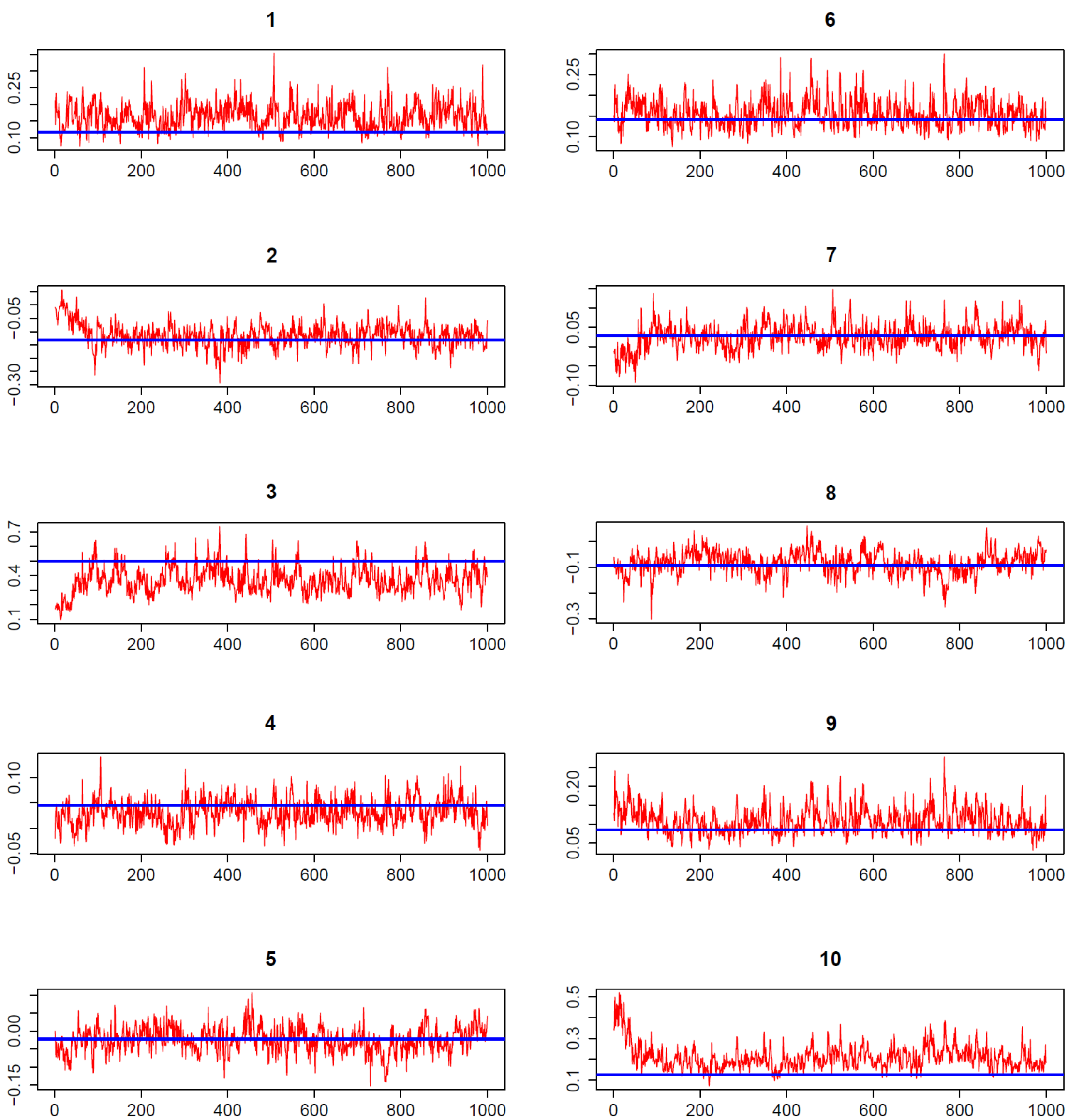

Table IV.14: Trace plots of the posterior draws of each element of the Σ matrix for occasion type 1.

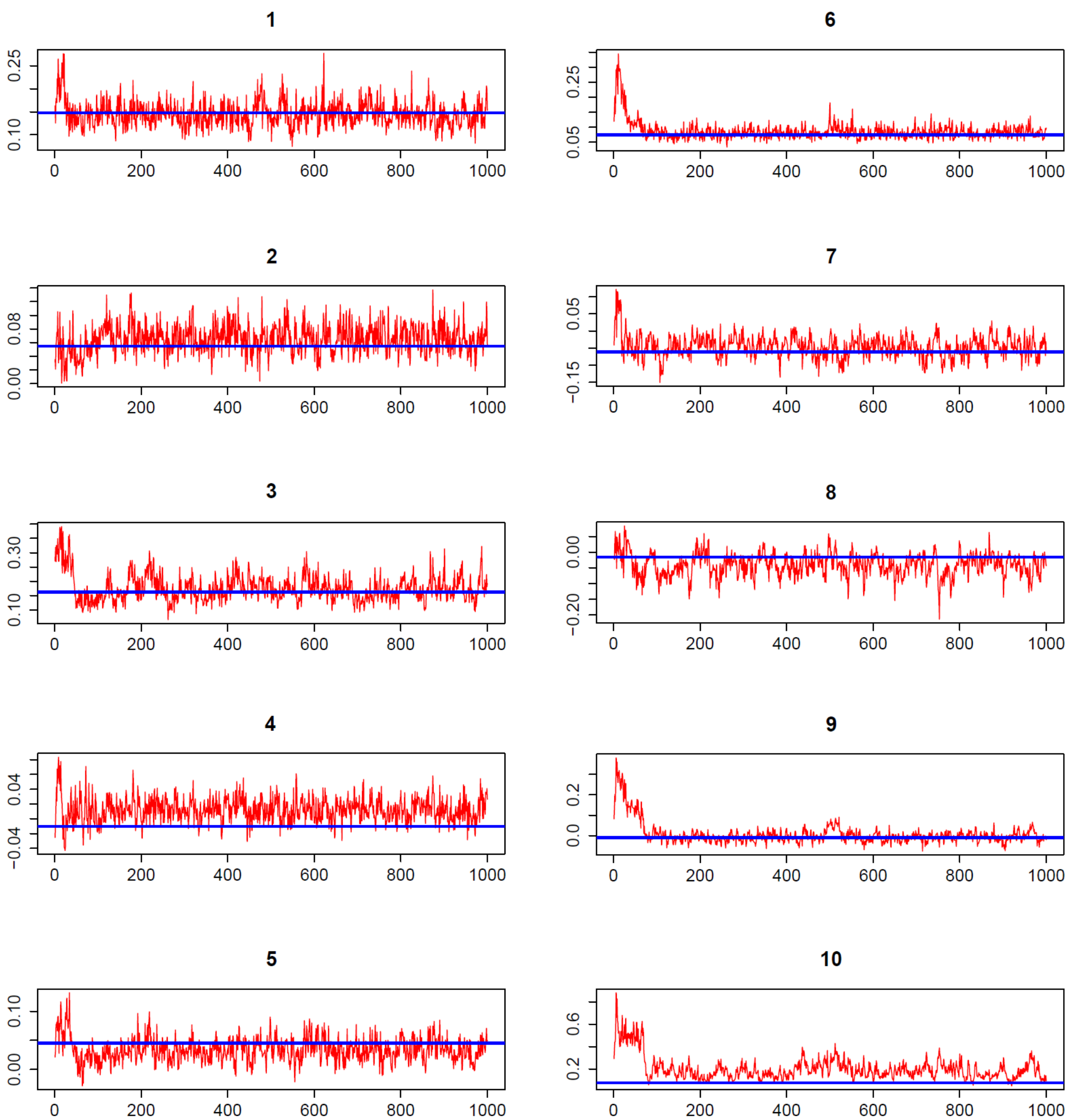

Table IV.15: Trace plots of the posterior draws of each element of the Σ matrix for occasion type 2.

REFERENCES

Allenby, G. M., N. Arora, and J. L. Ginter (1998, Aug.). On the heterogeneity of demand. *Journal of Marketing Research* (3), pp. 384–389.

Allenby, G. M. and P. E. Rossi (1998). Marketing models of consumer heterogeneity. *Journal of Econometrics 89*(12), 57 – 78.

Andrews, J. C., S. Burton, and R. G. Netemeyer (2000). Are some comparative nutrition claims misleading? the role of nutrition knowledge, ad claim type and disclosure conditions. *Journal of Advertising 29*(3), 29–42.

Bagwell, K. (1992). Pricing to Signal Product Line Quality. *Journal of Economics and Management Strategy 1*, 151–174.

Bagwell, K. and M. H. Riordan (1991, March). High and Declining Prices Signal Product Quality. *American Economic Review 81*(1), 224–39.

Bar-Isaac, H., G. Caruana, and V. Cunãt (2010). Information gathering and marketing. *Journal of Economics & Management Strategy 19*(2), 375–401.

Bar-Isaac, H., G. Caruana, and V. Cunãt (2012). Information gathering externalities for a multi-attribute good. *The Journal of Industrial Economics 60*(1), 162–185.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics 25*(2), pp. 242–262.

Bester, H. and K. Ritzberger (2001). Strategic pricing, signalling, and costly information aquisition. *International Journal of Industrial Organization 19*, 1347–1361.

Cho, I.-K. and D. M. Kreps (1987, May). Signaling Games and Stable Equilibria. *The Quarterly Journal of Economics 102*(2), 179–221.

Daughety, A. F. and J. F. Reinganum (1995, December). Product Safety: Liability, R & D, and Signaling. *American Economic Review 85*(5), 1187–1206.

Daughety, A. F. and J. F. Reinganum (1997). Everybody Out of the Pool: Products Liability, Punitive Damages, and Competition. *Journal of Law, Economics, and Organization 13*(2), 410–432.

Daughety, A. F. and J. F. Reinganum (2005). Secrecy and Safety. *The American Economic Review 95*(4), 1074–1091.

Daughety, A. F. and J. F. Reinganum (2007). Competition and Confidentiality: Signaling Quality in a Duopoly when There is Universal Private Information. *Games and Economic Behavior 58*, 94–120.

Daughety, A. F. and J. F. Reinganum (2008a). Products Liability, Signaling and Disclosure. *Journal of Institutional and Theoretical Economics 164*(1), 106 – 126.

Daughety, A. F. and J. F. Reinganum (2008b). Imperfect Competition and Quality Signaling. *RAND Journal of Economics 39*, 163–183.

Daughety, A. F. and J. F. Reinganum (2008c). Communicating quality: a unified model of disclosure and signalling. *RAND Journal of Economics 39*(4), 973–989.

Diebolt, J. and C. P. Robert (1994). Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(2), pp. 363–375.

Downs, J. S., G. Loewenstein, and J. Wisdom (2009, May). Strategies for promoting healthier food choices. *The American Economic Review 99*(2), pp. 159–164.

Dranove, D. and G. Z. Jin (2010). Quality Disclosure and Certification: Theory and Practice. *Journal of Economic Literature 48*(4), 93563.

Dubé, J.-P. (2004). Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks. *Marketing Science 23*(1), 66–81.

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica 80*(5), 2231–2267.

Ellingsen, T. (1997). Price Signals Quality: The Case of Perfectly Inelastic Demand. *The Journal of Industrial Organization 16*, 43–61.

Fishman, M. J. and K. M. Hagerty (2003, April). Mandatory Versus Voluntary Disclosure in Markets with Informed and Uninformed Customers. *Journal of Law, Economics and Organization 19*(1), 45–63.

Gentzkow, M. (2007). Valuing New Goods With a Model of Complementarity: Online Newspapers. *American Economic Review 97*, 713–744.

Grossman, S. J. (1981). The Informational Role of Warranties and Private Disclosure about Product Quality. *Journal of Law and Economics 24*(3), pp. 461–483.

Janssen, M. and S. Roy (2010). Signaling Quality Through Prices in an Oligopoly. *Games and Economic Behavior 68*, 192–207.

Jovanovic, B. (1982). Truthful Disclosure of Information. *The Bell Journal of Economics 13*(1), 36–44.

Kamakura, W. A. and G. J. Russell (1989, Nov.). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research 26*(4), pp. 379–390.

Kozup, J. C., E. H. Creyer, and S. Burton (2003, Apr.). Making healthful food choices: The influence of health claims and nutrition information on consumers' evaluations of packaged food products and restaurant menu items. *Journal of Marketing 67*(2), pp. 19–34.

McCulloch, R. E., N. G. Polson, and P. E. Rossi (2000). A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics 99*(1), 173 – 193.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142.

Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics 12*(2), pp. 380–391.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica 69*(2), 307–342.

Nevo, A., D. Rubinfield, and M. McCabe (2005). Academic Journal Pricing and the Demand of Libraries. *American Economic Review 95*, 447–452.

Newton, M. A. and A. E. Raftery (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(1), pp. 3–48.

Raghunathan, R., R. W. Naylor, and W. D. Hoyer (2006, Oct.). The unhealthy = tasty intuition and its effects on taste inferences, enjoyment, and choice of food products. *Journal of Marketing 70*(4), pp. 170–184.

Reynarts, J., R. Varadhan, and J. C. Nash (2012). Enhancing the convergence properties of the blp (1995) contraction mapping. *Vives Discussion Paper 35*.

Rossi, P. E. and G. M. Allenby (2003). Bayesian statistics and marketing. *Marketing Science 22*(3), 304–328.

Varki, S. and P. K. Chintagunta (2004). The Augmented Latent Class Model: Incorporating Additional Heterogeneity in the Latent Class Model for Panel Data. *Journal of Marketing Research 41*(2), 226–233.