Essays in Industrial Organization

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Pik Yi Lydia Cheung

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Patrick L. Bajari, Advisor

August 2012

Acknowledgements

I thank my advisor Patrick Bajari for his excellent guidance. I also thank the members of my committee: Amil Petrin, Thomas Holmes, and Joel Waldfogel for their invaluable help.

Dedication

For my parents.

For K, my reason and reward.

Abstracts

The Upward Pricing Pressure Test: An Empirical Examination

The Upward Pricing Pressure (UPP) test developed by antitrust economists Joseph Farrell and Carl Shapiro marks a new era in antitrust and provides an alternative to the traditional concentration-based tests in merger analysis. In addition to being free of market definition, the UPP's appeal lies in its ease of use: one simple formula indicates whether a merging firm has an incentive to increase prices post-merger. This paper first establishes the theoretical relationship between the UPP and the standard structural merger simulation, namely, that the UPP is a "single-product merger simulation" that ignores the re-equilibration of all other endogenous variables except that product's own price. To assess the consequence of this simplification, I compute "true" UPP values for a cross-section of airline markets using structurally estimated price elasticities, and confront them with the "gold standard" of a merger simulation. I examine the predictive accuracy of both the *sign* and *magnitude* of the UPP. I find that it gives wrong sign predictions to an average 10% of the observations, and its value has an average correlation of 0.92 with the structurally simulated price changes. However, since this test is meant to bypass a complicated demand estimation, I then use the example of a simple logit demand to illustrate the consequence of using inaccurate demand-side inputs in the UPP: the test will give a wrong sign prediction over a much larger range of cost synergies. Lastly, I discuss the pass-through conditions for Farrell and Shapiro's proposition, demonstrate empirically that they are not innocuous, and show that their violation can lead to false positive results (type I errors) in the UPP.

Brand Portfolio and Consumer Learning

I present a structural model in which consumers learn about the unobserved quality of a brand through purchases of its products in different categories, and examine whether the variety in a brand's portfolio—whether the brand concentrates on very similar products or covers a wide range of product categories—affects consumer learning. I model consumer learning explicitly using a Bayesian updating mechanism. I use supermarket scanner data in the salted snack category to estimate an alternative-specific conditional (ASC) logit model of demand, incorporating brand portfolio variation and

consumer learning as reduced form inputs to the consumer's brand choice problem. The estimation result verifies that a broad portfolio, consisting of a large number of snack categories, decreases the probability of purchase, holding other demand factors constant. This suggests that a varied brand portfolio is less advantageous for consumer learning than a concentrated portfolio. A forthcoming accompanying paper will estimate the full structural Bayesian model and conduct counterfactual experiments.

Contents

List of Tables

List of Figures

1 The Upward Pricing Pressure Test: An Empirical Examination

1.1 Introduction

The year 2010 marked a distinct shift in merger analysis from market concentration to price effects. In contrast to the traditional approach that relies on the increase in market shares, this new approach focuses on various price changes induced by the merger. As leaders in this paradigm shift, the Department of Justice and the Federal Trade Commission recently released the new *Horizontal Merger Guidelines*, and the Competition Commission and the Office of Fair Trading in the U.K. revised their *Merger Assessment Guidelines*. Similar revisions are in progress in the European Union and New Zealand. These revisions have substantial impact on businesses and the antitrust community: the U.S. antitrust agencies received 1166 Hart-Scott-Rodino pre-merger notifications in the year 2010, each of which involved a transaction value of at least $50 million.[1] Joseph Farrell of the FTC and Carl Shapiro of the DOJ directed this shift in merger analysis by their proposal of the Upward Pricing Pressure (UPP) test. It identifies a firm's incentive to raise prices post-merger by comparing its incentive to increase prices due to lost competition and the opposing incentive to decrease prices due to cost synergies. The UPP test has since generated great interest among antitrust practitioners and economists alike and has been incorporated into the two revised *Guidelines*. This paper is the first to examine the empirical performance of the UPP test. I compare the UPP's predictions against those from a standard structural merger simulation and explore potential challenges in its application, including the use of inaccurate diversion ratios and the violation of pass-through assumptions.

The UPP is designed to facilitate the antitrust community's shift away from the traditional merger analysis practice based primarily on market definition and market shares. In the previous (1992) edition of the *Horizontal Merger Guidelines*, for example, the stated procedure in analyzing a merger is to first define the boundary of the antitrust market, calculate the market shares of respective firms or brands, and finally,

[1] Department of Justice and Federal Trade Commission (2010a)

compute the increase in a market concentration index (such as the Herfindahl index (HHI) or the four-firm concentration ratio) caused by the merger. Thresholds are assigned to the index based on pre-merger number of firms and/or pre-merger market concentration. This approach based on market shares has a few problems. Firstly, consumer welfare can increase with concentration in some theoretical settings. Farrell and Shapiro (1990) give an example where consumer or total surplus is enhanced with an increase in market concentration: when production is shifted towards larger, more efficient firms. Secondly, it is based on the old Structure-Conduct-Performance approach in I.O., which researchers have shown to be problematic. When market structure is potentially endogenous, one cannot easily establish a causal relationship from observed structure to performance. Thirdly, the choice of market boundary is often objective; any two differentiated products may arguably have some degree of substitutability for certain consumers. Yet, market definition has been seen as the core of the debate in some cases, for example *FTC vs. Whole Foods (2007)*. Whether the relevant antitrust market is the narrowly defined "premium natural and organic supermarkets" or, more broadly, "all supermarkets" affects the measurement of market shares drastically. In contrast to the traditional HHI test that is based on Cournot competition between homogeneous goods, the new UPP test is based on Bertrand theory that accommodates differentiated products, which is deemed more appropriate for most consumer products. Another purpose of the UPP is to serve as a fast, easy preliminary merger screen that identifies anticompetitive merger markets without heavy data requirement. The antitrust agencies in the U.S. are given 30 days to analyze a potential merger upon receiving its notification, with little data available before deciding whether to issue a second request. The UPP a fitting tool in these regards: firstly, only one simple inequality needs evaluated for each merging firm; secondly, it requires data on substitution patterns, prices, and marginal costs only. In bypassing a structural demand estimation, the UPP has the additional appeal of "transparency" and independence from demand functional form assumptions, eliminating the possibility of functional form mis-specification.

In this paper, I first establish the theoretical relationship between the UPP and the standard structural merger simulation: the UPP is a much-simplified merger simulation that allows only the price of *one* product to re-equilibrate, while holding all other endogenous prices, quantities, and elasticities constant at pre-merger levels. In other words, when the UPP uses "true" substitution patterns computed from estimated demand parameters, it differs from a merger simulation only because of the former's non-reequilibration assumptions. I then use the set of overlapping routes in the America West–US Airways merger (2005) as an empirical laboratory to test the performance of the UPP against merger simulations. I demonstrate that when the UPP is used *in conjunction* with a structural demand estimation, the UPP produces accurate predictions, both in sign and magnitude. Thus the non-reequilibration assumptions inherent in the UPP in fact do not severely affect the sign or magnitude of its predictions. This means that the UPP can be used to substitute the computation of the post-merger equilibria (from the set of firms' post-merger first order conditions).

Secondly, I explore the correlation between the *magnitude* of the UPP against the magnitude of the simulated price change, as the percentage of merger-induced cost reduction e changes. I show empirically that this correlation increases with cost synergy e. I explain this by decomposing the post-merger price change into two components, analyzing the severity of the approximation errors in both, and the change in relative importance between them as e increases. Overall, across the full range of values of e (from zero to 0.1) that I tested, the correlations between the magnitudes of the UPP and the simulated price changes are remarkably high; the magnitudes of these two variables are also similar in range. These observations should encourage the use of the magnitude of the UPP as an approximated price increase.

Thirdly, I investigate the performance of the UPP *without* accurate demand-side estimates, because Farrell and Shapiro do not originally intend the UPP to be used together with a structural demand estimation. Using simple logit substitution patterns (that do not require the estimation of a demand model) instead, I demonstrate that the test computed with imperfect demand-side inputs gives a wrong sign prediction over a much larger range of cost synergies e. This paper is the first to document this phenomenon, and my explanation relates the UPP to the literature on cost pass-through. I show that because pass-through is high and almost constant throughout the relevant range of costs, there always exists a problematic range of e where the UPP gives the wrong sign prediction. I also illustrate this relationship between demand inputs, cost synergy e, and sign prediction graphically.

Fourthly, I empirically show that the pass-through assumptions in Farrell and Shapiro's main proposition are not innocuous. I point out that these assumptions are equivalent to requiring the two merging products to be strategic complements post-merger. It has been theoretically proven that two differentiated products in a Bertrand model can potentially be strategic complements or substitutes, depending on both the firms' cost structures and the demand model used. In particular, in the case of constant marginal cost assumed by the UPP, both strategic relationships are possible. I demonstrate with my merger simulations that the violation of the strategic complement relationship increases the UPP's tendency to produce false positive results (type I errors). This is intuitive because, when two goods are strategic substitutes, an aggressive behavior (e.g. a price increase) from one good will lead to an opposite behavior (e.g. a price decrease) in the other. This pass-through assumption has so far received little attention in the discussion of the UPP. I also generalize Farrell and Shapiro's proposition to the case when the pre-merger market is larger than a duopoly, which leads to a larger set of pass-through assumptions. I show empirically that the proposition's conclusion may not hold when this expanded set of assumptions is not satisfied.

Lastly, I compare the predictions from the HHI (Herfindahl-Hirschman Index), the traditional approach based on market shares, against the new UPP and merger simulation. I first document that when the UPP and merger simulation use Farrell and Shapiro's "default" 10% cost deduction, they flag few markets that are deemed anticompetitive by the HHI, using thresholds from either the 1992 or 2010 *Guidelines*. Despite their different conclusions, I show that the increase in HHI is nonetheless positively but moderately correlated with the former two predictions when they do not include a cost reduction, because the HHI has no inherent capacity to take into account changes in costs. In this sense, the UPP is a more flexible tool (that also requires more data inputs) than the HHI. This correlation between the UPP and HHI establishes a rough link between increases in concentration and prices.

This paper proceeds as follows. Section 2 reviews the literature on the UPP. Section 3 lays down the Bertrand supply model, establishes the theoretical relationship between the UPP and merger simulation, and explains the relationship between the pass-through assumptions and the strategic relationship between the merging goods. Section 4 specifies the random coefficient structural demand model, introduces the dataset used, and presents the demand estimation results. Section 5 examines merger simulation results and compares it against prediction by the UPP, in both sign and magnitude. In particular, it investigates the case when the UPP uses imperfect demand-side inputs. It also tests the assumption behind Farrell and Shapiro's proposition and shows the consequence when it fails. Section 6 explores the correlation between the UPP and the HHI. Section 7 concludes.

1.2 The UPP Literature

Researchers had used first order conditions to assess merger effects before the UPP was coined. Hausman, Leonard, and Zona (1994) estimate a multi-level demand system for beer and derive a formula for the post-merger price change in terms of marginal costs and the pre-merger price margins. Werden (1996) subsequently uses first order conditions to recover the cost synergy necessary to overcome the incentive to raise prices, given the diversion ratio and pre-merger price margins, without assuming any functional form of demand. His objective is very close to the spirit of the UPP: the computation of upward price incentives can be turned into an exercise of finding the cutoff in cost reduction that just balances it. Shapiro (1996) identifies the diversion ratio and price margins as the key variables that determine the price effects from a merger. These earlier studies lay the foundation for the UPP.

Ever since the release of the first draft of Farrell and Shapiro (2010) in 2008, the UPP has generated wide attention among antitrust practitioners and academic economists alike. Most notably, Jaffe and Weyl (2010) generalize the UPP formula to allow multi-product firms and non-Bertrand firm conduct, such as Cournot competition (Nash-in-quantities) and consistent conjectures, in response to a few early critiques of the test.

However, the increased data requirement by these generalized formulas—sometimes the complete own- and cross-price elasticity matrix between *all* goods in the market—means that these generalized UPP's are almost impossible to compute without a structural demand estimation. While these generalized formulas fill a gap in the theoretical research on the UPP, they deviate from Farrell and Shapiro's intention of bypassing a demand model and limit the ease of computation relative to a full merger simulation.

Another group of researchers, such as Shapiro (2010) and Hausman, Moresi, and Rainey (2010), popularize the UPP test by deriving explicit formulas for an exemplary 2-firm, 2-product case, assuming simple linear and constant elasticity demands. Both study further simplify the UPP by assuming no cost synergies and call the resultant statistic the Gross Upward Pricing Pressure Index (GUPPI). A few researchers propose modifications to the original UPP such that it is closer to a full-blown merger simulation in various ways. For example, the original Farrell and Shapiro (2010) include a more complicated UPP formula that accounts for bilateral (instead of unilateral) sales diversion between the two merging goods. Schmalensee (2009) suggests including the cost synergies of both merging products in the UPP, instead of only the product whose price increase is under consideration. Simons and Coate (2010) advocate including a pass-through term to the UPP that is very similar in spirit to Jaffe and Weyl (2010)'s proposition.

Despite the wide interest on the UPP, empirical examination of the test has been extremely scarce. An early attempt by Walters (2007) highlights the difficulty in estimating diversion ratios (between supermarkets in the U.K.) without much consumption data, prior knowledge, or a structural demand model. He finds wide variance between estimated and surveyed diversion ratios. Varma (2009) is the first study that analyzes potential policy implications of the UPP. He investigates whether the UPP will lead to more or less merger enforcement than the traditional test based on market concentrations and HHI. By simulating hypothetical markets, he concludes that the UPP test is likely to flag more markets as anticompetitive than the HHI test, under typical market definitions and adopting a 10% "default" cost reduction as Farrell and Shapiro suggest. Mergers that are most likely flagged under the UPP but not the HHI test are those where the merging products do not have large pre-merger market shares, but are nonetheless considered first and second choices by a group of consumers. It will be very instructive to repeat this study in the future but with actual UPP and HHI values used by the antitrust agencies, together with an assessment on welfare change on realized mergers.

1.3 Theoretical Comparison between UPP and Merger Simulation

In market $t = 1, ..., T$, there are $a = 1, ..., A_t$ firms and a total of $j = 1, ..., J_t$ differentiated products. Denote the set of products produced by multi-product firm a by $\mathcal{J}_{a,t} \subseteq J_t$. Each market t has market size M_t, which is a measure of potential full consumption and is usually proportional to the relevant population size. When the equilibrium is modeled as Nash-in-prices, the profit function of firm a is given by

$$\Pi_a = M_t \sum_{j \in \mathcal{J}_a} (p_j - c_j) s_j(p) - C_j,$$

where $s_j(p)$ is the endogenous market share of good j, and C_j is the fixed cost of producing good j. The marginal cost c_j is assumed to be constant for simplicity. The first order necessary condition that determines the equilibrium strategy p_j^* is

$$s_j(p) + \sum_{k \in \mathcal{J}_a} (p_k - c_k) \frac{\partial s_k(p)}{\partial p_j} = 0.$$

This set of J equations in J unknowns is often expressed in matrix notation as

$$s(p) + \Omega(p - c) = 0,$$

where Ω is the $J \times J$ matrix of partial derivatives, with each element

$$\Omega_{kj} = \begin{cases} \frac{\partial s_k(p)}{\partial p_j}, & \text{if } \exists a \text{ s.t. } k, j \in \mathcal{J}_a \\ 0, & \text{otherwise.} \end{cases}$$

This set of equations that completely defines the firms' behavior and market equilibrium can be inverted to back out structural variables, such as the (assumed constant) marginal costs, or manipulated to compute new equilibria under various counterfactuals. A merger simulation that investigates the *unilateral* pricing impact of a merger, where the new post-merger equilibrium is computed by changing the ownership structure in Ω, is one of the many possible counterfactuals. Another counterfactual, likely combined with a merger simulation, is an investigation of pass-through, where the marginal costs are changed to see how much equilibrium prices respond. These counterfactuals are, essentially, comparative statics exercises.

Consider a merger of good j with good k, each produced by single-product firms for simplicity. (This analysis can easily generalize to multi-product firms, as Jaffe and Weyl (2010) demonstrate.) I now establish the theoretical relationship between the UPP and a structural merger simulation: UPP_j is a "single-product merger simulation" that solves for the change in product j's post-merger price only, while ignoring the re-equilibration of all other endogenous variables, including all competitors' prices $p_{-j} = \{p_l : l \neq j\}$, own and competitors' quantities $D_j(p), \forall j$, and the partial derivatives $\frac{\partial D_j(p)}{\partial p_j}$ and $\frac{\partial D_{-j}(p)}{\partial p_j}$. It also assumes that all competitors' costs c_{-j} are held constant at pre-merger levels.

This follows directly from our equilibrium definitions. Assume that the pre-merger equilibrium p^* and post-merger equilibrium p^{**} are Nash-in-prices. Good j's pre- and post-merger equilibrium prices and market shares satisfy its pre- and post-merger first order conditions respectively:

$$\begin{aligned}
&(p_j^* - c_j) \left.\frac{\partial D_j(p)}{\partial p_j}\right|_{p^*} + D_j(p^*) = 0, \text{ and} \\
&(p_j^{**} - (1-e)c_j) \left.\frac{\partial D_j(p)}{\partial p_j}\right|_{p^{**}} + D_j(p^{**}) + (p_k^{**} - c_k) \left.\frac{\partial D_k(p)}{\partial p_j}\right|_{p^{**}} = 0,
\end{aligned}$$

where e is the percentage reduction to marginal cost c_j post-merger. Assume that p_j is the only endogenous variable that re-equilibrates. All other endogenous terms are held constant at pre-merger values:

1. $D_j(p^*) = D_j(p^{**})$
2. $p_k^* = p_k^{**}$
3. $\left.\frac{\partial D_j(p)}{\partial p_j}\right|_{p^*} = \left.\frac{\partial D_j(p)}{\partial p_j}\right|_{p^{**}}$
4. $\left.\frac{\partial D_k(p)}{\partial p_j}\right|_{p^*} = \left.\frac{\partial D_k(p)}{\partial p_j}\right|_{p^{**}}$

The difference between the pre- and post-merger first order conditions gives the UPP formula:

$$UPP_j = (p_k^* - c_k) \underbrace{\frac{\partial D_k(p)}{\partial p_j}\bigg|_{p^*} \cdot \left(\left| \frac{\partial D_j(p)}{\partial p_j}\bigg|_{p^*} \right| \right)^{-1}}_{\text{diversion ratio } D_{jk}} - e \cdot c_j$$

Thus if $UPP_j > 0$, then $p_j^{**} > p_j^*$.

Under general conditions, both the quantity $D_j(p)$ and the derivatives $\frac{\partial D_j(p)}{\partial p_j}$ and $\frac{\partial D_k(p)}{\partial p_j}$ depend on the endogenous prices; they thus re-equilibrate with price during a merger simulation. When a demand model is assumed, some or all of these terms may adopt an explicit functional form. Farrell and Shapiro thus "free" the UPP from any demand functional form assumption by imposing that all terms except p_j do not re-equilibrate.[2] Compared with an equivalent merger simulation, the UPP sacrifices prediction accuracy for computational convenience and a smaller data requirement. Several researchers have suggested various ways to use the UPP stated above. The original Farrell and Shapiro (2010) advocate using the its *sign* only, instead of its magnitude, for merger prediction, because the UPP gives the *incentive* of price change. They also claim that predicting the sign of price change is more robust than predicting its magnitude. Pakes (2010) suggests using the *magnitude* of the UPP as a crude approximation to the simulated price change. Finally, Jaffe and Weyl (2010) use the UPP, together with its first derivative, to approximate the simulated price change in the same manner as the first step of Newton's method. I will examine the predictive accuracy of all three interpretations of the UPP in the next section.

There are two main sources of approximation error in the (magnitude of the) UPP compared with its equivalent merger simulation, one in the first term of the UPP and one in its second term. Firstly, if one uses "true" diversion ratios from an estimated demand model in the UPP, the UPP only differs from a merger simulation

[2]As other ways to generalize the UPP test, many alternatives are possible on what variables to hold constant in good j's UPP test. For example, Schmalensee (2009) modifies Farrell and Shapiro's UPP by including synergies for both c_j and c_k, as well as a corresponding diversion ratio from good k back to good j. This modification brings the UPP theoretically closer to a partial merger simulation, yet some of the aforementioned problem of non-reequilibration persists. Otherwise, all else equal, this modification leads to the UPP inequality to hold less frequently: a market is less likely to be flagged anti-competitive when the default cost synergy is applied symmetrically to both merging goods. Thus, this modification can potentially change the UPP prediction from a positive to a negative.

in its non-reequilibration assumption. The magnitude of price re-equilibration among non-merging firms in Bertrand conduct is typically small; the merging partner's price re-equilibration is small when its market share is large relative to the firm in question. However, the sign of this approximation error is often difficult to determine, and depends largely on how the price of the merging partner's good will react to an increase in own good's price pre-merger. This, in turn, depends on the strategic relationship between the two goods. The UPP can produce both *false positive* and *false negative* outcomes, which my empirical results will demonstrate. Secondly, the UPP ignores a measurement of pass-through—the rate at which a reduction in one's cost translates to a reduction of one's price. Having no pass-through term in the second term of the UPP, it implicitly assumes a pass-through rate of one. As I demonstrate below, the Bertrand model often exhibits a very high pass-through rate close to one across a wide range of prices and costs. Thus, neither of these two sources of approximation error are severe. The primary source of error in an empirical implementation of the UPP is likely to be in the use of inaccurate diversion ratios.

As with other first order approximations, the UPP's deviation from a simulated result also depends on how "small" and local the change $(p_j^{**} - p_j^*)$ is. In a structural demand model, it also depends on the rate of change of its derivative around its pre-merger value. Since a full merger simulation (with no cost synergies) must lead to $p_j^{**} > p_j^*$, one can conclude that $D_j(p^{**}) < D_j(p^*)$ with a downward-sloping demand curve, yet the relative magnitudes of $\left.\frac{\partial D_j(p)}{\partial p_j}\right|_{p^*}$ and $\left.\frac{\partial D_j(p)}{\partial p_j}\right|_{p^{**}}$ cannot be measured without a specification of the demand functional form. (If, for example, the demand curve is globally convex, as in the case of a constant elasticity demand, then one can conclude that $\left.\frac{\partial D_j(p)}{\partial p_j}\right|_{p^{**}} > \left.\frac{\partial D_j(p)}{\partial p_j}\right|_{p^*}$, both bearing negative signs.) The relative magnitudes of $\left.\frac{\partial D_{-j}(p)}{\partial p_j}\right|_{p^*}$ and $\left.\frac{\partial D_{-j}(p)}{\partial p_j}\right|_{p^{**}}$ are even more difficult to determine. It suffices to say that the smaller the change in price $(p_j^{**} - p_j^*)$, the more accurate is the UPP's magnitude. Thus, the UPP performs best when the two merging firms do not have large market share pre-merger, or that the merging products are not the most similar in the market.

1.3.1 UPP and Pass-Through Assumptions

So far we have interpreted the UPP of a single merging firm in isolation from that of its merging partner. Interpreting UPP values of two merging firms separately can lead to inconclusive results, for example, when $UPP_j > 0$ and $UPP_k < 0$. Notice that this does *not* imply simulated results of $\Delta P_j^{sim} > 0$ and $\Delta P_k^{sim} < 0$, due to the respective non-reequilibration assumptions discussed above. In particular, the result $UPP_j > 0$ relies on the assumption $\Delta P_k^{sim} = 0$, while $UPP_k > 0$ relies on the assumption $\Delta P_j^{sim} = 0$; thus $UPP_j > 0$ and $UPP_k < 0$ are derived from *different* assumptions. Although UPP_j and UPP_k are always based on different assumptions no matter their sign, a stronger result can be achieved when both are positive. Farrell and Shapiro (2010)'s Proposition 1 establishes the condition in a *duopoly-to-monopoly* merger where the post-merger equilibrium prices of *both* merging firms j and k will rise: when *both* UPP_j and UPP_k are positive. This conclusion requires non-negative post-merger pass-through assumptions of both own's cost and merging partner's cost among the two merging firms.[3] Specifically, for duopolists j and k, the proposition requires positive pass-through $\frac{\partial p_j^{**}}{\partial c_j} > 0$ and $\frac{\partial p_k^{**}}{\partial c_k} > 0$ for own costs, and non-negative pass-through $\frac{\partial p_j^{**}}{\partial c_k} \geq 0$ and $\frac{\partial p_k^{**}}{\partial c_j} \geq 0$ for merging partner's costs. Because differentiated product firms in Bertrand competition always choose price according to the inverse elasticity pricing rule, and the own-price elasticity at the profit-maximizing equilibrium is always elastic, positive price margins guarantee that the conditions of positive pass-through of own costs ($\frac{\partial p_j^{**}}{\partial c_j} > 0$ and $\frac{\partial p_k^{**}}{\partial c_k} > 0$) always hold.[4] However, assumptions on the pass-through of merging partner's costs are much less obvious.

As Weyl and Fabinger (2009) point out, one can in fact relate the above pass-through assumptions to the literature on strategic complements and substitutes first introduced by Bulow, Geanakoplos, and Klemperer (1985): they require the products of

[3] My terminology in this paper on various kinds of pass-through (merger, pre-merger, post-merger) follows that of Jaffe and Weyl (2010).

[4] Technically, this is true when all other prices—including prices of own firm's other products (in the case of multi-product firms) and prices of other firms' products—are held constant. When all prices are allowed to re-equilibrate, it is theoretically possible in Bertrand oligopoly that a good's own (post-merger) equilibrium price is decreasing in its own cost, when the re-equilibration of other prices causes its demand to be more elastic. I thank Thomas Holmes for this clarification. Empirically, this hardly affects the conclusions of my tests on pass-through assumptions.

the two merging duopolists to *not* be strategic substitutes post-merger.[5] To the extent that post-merger pass-through may be very similar to pre-merger pass-through (Jaffe and Weyl (2010)), Bulow, Geanakoplos, and Klemperer (1985) also establish that the strategic relationship between two firms depends on both the marginal cost and the demand functional form: strategic complementarity is not guaranteed in all differentiated product Bertrand settings. In the simple case when marginal cost is constant, the strategic relationship between two goods depends on how own-price elasticity responds to merging partner's change in price, and both strategic complements and substitutes are possible. If costs are increasing in quantity, the two goods will be strategic complements; if costs are decreasing in quantity, they will be strategic substitutes. The researcher thus needs to test the pass-through assumptions empirically if he wishes to make joint predictions from the UPP values of both firms. However, even if the UPP of one of the merging firms is to be interpreted alone, deviation from the above pass-through assumptions still potentially affects the UPP. Empirically, it may lead the UPP to produce false positive results (type I errors) because the re-equilibration that is ignored by the UPP will cause the simulated post-merger price to decrease when the merging partner's price increases.

Even more practically important is the case when a merger market consists of more than two firms pre-merger. For the same conclusion to Farrell and Shapiro (2010)'s Proposition 1 to hold, the own- and cross-product pass-through assumptions have to hold between *all* product-pairs in the market, including both merging and non-merging ones. In other words, satisfaction of the strategic complement assumption between the two merging products alone is not enough; all products in the market have to be strategic complements. In section 1.5.3, I test whether my sampled markets (all consisting of more than two firms pre-merger) satisfy these assumptions and the consequence it has on the joint interpretation of the UPP.

[5] Given $\frac{\partial p_j^{**}}{\partial c_j} > 0$, $\frac{\partial p_k^{**}}{\partial c_k} > 0$, $\frac{\partial p_j^{**}}{\partial c_k} \geq 0$, and $\frac{\partial p_k^{**}}{\partial c_j} \geq 0$, for merging products j and k. Then $\frac{\partial p_j^{**}}{\partial p_k^{**}} = \frac{\partial p_j^{**}}{\partial c_j} \left(\frac{\partial p_k^{**}}{\partial c_j} \right)^{-1} \geq 0$, and $\frac{\partial p_k^{**}}{\partial p_j^{**}} = \frac{\partial p_k^{**}}{\partial c_k} \left(\frac{\partial p_j^{**}}{\partial c_k} \right)^{-1} \geq 0$. The two products are strategic substitutes if $\frac{\partial p_j^{**}}{\partial p_k^{**}} < 0$ and $\frac{\partial p_k^{**}}{\partial p_j^{**}} < 0$. If the pass-through of merging partner's costs are strictly positive, then the merging products are strategic complements.

1.4 Demand

I use a discrete-type random coefficient nested logit model of demand, following Berry, Carnall, and Spiller (2006) (henceforth BCS) and Berry and Jia (2008). Each market is populated by agents indexed by i, each belonging to one of the customer types r. The conditional indirect utility of consumer i of type r choosing product j in market t is given by

$$u_{ijt} = x_{jt}\beta_r - \alpha_r p_{jt} + \xi_{jt} + \nu_{it}(\lambda) + \lambda\epsilon_{ijt},$$

where x_j is a K-dimensional vector of observable product characteristics, p_{jt} is the price of product j in market t, ξ_j is an unobserved (by the econometrician) product characteristic, ν_i is a random error that differentiates inside from outside goods, $\lambda \in [0, 1]$ is the nested logit parameter, and ϵ_{ijt} is an error term. The single nest used consists of all inside goods. In the discrete-choice differentiated product setting, each agent i chooses a single good j that gives her the highest utility. The utility from purchasing the outside good $j = 0$ is normalized to zero. I follow BCS in choosing the number of types $r = 2$, representing leisure and business travelers. I henceforth denote the population weights of these two consumers types as γ and $1-\gamma$. The overall market share of product j in market t is then

$$s_{jt}(x_t, p_t, \xi_t) = \gamma s_{jt1} + (1-\gamma)s_{jt2}, \text{ where}$$

$$s_{jtr}(x_t, p_t, \xi_t) = \frac{\exp(\frac{\delta_{jtr}}{\lambda})D_{rt}^{\lambda-1}}{1 + D_{rt}^{\lambda}}, \forall r = 1, 2,$$

is the type-specific market share, $\delta_{jtr} = x_{jt}\beta_r - \alpha_r p_{jt} + \xi_{jt}$ is the average utility of good j in market t to a customer of type r, and $D_{rt} = \sum_{k=1}^{J} \exp(\frac{\delta_{ktr}}{\lambda})$ is the inclusive value, summed over all inside goods within the nest. By using types to model consumer heterogeneity, preferences for product characteristics are correlated. It is important to model this correlation because, for example, a business traveler who is not very price-sensitive is likely to have a stronger preference for direct flights and amenities at the airport as well. The BCS random coefficient model simplifies to a simple nested logit model when $r = 1$.

1.4.1 Data

The merger between US Airways and America West was completed in 2005 after a short investigation by the DOJ. The antitrust agencies did not challenge this merger because the two airlines' route networks were complementary instead of overlapping: America West concentrated its operation in the western U.S., with hubs in Phoenix (PHX) and Las Vegas (LAX), while US Airways was active on the east coast, with hubs in Charlotte (CLT) and Philadelphia (PHL). US Airways entered Chapter 11 bankruptcy protection in 2002 and 2004, and its financial situation had prevented United Airlines to acquire it in 2001. America West's acquisition brought US Airways out of bankruptcy, and the resultant merged company retained the name of the acquired firm. After the merger of the two complementary route networks, the new US Airways became the fifth largest domestic carrier. The antitrust agencies believe that this merger is an example with credible cost synergies and consumer benefit (McDonald (2005)).

The primary dataset used is the DB1B by BTS RITA (Bureau of Transportation Statistics, Research and Innovative Technology Administration). This is a ten percent sample of all itineraries enplaned every quarter. Appendix A details my itinerary selection criteria, aggregation process to the airline-market level, and the definitions of the characteristics variables I constructed. I used the eight pre-merger quarters, from 2003Q3 to 2005Q2, for demand estimation. Secondly, I obtain the annual population of each MSA (metropolitan statistical area) from the U.S. Census. The market size of each city-pair is then given by the geometric mean of the populations of the origin and destination MSA's. Thirdly, I obtain the monthly consumer price index for urban consumers' transportation (not seasonally adjusted) from the BLS (Bureau of Labor Statistics). I then use this CPI to adjust all itinerary fares to my base year of 2003 (January) dollars. Table 1 presents summary statistics of key variables in my dataset.

1.4.2 Demand estimation results

Table 2 shows estimates of the BCS random coefficient demand model. (See appendix C for estimation procedure and model identification.) Estimates of a simple nested logit model is also included for comparison; the similarity between the nested logit

Table 1: Summary statistics of key variables

Complete dataset ($N = 124852$)	Mean	Std	Min	Max
Itinerary fare (2003$)	346.08	109.08	53.95	1406.16
Passengers (count)	154.27	308.59	10	4527
Direct flight (%)	27.85	42.81	0	100
Distance (miles; passenger-weighted)	3089.88	1718.58	134	11872
Passenger share at origin airport (%)	13.71	14.62	0.019	94.65
Passenger share at destination airport (%)	13.38	14.55	0.0060	94.65
Market share s_j	$5.8e^{-5}$	$1.30e^{-4}$	$6.46e^{-7}$	0.0062
Market size M	3060724	2053482	221160	$1.55e^{7}$

Note: An observation is an airline–origin airport–destination airport–quarter tuple. A route is defined as an origin airport–destination airport pair. A market is defined as an origin MSA–destination MSA pair. Thus, a market contains multiple routes when a MSA contains multiple airports. Eight pre-merger quarters are used, from 2003Q3 to 2005Q2, in the complete dataset.

model's coefficients and those for the dominant customer type (type I) in the BCS model validates the BCS estimates. All coefficients bear the expected signs, reasonable magnitudes, and are statistically significant. For example, to quantify the coefficient's magnitudes using results from the first customer type, an increase in the percentage of direct flights of an airline on a route by 10% will increase the average quality of the product by 0.17, which is equivalent to a decrease in airfare by $22.05. An increase of 500 miles during the flight corresponds to a decrease in price of $13.55. In the simplifying assumption that passenger share at the origin airport (a proxy for an airline's "airport presence" and level of customer service at check-in) is exogenous, an increase in the passenger share at origin airport of an airline on a route by 10% will increase the average quality of the product by 0.179, which is equivalent to a decrease in airfare by $23.22.[6]

The last two columns show the two customer types have distinct preferences. Since type 2 consumers have a much lower price sensitivity than type 1 consumers, one can consider type 1 consumers as leisure travelers and type 2 consumers are business travelers. This categorization is reasonable given the estimated value of γ, which gives

[6] Airline passenger shares at origin and destination airports are, of course, endogenous in reality. These passenger shares are all but statistics of endogenous market shares of all routes that involve these airports and belong to the airline in question. However, the assumption of exogeneity may not be too severe because my demand model treats each route-market as independent. Since an airport-presence variable depends on the market share of *all* routes that involve that airport, the effect of *one* route's market share on the variable may be minimal.

Table 2: Demand estimates

	Nested logit[a]	BCS nested logit[b]	
		Type I	Type II
Itinerary fare[c]	-0.00499	-0.00771	-0.00117
	$(7.452e^{-4})$	$(1.752e^{-4})$	$(6.88e^{-5})$
Direct flight	0.0145	0.0171	0.0954
	$(3.98e^{-4})$	$(4.255e^{-4})$	$(7.65e^{-5})$
Distance	$1.24e^{-4}$	$2.098e^{-4}$	-0.0170
	$(2.46e^{-5})$	$(1.5e^{-5})$	$(4.674e^{-4})$
Psngr. share at origin airport	0.0118	0.0179	0.0527
	$(9.697e^{-4})$	(0.00141)	$(5.97e^{-5})$
Psngr. share at destination airport	0.00519	0.00952	0.00877
	$(8.264e^{-4})$	(0.00108)	$(4.07e^{-5})$
constant	-8.573	-8.570	-8.573
	(0.169)	$(9.49e^{-5})$	$(5.27e^{-5})$
Nested logit parameter (λ)	0.534	0.604	
	(0.00964)	(0.00176)	
Percentage in population (γ)	–	0.961	0.039
	–	$(8.49e^{-5})$	$(8.49e^{-5})$
N	5207	5207	
Average own-price elasticity	-2.767	-3.822	-0.580
Average cross-price elasticity	0.180	0.210	0.032

Note: Standard errors are reported in parentheses.

[a] The dependent variable is $ln(s_{jt}) - ln(s_0)$. All variables are statistically significant at the 1% level.

[b] Standard errors for the BCS demand parameters are computed using bootstrap.

[c] A Hausman instrument of prices is used to control for the endogenous itinerary fare. This instrument is defined by the simple average of itinerary fares, in 2003 dollars per mile, over all *other* observations in the same market. The unit of dollar *per mile* is used when constructing the Hausman instrument to allow for fair comparison between itinerary fares of different route lengths. Appendix A shows the first stage regression results of the endogenous variable on the exogenous and instrumental variables.

the percentage of leisure travelers in the population as 96%. Other coefficients also show that business travelers have a much stronger preference for direct flights and the airline's presence in the origin airport, which favors the use of frequent flier plans and serves as a proxy for customer service at check-in. Not surprisingly, neither customer type value the airline's presence at the destination airport much, as both customer types have similarly low coefficients. Almost all coefficients for both consumer types bear the same sign as the simple nested logit estimates. A reasonable exception is the "distance" variable for the business travelers to bear a negative sign.

1.5 Results

1.5.1 UPP and Merger Simulation

In this section, I compare the UPP's predictions against their structurally simulated price changes. Using the price elasticities implied by the BCS random coefficient demand estimates, I first back out the pre-merger marginal cost of each good using the firms' pre-merger first order conditions. With these marginal costs, I conduct a series of merger simulations for each market to obtain simulated post-merger prices. In particular, I consider a range of cost synergies e between zero and 0.1 to obtain both positive and negative simulated price changes, for the same set of markets. In each set of simulation, the same cost synergy value e is applied to the marginal costs of both America West and US Airways. This setup enables me to test whether the UPP creates any false positive (type I error) or false negative (type II error) predictions. For each airline-market observation, I also use the (pre-merger) own- and cross-price elasticities implied by the estimated demand model, and changes in marginal costs to compute a corresponding UPP value. It is important to emphasize that each UPP value I compute in this manner corresponds to a structural merger simulation in the *closest* possible way: namely, that the diversion ratio used in the UPP is the "true" one implied by the estimated structural demand model, and cost synergy e used in the UPP is identical to that in the merger simulation. Thus, the only source of deviation of the UPP from the simulated price change comes from the UPP's theoretical simplification in non-reequilibration. The use of estimated price elasticities is, of course, not the only way to compute the diversion ration in the UPP; in fact, Farrell and Shapiro originally intend their merger screen to bypass a structural demand estimation so to be quick to use. Using diversion ratios other than the "true" one will most likely lead the UPP to deviate further from its corresponding merger simulation result than my findings below. The consequence of this deviation is impossible to summarize because the manner in which practitioners will estimate the diversion ratio non-structurally will be case-specific. However, I do consider logit diversion ratios below that do not require demand estimation, and test how well this "practical" model emulate the BCS

Table 3: Tabulation of sign predictions by the UPP

cost	*UPP:*	positive	negative	positive	negative
synergy e	*Simulated price change:*	positive	negative	negative	positive
0		256	0	0	0
0.02		117	122	6	11
0.04		78	145	26	6
0.06		47	181	24	4
0.08		32	201	22	1
0.1		17	219	19	1

No. of observations $N = 256$ *for all scenarios.*
The four categories of UPP predictions are, in order: true positive, true negative, false positive (type I error), and false negative (type II error).

outcomes.

First I assess the UPP's prediction on the *sign* of post-merger price change, for both America West and US Airways in my set of overlapping route markets. Table 3 tabulates the UPP's sign predictions against the sign of the corresponding simulated price changes. As cost synergy e increases, more simulated price changes switch from positive to negative as expected, as the incentive to decrease price due to lowered cost dominates the incentive to increase price due to lost competition. In this set of markets and simulations, the UPP predicts the correct sign of simulated price change at least 87% of the time. Among the observations with wrong sign predictions, there are usually more false positives (type I errors) than false negatives (type II errors) in my markets.

Figure 1 is a scatter plot that shows how the UPP and the structurally simulated price change transition from positive to negative values as the cost synergy e increases from 0 to 0.1. For each airline-market-level observation, all six pairs of UPP and simulated values, each pair generated with a different level of cost synergy e, are included. Every two consecutive points within each airline-market observation are connected with a straight line. Points in the upper-left quadrant represents false positive results (type I errors), while points in the lower-right quadrant are false negatives (type II errors). A striking observation is that the overall line of each airline-market observation, each consisting of six connected points, itself resembles a straight line very much. This is evidence that the ratio between the rate at which own-cost reduction is passed through to the UPP, and the rate at which the same cost reduction is passed through to the

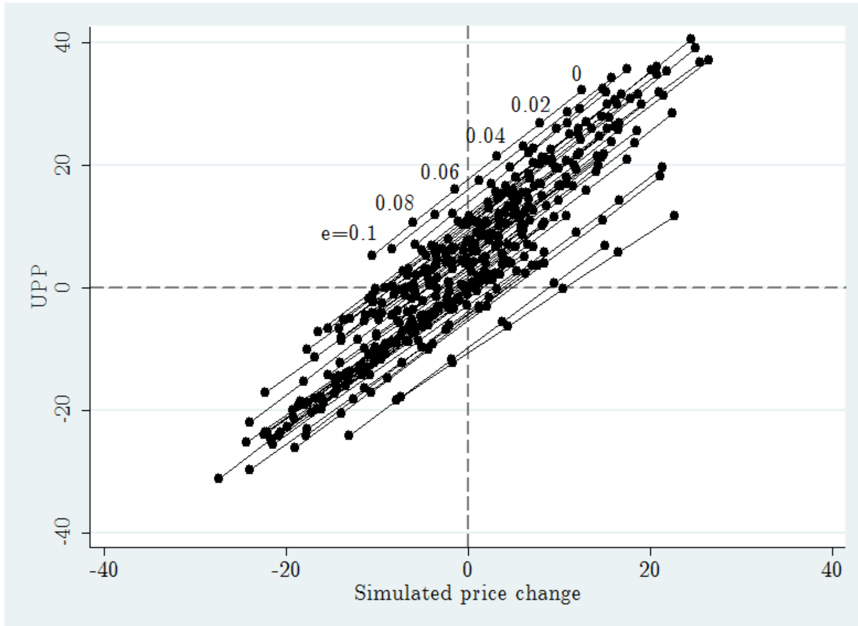

Figure 1: Scatter plot between the UPP and corresponding structurally simulated price changes, showing transition from positive to negative regions as the cost synergy e increases from 0 to 0.1. For each airline-market observation, all six pairs of values, each pair generated with a different level of cost synergy e, are included, and each pair of consecutive points is connected with a straight line. Only a subset of airline-market observations are plotted above to avoid over-crowding.

simulated price, is almost constant across the values of cost synergy e considered. More importantly, the slopes of these lines are all very close to one. In other words, the UPP is remarkably accurate in approximating the own-cost pass-through rate of a structural merger simulation, independent of the particular magnitude of the cost reduction used. The consequence of this observation in relation to the UPP's *sign* prediction is that for each airline-market observation j, there exists necessarily a range of values of cost synergy e where the UPP will give a wrong sign prediction, as long as the first term of the UPP_j, $D_{jk}(p_k - c_k)$, deviates from the corresponding simulated price increase (when $e = 0$). Graphically, if one is to pick a point in the first quadrant that does not lie on the 45-degree line, and to extend a 45-degree line from this point to the lower-left quadrant (to trace the UPP and simulated price predictions as e increases), some part of this line must lie in either the upper-left or lower-right quadrant. It can be seen from the graph that this range of cost synergy e that produces a wrong sign prediction can

Table 4: AVERAGE SIMULATED PRICE CHANGE AND UPP UNDER SIX VALUES OF COST SYNERGY e

cost synergy e	Mean simulated price change	Mean UPP	Corr.	Mean abs. difference
0	\$7.79	10.24	0.8985	3.5409
0.02	\$3.10	5.13	0.9053	3.4157
0.04	−\$1.58	0.022	0.9127	3.3430
0.06	−\$6.26	−5.09	0.9202	3.3145
0.08	−\$10.92	−10.20	0.9275	3.3287
0.1	−\$15.59	−15.31	0.9344	3.3686

No. of observations $N = 256$ for all scenarios.

be large, for some observations. (The distance between any two consecutive points on the graph represent a 2% change in cost.) The size of this problematic range of e depends on how far the first term of the UPP_j, $D_{jk}(p_k - c_k)$, deviates from the simulated price increase (when $e = 0$). This in turn depends on two factors: firstly, the quality of the diversion ratio used to compute the UPP in comparison to the actual substitution patterns; secondly, the severity of the UPP's non-reequilibration assumption in each market. It suffices to emphasize that the UPP will produce a wrong sign prediction at *some* values of cost synergy e. Figure 1 also suggests that an airline-market observation is unlikely to give both false positive and false negative predictions, but only either one, as e changes. This is because a 45-degree straight line cannot pass through both the upper-left and lower-right quadrants.

I now turn to the *magnitudes* of the UPP as an approximation to the magnitudes of simulated price increases. The summary statistics in table 4 show that their averages are remarkably close. As the cost synergy e increases from zero to 0.1, both the mean simulated price change and the mean UPP decrease expectedly. In my set of markets, the magnitudes of the UPP show a slight tendency to over-predict the price increase (or under-predict the price decrease). Despite this slight difference in overall magnitudes, the correlation between these two variables are consistently high. This high correlation implies that the UPP is informative in flagging the markets with largest potential price increase among a cross-section of markets, although each UPP value may not be directly translated into a price change.

A second interesting observation is that the correlation between the UPP and simu-

Table 5: REGRESSION OF THE UPP ON SIMULATED PRICE INCREASE

cost synergy e	coeff. b	const.	R^2
0	1.208	0.827	0.8074
	(0.037)	(0.415)	–
0.02	1.221	1.343	0.8196
	(0.036)	(0.310)	–
0.04	1.229	1.963	0.8330
	(0.035)	(0.287)	–
0.06	1.232	2.620	0.8467
	(0.033)	(0.344)	–
0.08	1.231	3.252	0.8603
	(0.031)	(0.434)	–
0.1	1.227	3.813	0.8732
	(0.029)	(0.528)	–

No. of observations $N = 256$ *for all scenarios.*
Standard errors are presented in parenthesis.

lated price change increases as the cost synergy e increases. This in fact relates to the UPP's remarkable accuracy in approximating the own-cost pass-through rate of a structural merger simulation, as mentioned above. Specifically, the UPP has two sources of approximation errors relative to a simulated price change, that correspond to the two terms in the UPP: firstly, that the UPP approximates the increase in price due to loss in competition by ignoring the re-equilibration of all other endogenous variables, except its own price; secondly, that the UPP approximates the decrease in price due to cost reduction by assuming an own-cost pass-through rate of one. As the analysis of figure 1 above explains, the second source of approximation error is very small and is independent of the size of cost synergy e. Then, as e increases, the fall in price due to cost reduction increases, and the UPP's approximation error from the first source (non-reequilibration) decreases in proportion to the overall price change, thus giving the increasing correlation between the UPP and simulated price change with e. To further investigate their relationship, I regress the UPP on its corresponding simulated price change, for each cost synergy value e. Table 5 shows these regression results. Both the coefficient and the constant term are statistically significant for all values of e. The increase in R^2 with cost synergy e echoes the result above where the correlation between these two variables increases with e.

Lastly, to ascertain that the (average) slope of the lines in figure 1 is close to one, I approximate the post-merger pass-through rate using consecutive values of simulated prices as cost reduces successively with synergy e. Since these pass-through rates are estimated using post-merger equilibria (each with a different cost synergy value e), the pass-through rates I obtain are post-merger pass-through. Although, as Jaffe and Weyl (2010) clarify, that the post-merger pass-through rate is theoretically different from the merger pass-through rate, the two can be similar empirically. The approximations in table 6 are indeed very close to one, and almost stay constant throughout the range of cost synergy e considered. This is coherent with the observation that lines in figure 1 have slopes close to one.

This approximation of the merger pass-through can then be used to translate pricing pressure to an actual change in price. Jaffe and Weyl (2010)'s Theorem 1 establishes that the magnitude of simulated price change can be approximated by the UPP multiplied by the merger pass-through rate. The approximated price change ΔP generated this way thus has two sources of approximation errors: firstly, that the first term of the UPP (with "true" diversion ratios) approximates the increase in price due to loss in competition by ignoring the re-equilibration of all other endogenous variables, except its own price; secondly, that the pass-through rate at the post-merger equilibrium does not equal that at the pre-merger equilibrium, or the overall merger pass-through. All pass-through rates are functions of the curvature of demand; in most commonly used functional forms of smooth demand functions, the change in curvature increases with distance traveled. Thus, it is possible that the second source of approximation error increases when the simulated price is far away from the pre-merger price, either because of no cost synergy (thus simulated price is above pre-merger price) or a large cost synergy (thus simulated price is far below the pre-merger price). The very small changes in pass-through values in table 6 indicates that this source of approximation error is minimal. Indeed, a comparison between the last column of table 6 and the last column of table 4 shows that the approximated price change computed using Jaffe and Weyl (2010)'s Theorem 1 (that uses a pass-through rate) has a smaller mean absolute difference from the structurally simulated price change, than one without. This

Table 6: APPROXIMATED POST-MERGER PASS-THROUGH

cost synergy e	Mean approx. post-merger pass-through[a]	Mean abs. difference[b]
0	–	–
0.02	0.9167	2.9409
0.04	0.9151	2.9207
0.06	0.9136	2.9174
0.08	0.9122	2.9253
0.1	0.9108	2.9376

No. of observations $N = 254$ for all scenarios.

[a] For example, post-merger pass-through at $e = 0.02$ is approximated by $\frac{P_{j,00}^{sim} - P_{j,02}^{sim}}{0.02 \times c_j}$, where $P_{j,00}^{sim}$ is the simulated post-merger price for good j when $e = 0$, and $P_{j,02}^{sim}$ is the simulated post-merger price for good j when $e = 0.02$, for each airline-market observation.

[b] This is the mean absolute difference between the approximated price change computed according to Jaffe and Weyl (2010)'s Theorem 1, and the simulated price change. The former is, in turn, computed by multiplying the UPP by the approximated post-merger pass-through in the second column.

discussion on pass-through again points to the importance of having good diversion ratio estimates, as all other sources of approximation error in the UPP are relatively insignificant.

1.5.2 UPP without demand estimation

The UPP values I have considered so far are by definition as close to the respective simulated price changes as possible, because the diversion ratios used in the UPP are "true" values computed from the own- and cross-price elasticities of the estimated structural demand system. This is, of course, not the only way to estimate diversion ratios, and certainly not one that Farrell and Shapiro (2010) advocate, due to its time and data requirement. Here I derive diversion ratios using the much simpler logit demand model.[7] It does not require any estimation because substitution patterns are simply given by observed market shares, due to its IIA (Independence of Irrelevant Alternatives) property. Thus, logit diversion ratios do not require any more data than

[7]Specifically, this is a logit demand model defined *without* the outside good $j = 0$. This is appropriate for my definition of the market size M, which is the geometric mean of the populations of the origin and destination metropolitan areas. This results in a very large share of the outside good (i.e. percentage of people not flying); in fact it is always larger than 0.99. Using observed market shares $s_j, \forall j \geq 1$ only will avoid the strong bias in substitution towards the outside good, giving a diversion ratio between inside goods close to zero. This approach is also reasonable in the general case because sometimes the total market size M is difficult to measure, and is only required for a structural demand estimation.

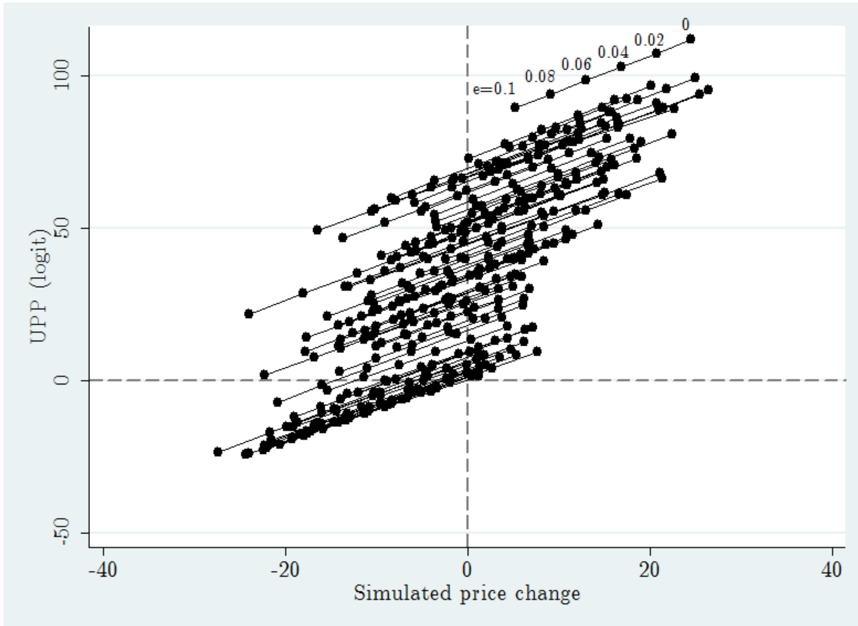

Figure 2: Scatter plot between the UPP (computed using diversion ratios from logit elasticities) and structurally simulated price changes (using BCS demand), showing transition from positive to negative regions as the cost synergy e increases from 0 to 0.1. For each airline-market observation, all six pairs of values, each pair generated with a different level of cost synergy e, are included, and each pair of consecutive points is connected with a straight line. Only a subset of airline-market observations are plotted above to avoid over-crowding; the same subset is used as in figure 1.

what the UPP already demands.[8]

Figure 2 is a scatter plot similar to figure 1, except the UPP used is computed from logit diversion ratios. For each airline-market level observation, the graph traces the pair of values as cost synergy e increases from zero to 0.1 successively. Figure 2's striking difference from figure 1 demonstrates the grave consequence when the diversion ratios used deviate significantly from the true values. Graphically, if the first point (when cost synergy $e = 0$) of each airline-market observation is located far from the 45-degree line, a much longer portion of the extension of this line will lie in either the top-left or bottom-right quadrant. Consequently, UPP values computed using these imperfect diversion ratios will give wrong sign predictions for a much larger range of cost synergy

[8]Farrell and Shapiro have suggested using logit diversion ratios in an interview. See "Roundtable Interview with Joseph Farrell and Carl Shapiro," `http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/Feb10_FarrShapRT2_25f.authcheckdam.pdf`, p.4.

e. This result is not idiosyncratic to the specific airline markets or merger analyzed; it is general to all applications when the UPP tries to approximate a Bertrand merger simulation result with imperfect diversion ratios. For this set of observations, the range of problematic e includes Farrell and Shapiro's "default" 0.1 for many observations in figure 2, as many lines' last points (when $e = 0.1$) are located in the top-left quadrant, giving false positive predictions. This serves as a caution against Farrell and Shapiro's belief that the prediction of the *sign* of price change is inherently more robust than *magnitude* predictions (Farrell and Shapiro (2010), p.19). Either type of predictions require good diversion ratios in order to emulate a structural merger simulation well.[9]

As Pakes (2010) points out, Farrell and Shapiro's suggestion in obtaining diversion ratios from company documents and consumer surveys are not immune to problems. The most difficult aspect is to identify the preferences of the *marginal* consumer as opposed to the average consumer: the diversion ratio captures the switching pattern of the *next* consumer lost upon an infinitesimal increase in price. Another potential method to measure the diversion ratio $\frac{\partial s_j}{\partial s_k}$ is a regression of one good's quantity on another, assuming that the quantities of these goods are observed repeatedly, either across time or across markets. A naïve implementation of this regression is unlikely to capture the substitution pattern either because of endogeneity of these two quantities and the likely high correlation between them, even with adequate use of fixed effects and control variables, such as costs. It is an instructive exercise in the future to evaluate the performance of these alternative methods to estimate the diversion ratio against a structurally estimated one in analyzed mergers.

[9] The correlation between the UPP (computed using logit diversion ratios) and simulated price changes (from structural model with BCS demand) is still high, and is similar to values in table 4. Thus the *relative* magnitudes of the UPP are still informative. (This strong correlation is largely due to the fact that both logit and BCS demands belong to the same family of discrete choice models. The same correlation may not be achieved if one uses other means to estimate diversion ratios, such as business documents, consumer surveys, or regression on quantities.) However, the mean values (or levels) of the UPP and simulated price changes differ greatly, unlike the case in table 4. The UPP with logit diversion ratios over-predicts the post-merger price increase because logit demand exaggerates substitution between the merging products. This can also be seen in figure 2, in that most of the first points of the lines (when cost synergy $e = 0$) lie above the 45-degree line in the first quadrant.

1.5.3 Testing Pass-Through Assumptions

As mentioned in section 1.3.1, pass-through assumptions on own and merging partner's costs imposed by Farrell and Shapiro's Proposition 1 do not always hold theoretically, and their violation may affect the accuracy of the UPP, interpreted either individually or jointly between merging firms. Here I show that the strategic complement assumption is not empirically innocuous either. To test the post-merger pass-through assumptions on own and merging partner's costs, I impose a 10% increase in marginal cost on *one* of the merging partners at a time and simulate the resultant post-merger equilibrium. This set of equilibrium prices is then compared against the standard set of post-merger equilibrium prices where there is no change in costs. If equilibrium price p_j^{**} weakly increases with marginal cost c_k and vice versa for p_k^{**} and c_j, the two merging goods satisfy the pass-through assumptions. Furthermore, if p_j^{**} strictly increases with c_k and vice versa, the two merging products are strategic complements post-merger. Among my sampled markets, none satisfy the requirement of non-negative post-merger pass-through of merging partner's costs. Because the post-merger pass-through rate is not observable from pre-merger data, I consider the alternative of using the pre-merger pass-through rate, computed in an analogous manner but with pre-merger equilibrium prices p_j^* and p_k^*. These two pass-through rates always agree in sign and are very similar in magnitude. Thus my markets satisfy neither pre-merger pass-through nor post-merger pass-through requirements. As a consequence of the BCS model's violation of these assumptions, some UPP values that are calculated from "true" elasticities from the BCS model give false positive predictions, as table 3 shows. It is worth noting that this resultant prevalence of false positives in the UPP is in contradiction with Farrell and Shapiro's remark that the UPP otherwise normally has a tendency to produce false negatives (Farrell and Shapiro (2010), p.13). When all of its post-merger pass-through assumptions are satisfied, the UPP has a tendency to *under*-predict increase in price because it ignores a feedback when sales of the merging partner's good is diverted back to the own good.

I complete the comparison between pass-through rates by computing the merger pass-through rate for my markets and comparing it against the pre-merger and post-merger rates. The merger pass-through rate is computed by comparing the post-merger equilibrium prices p_j^{**} and p_k^{**}, under a 10% increase in marginal cost on *one* of the merging partners at a time, against the pre-merger equilibrium prices p_j^* and p_k^*, under original marginal costs. Because the increase in price generated in this manner is necessarily larger than the increases in the previous two computations (due to the inclusion of the merger effect), merger pass-through conditions are easier to satisfy than pre-merger or post-merger conditions. Among my sampled markets, 84 out of 128 satisfy the requirement of non-negative merger pass-through of own and merging partner's costs, which makes for a more meaningful comparison within my sample. Figure 3 plots the scatter between the simulated price change and the "true" UPP when cost synergy $e = 0.04$ (chosen such that there are both positive and negative price changes). There is a definitive pattern where points from markets that violate the merger pass-through conditions are concentrated on the upper-left half of the cluster, thus more likely to be found in the graph's upper-left quadrant (even as cost synergy e changes), resulting in a false positive prediction from the UPP. In fact, the majority of the points in the upper-left quadrant are from markets that do not satisfy the merger pass-through conditions.[10]

Now I turn to the joint interpretation of the UPP's for both merging products in relation to pass-through. Section 1.3.1 generalizes Farrell and Shapiro's Proposition 1 to the practical case when the pre-merger market is larger than a duopoly. For its conclusion to continue to hold, then, the cross-product pass-through between *all* products in the market (both merging and non-merging ones) have to be non-negative. All of my

[10]Here I also replicate table 3, reporting the tabulation between the two groups of markets (those that satisfy or violate pass-through between merging partners) separately. The same pattern as figure 3 is observed for all values of cost synergy e: markets violating merger pass-through conditions show a larger percentage of false positive results among observations with wrong sign predictions.

		Markets satisfying pass-through				Markets violating pass-through			
cost	UPP :	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.
synergy e	ΔP^{sim} :	pos.	neg.	neg.	pos.	pos.	neg.	neg.	pos.
0		84	0	0	0	172	0	0	0
0.02		70	6	1	7	47	116	5	4
0.04		49	19	10	6	29	126	16	0
0.06		30	42	8	4	17	139	16	0
0.08		19	50	14	1	13	151	8	0
0.1		11	64	9	0	6	156	10	0

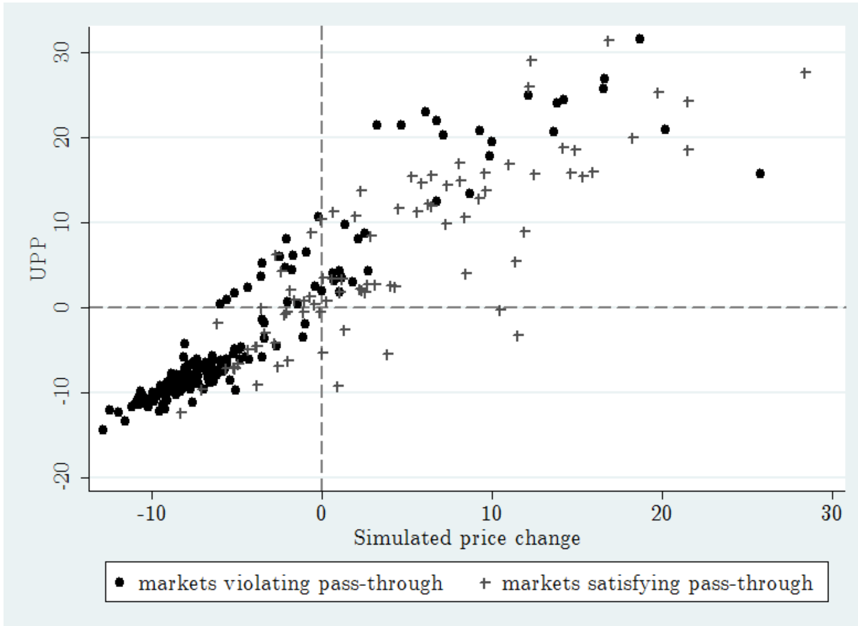

Figure 3: Scatter plot between the UPP and corresponding structurally simulated price changes, when cost synergy $e = 0.04$. The conditions considered here are merger pass-through between the two merging products only.

128 markets are larger than a duopoly pre-merger. From the same set of simulation test results above, none of them satisfy this more demanding set of requirements (although 84 of them have non-negative merger pass-through between the two merging products). Table 7 tabulates the number of markets that have positive UPP values for both merging firms, and among them, those that have positive simulated price changes for both merging firms. It shows that strategic complementarity (defined by positive merger pass-through) between *only* the two merging firms is not sufficient for the proposition's conclusion to hold: joint positive values do not always lead to joint positive simulated price changes. This is because prices of non-merging firms re-equilibrate during a merger simulation; in the case of strategic substitutes, they decrease in response to the merging firms' increase in prices. This, in turn, causes the merging firms' prices to increase *less* after full re-equilibration. The non-zero tabulations in the last column of table 7 also show that the aforementioned pass-through conditions are sufficient but not necessary.

Table 7: TEST OF FARRELL AND SHAPIRO'S PROPOSITION 1 ON NON-DUOPOLY PRE-MERGER MARKETS

cost synergy e	Markets satisfying merger pass-through between AW and US only[a]	
	$(UPP_{HP} > 0, UPP_{US} > 0)$	$(UPP_{HP} > 0, UPP_{US} > 0)$ and $(\Delta P^{sim}_{HP} > 0, \Delta P^{sim}_{US} > 0)$
0	42	42
0.02	29	28
0.04	18	10
0.06	1	0
0.08	0	0
0.1	0	0

[a] None of these markets satisfy the full set of pass-through assumptions between *all* (merging and non-merging) products in the market.

1.5.4 UPP's Deviation from Simulated Price Increase

One way to improve on a price prediction is to estimate its deviation from the "gold standard" as well, since this deviation may not be random. As emphasized above, the difference between the magnitudes of the UPP and the simulated price change (when cost synergy e is zero), other than the merger pass-through rate established by Jaffe and Weyl (2010), is attributed to non-reequilibration. In this section, I explore whether this approximation error due to non-reequilibration is related to product characteristics. I conduct an OLS regression of the magnitude difference between computed UPP and simulated price change on the four observed product attributes and the recovered unobserved product characteristic (ξ_j). By comparing the significant levels of the regressors, I can assess whether some product characteristics cause more re-equilibration than others. Table 8 shows two sets of regression results, where the first includes only observable characteristics, and the second includes the recovered unobservable ξ_j as well. Thus the first regression may be more suggestive if the practitioner wishes to bypass a demand estimation. The R^2 values from the two regressions are similar, both below 0.2, indicating that the overall explanatory power of these characteristics on the UPP's deviation is not high. Both regressions conclude that itinerary fare and airline presence in the origin airport are the most statistically significant. Both of their coefficients are positive, meaning that both itinerary fare and airline presence at the origin airport are correlated with the UPP's *under*-prediction of price increase.

Table 8: UPP deviation regression estimates

Itinerary fare	0.0260	0.0238
	(0.0069)	(0.0069)
Direct flight	-0.0064	0.0020
	(0.0155)	(0.0156)
Distance	0.0010	0.0016
	(0.0006)	(0.00006)
Psngr. share at origin airport	0.1156	0.1177
	(0.0412)	(0.0407)
Psngr. share at destination airport	0.0809	0.0840
	(0.0460)	(0.0454)
Unobservable ξ_j	–	0.09892
	–	(0.3709)
constant	-18.1558	-20.3851
	(3.6313)	(3.6838)
R^2	0.1478	0.1715
N	256	256

Note: Standard errors are reported in parentheses.
The dependent variable in both regressions is $(\Delta P_j^{sim} - UPP_j)$, where ΔP_j^{sim} is the simulated price increase of product j.

In other words, the re-equilibration (from both the merging partner and non-merging firms) that is ignored by the UPP will lead the simulated price to increase further (or decrease less) than the UPP indicates, and this re-equilibration is positively correlated with the itinerary fare and airline presence at the origin airport. This is reasonable because these two are the most important attributes that determine product similarity. Lastly, the unobservable ξ_j is positive and significant in the second regression. This is confirmation that unobserved quality has important impact on consumer choice.

1.6 UPP and HHI

As a preliminary merger screen that is easily computed with small data requirement, the UPP is intended as an alternative to the traditional approach based on market shares and changes in the HHI (Herfindahl-Hirschman Index), which is equally easy to compute and small in data input. In this traditional approach, the post-merger HHI is computed using pre-merger market shares, assuming that market shares do not change post-merger. Thresholds on this increase are then chosen relative to the pre-merger HHI; these thresholds had also been revised in the new 2010 *Guidelines*. Here, I

Table 9: Tabulation of markets according to HHI thresholds

Using 1992 thresholds:	Total no. of markets	Markets where $UPP_{AW} > 0$ or $UPP_{US} > 0$	Markets where $\Delta P_{AW}^{sim} > 0$ or $\Delta P_{US}^{sim} > 0$
$HHI_{pre} < 1000$	0	0	0
$1000 \leq HHI_{pre} \leq 1800$ and $\Delta HHI > 100$	5	0	0
$HHI_{pre} > 1800$ and $\Delta HHI > 50$	105	36	18
Using 2010 thresholds:	**Total no. of markets**	**Markets where $UPP_{AW} > 0$ or $UPP_{US} > 0$**	**Markets where $\Delta P_{AW}^{sim} > 0$ or $\Delta P_{US}^{sim} > 0$**
$HHI_{pre} < 1500$	1	0	0
$1500 \leq HHI_{pre} \leq 2500$ and $\Delta HHI > 100$	37	0	0
$HHI_{pre} > 2500$ and $\Delta HHI > 100$	66	36	18

No. of market-level observations is 128 *for all scenarios.*
HHI_{pre} denotes pre-merger HHI. ΔHHI denotes increase in HHI. UPP_j values are compuated with the "default" 10% reduction in cost to both merging firms. $j = AW$ denotes America West; $j = US$ denotes US Airways. ΔP_j^{sim} denotes the structurally simulated price increase of firm j, computed with the "default" 10% reduction in cost to both merging firms.

first explore whether the UPP and merger simulation will flag the same set of markets as the HHI. Specifically, I tabulate the number of markets flagged by both the 1992 and 2010 HHI thresholds, and the number of markets flagged by the UPP and merger simulation among them. I use Farrell and Shapiro's "default" 10% cost reduction when computing the UPP and simulated price increases; they flag a market as anticompetitive when either (or both) of the merging firm's value is positive. Table 9 shows that the UPP and merger simulation are flagging only a small portion of the markets that are flagged by the HHI. In particular, they fail to flag any of the markets that are deemed "moderately concentrated" in either year's HHI standard. This suggests that one may need to adjust the "default" cost reduction for each merger; a smaller value will lead to more agreement between the three screens in this case. Note that this result is opposite to Varma (2009), who suggests that UPP is a tighter screen than the HHI from his simulations. This is possibly due to the "narrow" definition of airline-route markets. Most consumer products will have many more competitors than the number of airlines in a route, which leads to lower HHI values.

I also compare the HHI's correlations with the UPP and simulated price changes respectively. Because the HHI is inherently unable to take into account the change in

Table 10: CORRELATION BETWEEN INCREASE IN HHI, SUM OF UPP'S, AND SUM OF SIMULATED PRICE INCREASES

corr.	$(\Delta P^{sim}_{AW} + \Delta P^{sim}_{US})$	$(UPP_{AW} + UPP_{US})$	ΔHHI
$(\Delta P^{sim}_{AW} + \Delta P^{sim}_{US})$	1	–	–
$(UPP_{AW} + UPP_{US})$	0.8985	1	–
ΔHHI	0.6053	0.5135	1

No. of market-level observations is 128 *for all scenarios.*
ΔP^{sim}_j denotes the structurally simulated price increase of firm j. $j = AW$ denotes America West; $j = US$ denotes US Airways.

post-merger marginal costs, I compare it against merger simulations and UPP values that are computed with *no* cost synergies ($e = 0$). Furthermore, to conduct the comparison on the market-level (since the HHI is by definition a market-level variable, while the UPP is an airline-market level variable), I compute the sums of the UPP values and simulated price changes from America West and US Airways in each market. In other words, I first compute the pre- and post-merger HHI, assuming that the post-merger market shares of all non-merging goods stay constant at the pre-merger levels. I then compare the increase in HHI against the *total* price increase across the two merging firms in the same market as predicted by a merger simulation and the UPP.

Table 10 shows the correlation matrix between the three market-level variables: increase in HHI, market-sum of UPP's for both America West and US Airways, and market-sum of simulated price increases for the two merging firms. Naturally, the UPP results have a stronger correlation with the sum of simulated price increases than the HHI, because the UPP is in fact a simplified merger simulation. The increase in HHI is positively but moderately correlated with the (market-sums of) UPP and simulated price increases, confirming that increase in concentration is correlated with increase in price. The same word of caution applies as before, that the UPP values are computed from "true" diversion ratios and therefore gives the best prediction possible short of a full-blown merger simulation. When one wishes to compute the UPP without a structural demand estimation, in the spirit of Farrell and Shapiro, the deviation between the UPP and a merger simulation result will widen. The correlation between HHI and these two variables will likely be lower than the values shown.

1.7 Conclusion

The invention and popularization of the Upward Pricing Pressure test marks a new era in the practice of antitrust when the traditional approach based on market definition and concentration is widely deemed problematic. Farrell and Shapiro's UPP test has appeal in being easy to compute, low in data requirement, and free of market definition. This paper first establishes the close theoretical link between the UPP and the standard structural merger simulation with Bertrand firm conduct and differentiated products. I show that the UPP is a much simplified merger simulation that ignores the re-equilibration of all other merging products' endogenous prices and quantities. Using the structural merger simulation as the "gold standard", I demonstrate that the UPP performs well in both sign and magnitude predictions when it uses the "true" diversion ratios implied by the estimated demand model. Thus, the UPP is not a bad substitution for the last step of merger simulation, namely the computation of the new equilibrium from the set of post-merger first order conditions. However, practitioners may not have access to these "true" diversion ratios because the UPP is originally intended to be an easy test that does not require a demand estimation. I therefore use the example of a simple logit demand to illustrate the consequence of using inaccurate diversion ratios in the UPP: the test will give a wrong sign prediction over a much larger range of cost synergies than before. This phenomenon is due to the almost equal pass-through rates between the UPP and a merger simulation, which are largely constant in the relevant range of costs. This finding is general to all merger analysis and is not idiosyncratic to the airline markets or the particular airline merger. The most severe source of approximation error in the UPP therefore comes in the use of inaccurate diversion ratios. I also direct attention to the pass-through assumptions in Farrell and Shapiro (2010)'s proposition, which has so far received little attention. It assumes that the post-merger equilibrium prices of both merging products have positive pass-through in its own cost, and non-negative pass-through in its merging partner's cost. I point out that these conditions are equivalent to requiring the merging goods to be strategic complements post-merger. It has been demonstrated theoretically that this assumption is not innocuous even in the familiar context of the Bertrand model.

I empirically show that when the merging products are strategic substitutes instead of complements, the UPP produces false positive results (type I errors). Furthermore, when the pre-merger market is larger than a duopoly, a much larger set of pass-through assumptions is required, and I show that their violation may lead the joint prediction of the UPP to fail. Lastly, I compare the UPP with the traditional HHI test, using both the 1992 and 2010 thresholds, and conclude that the HHI is a more severe test in both cases.

This study of the UPP is inherently limited because I do not have information on the actual mergers analyzed by the antitrust agencies and their UPP values used. In particular, accuracy of the diversion ratio used has the largest impact on the performance of the UPP. This paper mostly uses the "true" diversion ratio derived from the estimated demand model, which gives the best possible prediction in the UPP against my gold standard. The real test of this merger screen is, admittedly, not on its best possible value, but the value that is actually used by practitioners. It will be instructive to repeat this study in the future when the UPP has been applied broadly. More research is also needed on improving the accuracy in alternative estimation methods of the diversion ratio. Researchers should investigate the best available control variables (or alternative econometric techniques) that identify the substitution pattern of the marginal consumer in a regression between quantities, since quantities data is likely available in most cases. This regression outcome can then be compared against a customer survey result in a merger where the latter is available as well.

2 Brand Portfolio and Consumer Learning

2.1 Introduction

Most modern brands contain a very large number of products, and new product introduction and product repositioning are particularly prevalent in supermarket consumer goods. Brands are naturally concerned about how the composition of their product portfolio affects the overall perception of the brand. I model how products under a brand that are not direct competitors, but are nonetheless related, may affect the perception of the overall brand, hence the demand for its products. This paper does not investigate the effect of a newly introduced product on the existing brand per se, but a closely related topic: how the different components of a brand affect the overall perception of the brand. In particular, I compare brands in the salted snack category on the variety of the their product portfolios, and investigate the difference in consumer learning on an unobserved brand quality due to this difference in portfolios. Some brands, like Pringles and Lays, focus exclusively on a narrow or single product category: potato chips. Other brands, like Wise, have a truly big variety in their products offers, and at times carry all eight categories of salted snacks seen in my dataset. (Chapter 3 defines these eight snack categories.) A brand's product portfolio changes only slowly over time. I am able to exploit the exceptionally long time frame (six years) of my dataset and observe a variation in snack categories covered in most of the brands. This vital source of variation helps identify its effect on purchase probabilities, which in turn is determined by consumer's current belief of the brand's (unobserved) quality.

An innovation of this paper is the use of cross-category purchase within brand to gauge consumer learning. While I model brand competition within the potato chip category alone, I observe consumer purchases in other snack categories of these brands. My model acknowledges these other purchases as inputs to the consumer's learning process of the unobserved overall brand quality. Unlike my approach, most empirical papers on consumer loyalty, learning, or advertisement rely on data from a single product category only, such as laundry detergent, yogurt, or orange juice. The same can be seen in the big body of literature that revolves around the estimation of flexible, random coefficient discrete choice demand models, such as that of automobiles or ready-to-eat breakfast cereals. A likely cause is the unavailability of data in multiple related product categories, such as the eight salted snack categories in this paper. This is not to say that focusing on a single product category is always unreasonable. This depends on the task at hand and particular products investigated. For example, the purchase of an automobile is usually an independent endeavor where the consumer encounters few other transactions. In contrast, because of the sheer size and collection of modern supermarkets and the proximity of related products on supermarket shelves, consumers obtain a lot of information during the shopping process even when they purchase only a single item. Products that are not traditionally thought to be direct competitors (e.g. potato chips vs. pretzels) do not exist in physically isolated markets. The cross-category learning model in this paper is one hypothesis on how various product categories may exert an influence on each other in ways other than direct competition.

"Brand erosion" or "brand equity dilution" is a recurring theme in consumer retail and marketing, and one of the most discussed in relation to brand portfolio choice or composition. Dacin and Smith (1994) is closely related to this study in its theme and investigates the relationship between brand portfolio characteristics and consumers' brand evaluation. The authors use two laboratory experiments and a survey to gauge consumers' confidence in brands and their evaluation of brand extensions. The brand characteristics considered by these authors are similar to those in this paper, which include the total number of products carried by a brand, and a perceived quality variance of these products within the brand. Not all of their findings are replicable across the two methodologies: a positive relationship between the number of products in a brand and consumer confidence can only be found in the laboratory experiments, but not the survey. This is not in disagreement with this paper's findings: I find a negative and statistically significant effect on the product-count variable, but the marginal effect is small and its real economic significance is doubtful. However, when the portfolio quality variance decreases, these authors' methodologies unanimously conclude a positive relationship between the number of products and consumer evaluation of a brand extension. While the reduced form regressions in my paper does not explicitly account for quality variance and its effects, a forthcoming paper that estimates the full structural consumer learning model will. My full model allows different product categories within a brand to have different variances in an unobserved quality. Combined with variation in observed purchase patterns, one can then analyze the effect of quality variance on consumer learning and purchase behavior. Meyvis and Janiszewski (2004) is another, more recent study on product category extension in brands and the strength of the brand. Their hypothesis on the accessibility of brand category associations predicts that a brand with a more varied portfolio produces more successful extensions than a brand with a narrower offering. It presents another possible extension to my model in the analysis of categories under a brand: an empirical estimation of the "similarity" between these categories. The snack categories in my model are not inherently distinguished in the sense of, for example, a category-specific quality or attribute; my current model does not estimate a common, unobserved product characteristic for all pretzel products vs. potato chip products. While a category-specific dummy variable is commonly used in many kinds of regression analysis, a category-specific term is more difficult to incorporate in a Bayesian setting if the researcher still wants to model the learning of a cross-category common quality of the brand. A consumer in this model will have to disentangle the category-specific term from the common quality term of the brand before Bayesian updating can take place. This "disentangling" may be justified if the consumer is observed to have made multiple purchases across both categories and brands.

In this paper I present a structural model of consumer learning through multi-category purchases in the Bayesian setting, and estimate a discrete brand choice model that includes reduced form representations of brand portfolio variation and consumer cross-brand purchases. The earliest framework of consumer brand choice models, such as that of Guadagni and Little (1983), assumes that the consumers have complete information on brand and product attributes, and that they are myopic and maximize immediate utility only. A sizable literature has since develop to address these two limitations, by explicitly modeling individual consumer's learning process on some unobserved product attributes, and extending the static decision problem to a dynamic one that allows strategic, intertemporal tradeoffs.

Consumer learning is only one among a few broad applications of single agent dynamic problems. Modeling technology adoption or the switching cost from old to new environments is another such occasion, with examples such as Yang and Ching (2009) and Ryan and Tucker (2011). A static treatment of technology adoption will almost surely bias the estimates, because both the cost of adoption and the stream of benefits from adoption are bore over time. If the researcher erroneously assumes that the cost of adoption is bore entirely in the one single period when adoption occurs, the estimates will either give a too-low adoption cost (if this is the object to be estimated) or a too-high preference for the new technology, in order to justify the heavy cost.

Another application of the single agent dynamic programming problem is the modeling of consumer stockpiling behavior. Both consumer survey data and most people's personal experience indicate that the rate of consumption of supermarket products does not synchronize with the observed frequency of supermarket shopping trips. A consumer is likely to shop and stockpile for a time period that lasts further than his next shopping trip, especially when a product is on sale. This implies that if the researcher wrongly assumes that all products bought are to be consumed immediately in the same period, the estimates will indicate a price sensitivity or demand shock that is larger than reality. Recent papers on stockpiling include Hendel and Nevo (2006b), Hendel and Nevo (2006a), and Erdem, Imai, and Keane (2003).

Empirical research on consumer learning, loyalty, and habit formation has a long history. One of the biggest literatures on consumer learning is the study of the effects of advertising. In particular, various forms of advertising (such as "informative" vs. "prestige") are modeled to have different effects on a consumer's information set. Often, the consumer's updates on product attributes from purchase experience or advertising exposure are modeled in the Bayesian framework, and recent examples include Erdem and Keane (1996) and Ackerberg (2003). In Erdem and Keane (1996), consumers are unsure about a particular attribute ("quality") of the products. A consumer's information set includes parameters on his current belief of the distribution of this uncertain attribute, and learning is defined as Bayesian updating of these parameters when the consumer receives a new draw on this product's quality, either from a purchase or from an exposure to advertising. Since the "true" values of this product attribute is never known to a consumer, he makes purchase decisions based on the expected values of this attribute in each alternative, formed from his current beliefs on their distributions. Ackerberg (2003) tracks consumers' advertising exposures in newly introduced, non-durable grocery products. Its structural learning model explicitly takes into account two different effects of advertising, "informative" vs. "prestige." The "quality" of this new product, as suggested by "informative" advertising, enters the consumer's utility function directly. The consumer's current posterior mean on the intensity of advertising also enters his utility function and the author interprets this as "prestige" advertising or information signaling. The paper finds strong, significant effects for both types of advertising. The structural model allows the author to conduct welfare analysis, and he concludes that although information signaling is beneficial to the consumer, it comes at a much larger cost and is therefore socially inefficient.

2.2 Model

I consider a general model of demand in which consumer i chooses among J alternatives sequentially in each purchase occasion t. Each brand j may produce products in more than one product category. I model the consumer's decision problem within a single product category c only, although the consumer's purchases in other related categories in the set C form a crucial part of his expectation. Alternatives are defined as exhaustive and mutually exclusive brands within the category of interest. I do not include an "outside good" or "no purchase" option in the set J because I only model consumer choice when an actual purchase is recorded in the data, which I define as a purchase occasion. Each purchase occasion can be mapped back to a calendar week in the data. Because the panel data is unbalanced in terms of calendar weeks or purchase occasions, any purchase occasion, $j = 1$ for example, can occur in different calendar weeks among agents in the panel data.

2.2.1 Consumer Expected Utility

I start by defining the aspect of uncertainty in the consumer's problem and how it affects the consumer's decision. The consumer makes a discrete choice on brands among several available alternatives, choosing the option that gives him the highest utility. The conditional utility for consumer i in choosing alternative j in product category c at purchase occasion t is a function of both observed and unobserved attributes of the available alternatives:

$$U_{ijct} = -\alpha P_{jct} + \beta X_{jc} + \theta_1 A^E_{ijct} + \theta_2 (A^E_{ijct})^2 + e_{ijct} \tag{1}$$

Observed attributes include the price P_{jmt} and observed product characteristics X_{jm}. The term A^E_{ijct} denotes the consumer's experienced "quality" of the product, which is unobserved at the time when purchase decision is made, and only realized after consumption. The last term e_{ijmt} is the i.i.d. logit error term that reconciles any unexplained deviation between observed and predicted purchase patterns. The parameters $(\alpha, \beta, \theta_1, \theta_2)$ are the consumer's tastes for these respective attributes and objects to be estimated. The consumer's uncertainty lies in the term A^E_{ijct}, which will be decomposed into constant (but unknown) and random components. The constant component is the object of consumer learning, since the "true" quality is never known to the consumer or the econometrician. I consider a general learning model in which, even after purchases, quality is only experienced with noise:

$$A^E_{ijct} = A_j + \xi_{ijct} + \eta_{ijct}, \qquad \xi_c \sim \mathcal{N}(0, \sigma^2_{\xi c}), \qquad \eta_{ijct} \sim \mathcal{N}(0, \sigma^2_\eta) \tag{2}$$

The first term A_j is the unknown "true" quality of brand j, which are objects that the agents try to recover through their purchase experiences in each brand j. The next term ξ_{ijct} is a normally distributed, mean zero deviation in quality; each product category has its own variance $\sigma^2_{\xi c}$ on ξ_{ijct}. Finally, η_{ijct} is a normal, mean-zero disturbance term that is agent- and purchase occasion-specific. These three terms are never individually observed by the econometrician or the consumer, thus the consumer is never completely certain about the "true" quality A_j even after many purchases, although this "true" quality is unchanged over time. More observations on this attribute through purchases do improve the consumer's "guess" at the time of decision, and this form of consumer learning is modeled explicitly in a Bayesian manner detailed below. If these disturbance terms are absent, this general learning model reduces to a simple 1-period learning, in which the "true" quality of a brand is revealed to the consumer immediately after the first purchase. This more general learning model is justified when all relevant attributes of a product is not necessarily experienced within a short period of time after consumption. For my specific application on the choice between brands of potato chips, it could be argued that the consumer needs a longer horizon to experiment with food pairings or to realize the full health effects of particular choices, such as the difference in fat content between fried and baked potato chips. Going back to my general definition of experienced quality, I further group the two disturbance terms into one, because they are not individually observed:

$$A^E_{ijct} = A_j + \delta_{ijct}, \qquad \delta_{ijct} \sim \mathcal{N}(0, \sigma^2_{\delta c}) \tag{3}$$

The variance of δ_{ijct} is indexed by c because the variance of ξ_{ijct}, a component of δ_{ijct}, is also indexed by c. Continuing with the conditional utility function, the quadratic term involving A^E_{ijct} reflects the consumer's possible risk aversion towards an uncertain product attribute. If the consumer is indeed risk averse, the coefficient of the quadratic term θ_2 will be estimated to be negative. Lastly, since absolute levels of utilities are not identified, I assign one of the alternatives as the "base alternative" and normalize its utility level to zero in my estimation.

Since I focus on competition within a single snack category c (potato chips) only, the c subscript can be omitted in the agent's brand-choice problem within this single category:

$$U_{ijt} = -\alpha P_{jt} + \beta X_j + \theta_1 A^E_{ijt} + \theta_2 (A^E_{ijt})^2 + e_{ijt} \tag{4}$$

Because experienced quality A^E_{ijmt} is not realized before purchase decision, agent forms expectation on it conditional on current information set $I_i(t)$. Thus agent thus makes purchase decision based on the expected utility of each alternative j:

$$\begin{aligned} E[U_{ijt}|I_i(t)] = \quad & -\alpha P_{jt} + \beta X_j + \theta_1 E[A^E_{ijt}|I_i(t)] + \theta_2 \left(E[A^E_{ijt}|I_i(t)] \right)^2 \\ & +\theta_2 E\left[(A^E_{ijt} - E[A^E_{ijt}|I_i(t)])^2 \right] + e_{ijt} \end{aligned} \tag{5}$$

When the consumer is risk averse (i.e. when $\theta_2 < 0$), his expected utility for brand j will be a concave function of A_{ijt}^E and a linear function of the perceived variance in A_{ijt}^E. The information set $I_i(t)$ contains parameters that define the posterior distribution of A_j from time $t-1$. This is also the prior distribution used at time t when the agent evaluates his expected utilities, before the purchase decision is made and the experienced utility is realized.

Relating the above model to the structure of my data, I have two broad possible definitions of what constitutes an "alternative." Since the lowest level of observation in the data is the UPC (universal product code), I can theoretically model the consumer's problem as a choice between unique products. However, due to the very large number of unique UPC's observed in the data, the resulting thin observations in many of them, and the lack of good product attributes that distinguish between each single UPC, a choice model on UPC is difficult to estimate and likely suffers from mis-specification issues. Thus, I have chosen to model the consumer as choosing between brands instead when making a purchase (although I do observe the actual UPC bought under that brand). Aggregating the alternatives of choice helps with model specification because brand statistics can be computed from the lower-level data. These statistics are unique to each brand and therefore helps identify a consumer's brand choice. The brand statistics I will use include a consumer's purchase history with the brand, the number of unique UPC's it has, and the number of snack categories it covers. Detailed variable descriptions can be found in the following section.

2.2.2 Consumer Learning

I assume that consumers learn about the unobserved brand quality A_j following the Bayesian updating mechanism, where each experienced quality level from a purchase (in any product category covered by this brand) serves as a draw that gives an observation of A_j with noise. With each draw, the consumer then updates his prior belief on the distribution on A_j, and forms a posterior. This posterior is then next period's prior distribution. I first define the perception error or "surprise element" in brand quality as the difference between the actual quality experienced after consumption and expected quality formed before the purchase is made:

$$\nu_{ijct} = A^E_{ijct} - E[A^E_{ijct}|I_i(t)] \tag{6}$$

Because the experienced quality A^E_{ijct} is specific to each agent, category, and purchase occasion, the perception error will be so as well, although the underlying variable that agents try to recover, A_j, is only brand-specific. Also note that because the agent's beliefs $I_i(t)$ is updated every period, his expectation $E[A^E_{ijct}|I_i(t)]$ is updated accordingly as well. Thus, even if the same random experienced quality A^E_{ijct} is realized twice in the agent's lifetime, the perception errors will be different in these two purchase occasions.

The choice of normal distribution in the disturbance terms ξ_c and η_{ijct} to experienced quality greatly facilitates the computation of the posterior, because the theory on conjugate prior distributions can be applied. Since the Gaussian family is self-conjugate, when the likelihood function is Gaussian, using a normally distributed prior will ensure that the posterior distribution is also normal. An additional advantage is that the posterior distribution has a convenient closed-form expression, thus avoiding the need for a difficult numeric integration.

To emphasize cross-category learning of a common quality within a brand, both purchases in this and other categories within the same brand contribute to the posterior of A_j, which is also the prior at the next purchase occasion. Appealing to the theory of conjugate priors, the expected value of brand quality A_j is updated as follows:

$$E[A_j|I_i(t+1)] = E[A_j|I_i(t)] + \sum_c D_{ijct}\gamma_{ijct}\nu_{ijct} \tag{7}$$

where $D_{ijct}, \forall c$ are purchase indicators that equals one when a purchase is made by agent i in the product category of interest under brand j at occasion t, and zero otherwise. The weights γ_{ijct} assigned to each perception error term are Kalman gain coefficients, to be explained below. Brand qualities are assumed to be uncorrelated, thus each brand has its own independent updating mechanism similar to the above. Because "time" is indexed according to purchase occasion and not a calendar week, every iteration from $I_i(t)$ to $I_i(t+1)$ indeed represents a non-trivial revision in the posterior distribution of a brand's quality; specifically, it is the brand that was purchased at occasion t. Intuitively, the above updating equation states that if the agent has a negative perception error today (i.e., his expected value today was too high), from any of the categories he purchased within this brand, his expectation tomorrow will be adjusted downward, and vice versa. Because the "time index" t is defined as a purchase occasion instead of a calendar week, only one of the purchase indicators D_{ijct} will be equal to one at any purchase occasion, among all the product categories c available under brand j. The Kalman gain coefficients are inversely related to the variance of the disturbance terms:

$$\gamma_{ijct} = \frac{\sigma^2_{\nu ijct}}{\sigma^2_{\nu ijct} + \sigma^2_{\delta c}} \tag{8}$$

When $\sigma^2_{\nu ijct}$ is constant across all product categories c, the gain coefficient γ_{ijct} is decreasing in $\sigma^2_{\delta c}$, which is category-specific, because the variance of the disturbance term ξ_{ijct} in A^E_{ijct} is category-specific. In other words, more weight is given to observations in product categories that are known to have a smaller noise component, compared with a hypothetical purchase in another product category under the same brand, at that particular purchase instant. Also note that variance $\sigma^2_{\nu ijct}$ is specific to each purchase occasion t. It therefore is updated after every purchase for each individual i, along with his posterior beliefs. By the theory of conjugate priors in DeGroot (1970) and generalizing a similar formula in Erdem and Keane (1996), the variance of the perception error is given by

$$\sigma^2_{\nu ijct} = \frac{1}{\frac{1}{\sigma^2_{\nu 0}} + \sum\limits_{c \in C_j} \frac{\sum_{s=0}^{t} D_{ijcs}}{\sigma^2_{\delta c}}} \tag{9}$$

where $\sigma^2_{\nu 0}$ is the variance of of the perception error at $t = 0$, which can be exogenously chosen. The remaining terms in the denominator consists of one fraction for each product category available under brand j, and this set of categories under brand j is labeled C_j. Each fraction consists of the total number of purchases made in that category over the consumer's lifetime, divided by that category's disturbance variance $\sigma^2_{\delta c}$. Thus both the brand's total number of product categories offered and the agent's experience in each of these categories affect his perception error variance. One can recall that the agent's expected utility in equation (5) contains a term on the expected variance of the disturbance term. In other words, the components in $\sigma^2_{\nu ijct}$ can be thought of as reduced form explanatory variables in a consumer's decision rule. This justifies my use of the agent's history of purchase with brand, as well as the number of product categories a brand offers, in my regressions. In general, one can use polynomials of these statistics mentioned as regressors to better approximate a decision rule that involves $\sigma^2_{\nu ijct}$, since these statistics do not appear as simple linear functions in equation (9) either.

A possible extension in the future is to relate the different contributions from various categories in consumer learning to these categories' different characteristics. What makes a particular category more indicative of its brand's quality? Maybe it is the category's history within the brand, the frequency of category-specific advertising, the category's share in the brand's total number of products, or soft indicators such as the uniformity of packaging and trade dress. The researcher can explore whether there are statistically significant relationships between these category characteristics to the category disturbance variance $\sigma^2_{\delta c}$.

2.2.3 Consumer's Dynamic Optimization Problem

I now move on to the consumer's dynamic optimization problem across his planning horizon. Since the agent's state variables—his information set $I_i(t)$—is revised after each purchase, and affects all future utilities by the accuracy of his expected utility computations, it is reasonable to the presence of intertemporal decisions and tradeoffs, hence the need for a dynamic programming problem. Intuitively, a purchase may be made today for information seeking reasons, although that alternative may not give the highest expected utility with the agent's prior today. Formally, I assume that the consumer maximizes his expected present value of utility over a finite planning horizon T. Given a consumer's series of discrete decisions $\{D_{ijct}\}$, his discounted present utility can be written as

$$
\begin{aligned}
V_i(I_i(t),t) &= \max_{\{D_{ijct}\}} E\left[\sum_{\tau=t}^{T} \rho^{\tau-t} \sum_{c\in C}\sum_{j\in J_c} E[U_{ijc\tau}|I_i(\tau)]D_{ijc\tau}|I_i(t)\right] \\
&= \max_{j\in J_c} V_{ijc}(I_i(t),t),
\end{aligned}
\tag{10}
$$

where $D_{ijc\tau}$ is purchase indicator, $\rho > 0$ is discount factor, and $V_{ijc}(I_i(t),t)$ is the choice-specific expected value function. J_c is the set of brands available in product category c. The consumer is allowed to purchase from multiple product categories, although each category is considered a separate market. In other words, the consumer makes a purchase in the single brand that gives him the highest expected utility in a category in each purchase occasion, and brands (or products) across categories are not modeled as competitors. I do not model quantity choice in these purchases, and assume all purchases are of one unit. One can write the lifetime value function after today's choice is made, based on today's state variables $I_i(t)$. Using the choice-specific value function, one can form the recursive Bellman equation:

$$
\begin{aligned}
V_{ijc}(I_i(t),t) &= E[U_{ijct}|I_i(t)] + \rho E[V_i(I_i(t+1),t+1)|I_i(t), D_{ijct}=1], \\
&\forall t = 1,...,T-1.
\end{aligned}
\tag{11}
$$

The Bellman equation implicitly states that all future decisions will be made optimally following the new information obtained from today's purchase. After today's consumption, experienced utility will be realized, and the agent is able to form his posterior, and update his information set from $I_i(t)$ to $I_i(t+1)$. This new information set, as the prior distribution in next period $t+1$, is sufficient to for the agent to make his optimal brand choice next period. The choice-specific value function for period $t+1$ can thus be generated, and the econometrician can iterate the Bellman equation one period forward. An agent's entire lifetime discrete choice consumption path can be computed in this manner. At final period T, $V_{ijc}(I_i(T),T) = E[U_{ijcT}|I_i(T)]$. The second term above goes to zero because there is no more future period or future utility to consider.

Note that the expectation taken over a consumer's utility in all equations above are taken over the uncertain experienced brand quality A_j^E. It is not taken over the logit error term, which is known to the consumer (though not the econometrician) at the point of decision and therefore contains no uncertainty in the consumer's point of view. Logit probabilities are well known to have nice analytical forms, and this is not lost in our dynamic decision problem. When the random i.i.d. logit error in the consumer's utility is integrated out, this set of dynamic optimization equation gives choice probability in the "standard" logit form, albeit containing a recursive, forward-looking term inside the exponential for each alternative, in addition to the current expected value of utility from the observed and unobserved product attributes. The logit brand choice probability, for brand j in category c, is then

$$Pr_{jc}(I_i(t), t) = \int_{\nu} \frac{\exp(\overline{E}_{ijct})}{\sum_{k=0}^{J} \exp(\overline{E}_{ikct})} f(\nu) d\nu \tag{12}$$

As a reminder, $\nu = E[\widehat{A}_{ijmt}|I_i(t)] - A_{jm}$ is agent's perception error; it is unobserved by econometrician and hence is integrated out as well. The term inside the exponential gives the sum of immediate utility from this purchase occasion and the present value of all future occasions, optimized based on today's updated posterior distribution on brand qualities:

$$\begin{aligned} \overline{E}_{ijct} = \quad & -\alpha P_{jct} + \beta X_{jc} + \theta_1 E[A^E_{ijct}|I_i(t)] + \theta_2 \left(E[A^E_{ijct}|I_i(t)] \right)^2 \\ & + \theta_2 E\left[(A^E_{ijct} - E[A^E_{ijct}|I_i(t)])^2 \right] + \rho E[V_i(I_i(t+1), t+1)|I_i(t), D_{ijct} = 1] \end{aligned} \tag{13}$$

In other words, $\overline{E}_{ijct}$ contains all the terms in expected utility (5) minus the i.i.d. logit error term, plus discounted future utilities. The last term is the forward-looking component, present in every period except the last when the consumer makes a decision. Again, its presence means that the consumer may make intertemporal tradeoffs by exchanging immediate utility with future utility. For example, the ex ante expected utility of an alternative may not be the highest for a person, but nonetheless he may choose it today for exploratory purpose, because the new information he gains from this purchase may improve his future utilities, with the discovery of a high quality product. If the agent maximizes immediate utility only, the discount factor ρ equals zero and the dynamic programming problem reduces to a repeated static problem. Notice that in this special static problem, the consumer's state variable, his information set $I_i(t)$, is still updated after every purchase, and it still affects his future decisions through his expected utility computations. The subtle difference is that in the static problem the agent will have no incentive to make the kind of intertemporal tradeoff mentioned above. There is no strategic planning in making an inferior purchase today for the benefit of new information. Instead, the alternative that gives the highest expected utility today is always chosen, and any information gathered is a mere afterthought.

2.2.4 Identification

Intuitively, acquisition of new information changes consumer expectation and their brand choice decisions. Thus, identification of consumers' learning parameters comes from the change in purchase behavior after an observed number of purchases. At the extreme, if there is no learning involved in these purchases, or that the brands in fact have no unobserved characteristics to be learned (or that, econometrically, no brand-specific information can be extracted from the logit demand residuals), consumers' purchase behavior should not change over time after any number of observed purchases. In this extreme case, the consumers' believes are never updated and the estimated variance to the perception error should be zero. If the brand-specific quality does exist but is observed without noise, a consumer learns about it immediately after his first purchase. Thus, theoretically, consumer behavior should converge immediately after the first purchase and stay constant, controlled for other demand factors. The unobserved brand quality can be identified by comparing the purchase behaviors between consumers who have never made the purchase, versus those who have made the first purchase. The estimated variance to the perception error should again be zero. With general learning from noisy signals, consumers' believes about brand qualities will ultimately converge to the "true" values. The econometrician can then use the observed purchase behaviors of consumers who are observed to have learned through many purchases to estimate the "true" brand qualities. The variance in the perception error term is captured by how "fast" the consumer purchase behavior settles down or converges to that implied by the "true" qualities, and the difference in this variance across snack categories is identified by different learning "speeds" (or the number of purchases required till convergence) of two consumers who, say, exclusively purchase potato chips and pretzels, respectively.

A category with noisier signals of quality will lead to a slower convergence in consumer purchase behavior.

2.2.5 Reduced Form Implication

The static version of the above brand choice problem can be approximated as a multinomial logit model with each available brand as an alternative, brand-specific intercepts, and price and product characteristics as regressors. The expected quality and Bayesian learning component in the utility can be approximated with functions of individual consumers' past purchases and brands' portfolio components at that time. In this paper I estimate a static brand choice model with the aforementioned explanatory variables, forming an alternative-specific conditional (ASC) logit regression. A forthcoming accompanying paper will estimate the full fledged structural Bayesian learning model and conduct counterfactual experiments from its results. The dynamic version of the model does not lend itself to an easy reduced form simplification because of its state dependence every period and the possibility of intertemporal tradeoff.

In the static brand choice model I use an exponentially smoothed, brand-specific purchase history of each individual and the number of snack subcategories a brand covers as approximate measures of cross-category learning on brand quality. Although these variables are arguably crude, they allow me to verify whether a brand's portfolio spread has an effect on purchase probability. Using purchase history together with brands' UPC-count as control variables, I am able to compare the choice probabilities in a hypothetical situation where two brands are of the same size but but of different portfolio spreads. (For example, both brands A and B contain 50 unique products, but brand A's 50 products all belong to the potato chips category, while brand B's 50 products are spread between different snack categories.) It is important to control for a brand's size because this affects the brand's supermarket shelf space and its exposure to consumers, among other facts. Each brand is given its own brand-specific intercept. Lastly, I include two demographic variables, household size and family income, in the last specification, and allow these variables to have brand-specific coefficients. These demographic variables are not explicitly featured in the structural model but are included in the regression to test the statistical significance of the main variables of interest.

While a reduced form regression is much easier to estimate, a structural model provides the benefit of credible counterfactuals and more accurate out-of-sample forecasts. The difference lies in the reduced form decision rule that incorporates elements of the perception error variance (9) in a linear or polynomial fashion into the agent's expected utility (5). The coefficients estimated from this demand model with reduced form brand purchase history and brand portfolio spread as regressors will be functions of deep parameters, such as $\sigma^2_{\delta c}$. When one uses these demand parameters to generate an out-of-sample forecast in the changes of, say, a brand's portfolio components, one implicitly assumes that these demand parameters remain constant to these changes. However, it is evident in equation (9) that these deep parameters change with the state variable of the structural model, which depends on our variables of interest (say, a brand's portfolio components). In contrast, when one conducts a counterfactual experiment using an estimated structural model, one computes the new state variables under the new conditions in order to generate the forecast. This is an example of the Lucas's critique, that reduced form parameters are in fact not invariant to policy changes. Thus, a reduced form model should at best be used as a proof of concept, to verify that a variable of interest does have effect on outcomes. This is what I do in this paper. The accompanying paper will estimate the structural Bayesian learning model and estimate counterfactuals on changes in brands' portfolio components.

2.3 Data

The dataset was purchased from SymphonyIRI. This supermarket scanner dataset spans 31 major product categories over a six year period (2001-2007). Sales is recorded in the weekly frequency for each UPC (universal product code) under the included product categories. It covers 47 geographic markets in the U.S., and shopping venues include chain grocery stores, drug stores, and mass-market outlets. When a particular store chain is selected for inclusion in a geographic market, all stores under that chain in that market are included in the dataset. Although store chains are not identified by name, they are identified by unique numeric chain ID's. Each individual store is also identified over time by a unique ID, together with attributes such as geographic market, and open and close dates with each chain, should there be a merger or acquisition involving that store. In the scanner dataset, each observation is a store-week-UPC combination, with information on total unit sales, total dollar sales, and three marketing mix variables: feature, display, and a price reduction flag. All UPC's also come with product attributes including its parent company, vendor, brand, and a string variable containing a textual description of the product, including its name. Most categories also contain a category-specific volume-equivalent variable for easy comparison of volume between products within the same category. Certain categories also contain product additional attributes specific to its own; for example, products in the "potato chips" category has fields such as "cut" and "fat content."

Apart from the country-wide scanner data on weekly sale statistics, this dataset also includes a household-level panel from two BehaviorScan markets: Pittsfield, MA and Eau Claire, WI. This panel dataset covers the same six year period and 31 product categories. Each household is identified with a unique panel ID over time, and the time of each of their purchases is identified to a particular shopping trip within a week, where there can be multiple. In addition, each household is also associated with a big set of demographic variables, including combined pre-tax income, family size, race, and age group and education level of both the male and female heads of household, etc.

Because of the long time horizon and complete coverage of all purchases within the broad product categories, this panel dataset is a excellent resource for studying dynamic consumer behavior. The substantial length of time spanned (six years), longer than most other supermarket scanner datasets, is particularly conducive for research on consumer product brand portfolio, which changes only infrequently over time. I have chosen the "Salted Snack" category for investigation on the effect of brand portfolio composition because it contains products of a wide variety. The snack category contains eight subcategories: cheese snacks; corn snacks (no tortilla chips); other salted snacks (no nuts); pork rinds; potato chips; pretzels; ready-to-eat popcorn / caramel corn; and tortilla / tostada chips. Each UPC contained in the snack category is assigned to either one of these eight exhaustive, non-overlapping subcategories. I consider each of these subcategories as a separate market; for example, I only model competition between brands within the potato chip subcategory, but not competition between a potato chip brand and a caramel corn brand (unless there is another subcategory where both brands have a presence). Specifically, I define a brand's portfolio spread by counting the number of subcategories its products spans, over every 6-month period. (A 6-month period, as opposed to a shorter time frame for more frequent observations, is chosen to minimize the possibility of a product being offered by a brand but unobserved in the dataset due to lack of purchase, thus possibility skewing down the number of subcategories covered by a brand's portfolio and UPC-count.)

I use panel data from the Pittsfield market because, unlike Eau Claire, it does not have big-box stores (such as Walmart) in the neighborhood during the observed period that are excluded in the dataset, which might severely bias a model of consumer products competition. I choose the potato chips subcategory to model brand competition because it is the largest, and potato chip producing brands have a large variety in their portfolio spread. There are both grocery stores and drug stores in Pittsfield, and I include potato chip brands from both types of stores in my demand model because of the relatively small size of Pittsfield. In other words, I consider all potato chip brands available in Pittsfield to be in direct competition with each other, whether they are available in a grocery store or a drug store. I model a consumer's brand choice only, not his store choice.

A fine point on the definition of a "brand" is warranted here. Each UPC in the dataset has a field called "brand," but this is not the definition I use. My observation is that this definition given in the data is narrower than what is typically perceived as a core brand in snacks. For example, "Baked Lays," "Lays Kettle Cooked," "Lays Natural," "Lays Wow," and "Wavy Lays" are considered different brands in the dataset, while most consumers will simply associate them as "Lays." I therefore manually regroup all Lays "sub-brands" into one large "Lays" brand. Aggregation also eases convergence issues during model estimation due to thin observations in rarely occurring brand alternatives. After manually regrouping these given "brands," I have a total of

Table 11: Variation in Portfolio Spread Among Potato-Chip-Carrying Brands

Brand	Number of snack sub-categories carried by the brand											
	2001.1	2001.2	2002.1	2002.2	2003.1	2003.2	2004.1	2004.2	2005.1	2005.2	2006.1	2006.2
Bachman	6	4	4	5	5	5	5	5	5	5	5	4
Cape Cod	4	4	4	3	3	3	3	2	4	3	3	3
Cottage Fries	1	1	1	1	1	1	1	1	1	1	1	1
Gibbles	1	2	3	2	2	2	2	2	2	2	2	2
Herrs	1	5	4	2	2	2	2	2	1	1	1	1
Kettle	3	3	3	3	4	2	2	2	4	3	4	3
Lays	1	1	1	1	1	1	1	1	1	1	1	1
Pringles	1	1	1	1	1	1	1	1	1	1	1	1
Ruffles	1	2	2	2	2	1	1	1	1	1	1	1
State Line	2	2	1	1	1	1	1	1	1	1	1	1
Utz	6	6	5	5	5	5	5	5	5	5	6	8
Wachusett	1	1	1	1	1	1	1	1	1	1	1	1
Wise	7	6	8	7	7	7	7	6	6	7	7	7

Notes: In Pittsfield, MA. Tabulated every six months from 2001 to 2006. "2001.1" stands for the first six calendar months of year 2001, etc. The eight possible subcategories are: cheese snacks; corn snacks (no tortilla chips); other salted snacks (no nuts); pork rinds; potato chips; pretzels; ready-to-eat popcorn / caramel corn; and tortilla / tostada chips.

13 brands competing in the potato chips subcategory in my Pittsfield panel dataset. (I have discarded six brands which have extremely thin observations in the data to avoid difficulty in fitting their alternative-specific parameters.[11] All observations involving purchases in these six brands are dropped, and they are not included in the choice set of 13 available brands.) These brands span the full spectrum in terms of coverage in salty snack subcategories: some cover all eight subcategories mentioned above; some cover only one. Table 11 shows the variation on portfolio spread among the 13 brands over time every six months in the six year period observed.

More than half of the brands exhibit changes in the number of sub-categories covered over time, which I tabulate every six months. Herrs, for example, exhibits a gradual yet consistent decline in the range of snacks offered, going from offering cheese snacks, ready-to-eat popcorn, tortilla chips, and other salted snacks in 2001, to potato chips alone in 2005. Utz increases is offering from five categories in 2002 to all eight categories by the end of 2006, expanding into corn snacks, pork rinds, and ready-to-eat popcorn.

[11]These six brands are Grandpa Gibbles, Liebers, Mystic Chips, Nibble w Gibbles, Roberts American Gourmet, and Tastee.

Brands that exhibit no change in the number of subcategories offered during this six year period include Cottage Fries, Lays, Pringles, and Wachusett, all of which produce exclusively in the potato chip subcategory. It is possible that an observed drop in product offering is a result of regional or even chain-specific withdrawal, instead of a real, permanent change in that brand's portfolio, and that these withdrawn products might still be available in other nearby stores not covered in my dataset. Unfortunately, I cannot distinguish between these situations. An admittedly weak argument is that supermarkets are geographically concentrated in Pittsfield and residents are unlikely to make frequent shopping trips out of town. And because my dataset covers all major store chains in Pittsfield, any excluded stores within the area have limited marginal effect on the overall set of products available.

There are many possible levels of fineness in capturing a brand's portfolio spread. In tabulating the total *number* of snack subcategories offered as opposed to the actual *identities* of these subcategories, let alone the *ratio* in the number of unique UPC's between these subcategories within a brand, or even the product attributes of these UPC's, I am inevitably simplifying and losing some information. For example, the effect of a brand's presence in both potato chip and tortilla chip (two arguably more similar product) may be different from a brand's presence in both potato chip and pork rinds, yet this difference will not be captured in my model, because in both cases, the brand covers exactly two subcategories. One could further hypothesize that having an even or uneven presence in these two subcategories, say in terms of the number of unique UPC's offered, will have different effects on the brand's potato chip sale. These are valid concerns and viable paths for further research. The same can be said on the arguably crude UPC-count variable for each brand. This is introduced to control for the fact that bigger brands that offer more distinct products (whether concentrated in one subcategory or not) are likely to have more exposure to customers and, therefore, leads to higher probability of purchase. A unique UPC is given to a potato chip product unique to its brand, flavor, and size (among other possible attributes), and one can reasonably argue that different sizes (party-size vs. single-serve) of the same brand and flavor should not be considered different products, thus inflating the count of "truly" unique products. Unfortunately my dataset does not contain enough product attributes to allow me to identify "truly" unique products that differ only by size. Thus, to avoid further complication, I use a crude UPC-count as a measure of the size of a brand.

I consider every observed purchase of potato chips in the dataset as a "purchase occasion." I model each purchase occasion as the panelist making a choice among the 13 available brands. I therefore do not include the "outside good" as an alternative in my model, as I do not observe in the data when a consumer intends to make a potato chip purchase but finds the outside good to carry the highest utility, and thus ends up not making a purchase. This behavior is observationally equivalent to the consumer never having a demand for potato chip in the first place on that shopping trip. Since I do not observe a panelist's shopping list, I model *realized* demand and purchases only. For each panelist and each brand, the panelist's observed purchases thus form a raw series consisting of $\{0, 1\}$, where 1 indicates the panelist purchasing in that brand, 0 otherwise, where each element in this raw series is a single purchase occasion. I construct the panelist-brand-specific "brand history" variable as an exponentially smoothed series of this raw $\{0, 1\}$ series. Exponential smoothing gives more weight to recent purchases. I use a carry-over constant of 0.85 in my exponential smoothing, which is similar to the values used in Guadagni and Little (1983). Specifically, the exponentially smoothed series $\{bhist_n\}$ for panelist i of brand j is constructed as

$$bhist_n = 0.85 bhist_n + 0.15 x_{n-1}, \tag{14}$$

where

$$x_{n-1} = \begin{cases} 1 & \text{if purchase } (n-1) \text{ belongs to brand } k \\ 0 & \text{otherwise} \end{cases}$$

and where n indexes a purchase occasion observed in the data for panelist i.

I also construct a summary price statistic for each brand in each week, from the prices observed at transactions recorded in the data within that week. I first normalize all observed unit prices by the volume equivalent given in the dataset, to obtain a price-per-volume. I then compute the brand-week-specific averaged price using an unweighted average of the price-per-volume for all unique UPC's belonging to that brand observed

Table 12: DATA SUMMARY STATISTICS

variable	mean	std.dev.	min	max
Average price per volume[a]	3.915	0.705	1.99	6.351
Brand history[a]	0.492	0.308	$6.14e^{-34}$	1.000
Subcategory-count[a]	2.272	2.178	1	8
UPC-count[a]	53.682	16.844	2	76
Household income[b]	6.901	2.992	0	12
Family size[b]	2.631	1.334	1	6
No. of obs.	317174			
No. of cases[c]	24398			
No. of unique panelists	3683			

Notes: This data subset is sampled from panel dataset of Pittsfield, MA.
[a] Alternative-specific variable statistics, taken over obs. where $y = 1$.
[b] Panelist-specific variable statistics, taken over unique panelists.
[c] A case is defined as a unique purchase instant when a panelist is to choose a brand over a fixed set of available brands. Thirteen brands are available at each purchase occasion, as evidenced by the no. of obs. divided by the no. of cases.

in that week. I opt for an unweighted average, as opposed to weighting it by units sold, because realized sales is endogenous. An individual consumer does not observe realized sales when she makes her brand choice decision; but she does observe all available UPC's, their prices, and their volume. This unweighted averaged price-per-volume for each brand-week combination is used in my ASC logit model as the measure of price variation between brands.

In order to speed up convergence in model estimation, I use only a sample of potato chip purchase observations in the Pittsfield panel data. In particular, I use all observations occurring in the first week of every two months over the six year period. In sampling evenly across the six year period, I am not losing the variation in brand portfolio over time, and this helps identify its effect on the probability of purchase. Table 12 gives summary statistics of key variables used in my ASC logit estimation. For alternative-specific variables, their statistics are computed only over observations where the alternative is chosen ($y = 1$). For case-specific variables, which are time-invariant for each panelist, their statistics are computed only over unique panelist observations.

2.4 Results

I estimate a model of consumer brand choice in potato chips using McFadden (1974)'s alternative-specific conditional (ASC) logit model by maximum likelihood. An alternative is one of the 13 brands of potato chips available in Pittsfield, MA. The full model contains both alternative-specific variables (price, panelist's purchase history with the brand, the brand's subcategory-count, and its UPC-count) and case-specific (i.e. alternative-invariant within a case) variables (household income and family size). (The model reduces to the simpler multinomial logit if it contained only case-specific variables; it reduces to the conditional logit if it contained only alternative-specific variables.) In this model, each alternative-specific variable is given a single coefficient common across all alternatives, while each case-specific variable (including an intercept) is given 13 coefficients, one for each alternative. For example, a brand's weekly averaged price per volume is given a single effect, common across all brands, while household income is allowed to exert a different effect on each brand. This model is designed to capture the differentiation between brands as much as possible.

Table 13 shows the five ASC logit models I estimate. All of them include a set of 13 brand-specific intercepts; the last model also includes brand-specific coefficients for household income and family size. Coefficients for price and panelist's purchase history with the brand carry the expected signs: an increase in price in a brand decreases the probability of its purchase, while a stronger purchase history in a brand (due to more recent purchases) increases its purchase probability. Subcategory-count of a brand has a negative and significant effect on purchase probability. To control for the fact that some brands may have larger exposure to customers simply because they have more unique products to be displayed on shelves, I include a UPC-count for each brand in model (4). It has a small and negative coefficient, albeit statistically significant. (Both the subcategory-count and UPC-count variables are tabulated once every six months for each brand.) The negative signs of these two coefficients are suggestive that a more dispersed product portfolio and a larger number of products cause the brand to be chosen with lower probability, while holding its average price and the panelist's purchase history with it constant.

The sign and magnitude of the coefficients do not alter much between models. The 13 brand-specific intercepts are always (except that of Gibbles) statistically significant in the first four models, indicating that the brands are sufficiently differentiated. Moreover, most of them are positive, indicating that they are perceived as better than the base brand Wachusett (holding all else equal). This should be expected because Wachusett has the smallest number of observed purchases in the dataset used. In addition, the rank in magnitude between these intercepts is stable across the first four models. This rank is a reflection of the relative popularity between brands, as observed in the data in terms of purchase frequency. However, once the case-specific variables, household income and family size, and added to the regression on model (5), almost all of the alternative-specific coefficients become insignificant. While these demographic variables could be predictive on whether a panelist purchases potato chips, they seem to be insufficient in explaining a particular brand choice. Thus, I proceed with model (4) as my benchmark.

Since the ASC logit model is nonlinear, its coefficients cannot be directly interpreted as marginal effects on choice probabilities. I now take my benchmark model (4) and convert its coefficients into marginal effects. In particular, I focus on two key variables: panelist's purchase history with brand and the brand's subcategory-count. I tabulate a (13×13) marginal effects table for each variable, where an element (i, j) (on the ith row, jth column) in a table is the marginal change in the probability of brand i being chosen with an infinitesimal change in that variable for brand j products. All derivatives for the marginal effects are evaluated at the mean values of respective variables, thus they are marginal probability changes on the "average" brand with the "average" consumer (instead of the averaged effect evaluated at each observation). As tables 14 and 15 show, all own- and cross-brand marginal effects are statistically significant, albeit mostly of small magnitudes. They all have expected signs that agree with the regression results: elements on the diagonal of table 14 are positive and that of table 15 are negative, while the off-diagonal elements are of respectively opposite signs. In other words, a panelist's purchase history with a brand has a positive within-brand effect and a negative cross-brand effect, while a brand's subcategory-count has a negative within-brand effect and

Table 13: Alternative-Specific Conditional Logit Regression Results

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)		
Alt.-specific variables:							
Price	−0.598	−0.608	−0.609	−0.630	−0.628		
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
Brand history	–	3.629	3.629	3.629	3.613		
	–	(0.000)	(0.000)	(0.000)	(0.000)		
Subcat.-count	–	–	−0.101	−0.057	−0.058		
	–	–	(0.000)	(0.010)	(0.009)		
UPC-count	–	–	–	−0.009	−0.009		
	–	–	–	(0.000)	(0.000)		
Alt.-specific intercepts:					Alt.-specific coefficients:		
					Income	*Family size*	*intercept*
Bachman	−1.020	−1.430	−1.010	−1.083	−0.017	0.017	−0.990
	(0.004)	(0.000)	(0.003)	(0.003)	(0.896)	(0.956)	(0.334)
Cape Cod	4.152	3.533	3.752	3.934	0.099	0.104	2.910
	(0.000)	(0.000)	(0.000)	(0.000)	(0.310)	(0.647)	(0.000)
Cottage Fries	0.868	0.772	0.751	0.802	−0.013	−0.032	0.949
	(0.005)	(0.012)	(0.015)	(0.009)	(0.905)	(0.903)	(0.270)
Gibbles	−0.245	−0.301	−0.203	−0.231	0.014	−0.064	−0.185
	(0.414)	(0.316)	(0.500)	(0.443)	(0.898)	(0.804)	(0.826)
Herrs	1.316	1.194	1.430	1.504	−0.108	0.326	1.305
	(0.000)	(0.000)	(0.000)	(0.000)	(0.304)	(0.179)	(0.109)
Kettle	2.341	2.225	2.410	2.565	0.071	0.121	1.726
	(0.000)	(0.000)	(0.000)	(0.000)	(0.479)	(0.603)	(0.027)
Lays	5.317	3.468	3.455	4.029	0.0355	0.196	3.244
	(0.000)	(0.000)	(0.000)	(0.000)	(0.715)	(0.387)	(0.000)
Pringles	3.925	3.152	3.138	3.481	0.017	0.271	2.599
	(0.000)	(0.000)	(0.000)	(0.000)	(0.860)	(0.233)	(0.001)
Ruffles	3.800	3.340	3.369	3.605	0.065	0.254	2.417
	(0.000)	(0.000)	(0.000)	(0.000)	(0.503)	(0.266)	(0.001)
State Line	0.694	0.587	0.573	0.594	−0.044	0.255	0.191
	(0.013)	(0.037)	(0.042)	(0.035)	(0.662)	(0.280)	(0.809)
Utz	2.969	2.335	2.765	2.905	0.031	0.166	0.191
	(0.000)	(0.000)	(0.000)	(0.000)	(0.662)	(0.280)	(0.809)
Wachusett	–	–	–	–	–	–	–
	–	–	–	–	–	–	–
Wise	4.013	2.952	3.527	3.749	0.239	0.191	3.063
	(0.000)	(0.000)	(0.000)	(0.000)	(0.806)	(0.401)	(0.000)
Log Likelihood	−36064	−27617	−27604	−27589	−27527		
No. of obs.	243568	243568	243568	243568	243568		
No. of cases	24398	24398	24398	24398	24398		

Notes: Wachusett, the brand with the fewest purchases, is chosen as the base alternative. P-values are in parentheses.

a positive cross-brand effect. It is also worth noting that marginal effects involving the Lays brand have larger magnitudes than others in these two tables. This reflects the fact that Lays is the largest brand in terms of number of observed purchases; its large brand-specific intercept in table 13 demonstrates its popularity. Lays is thus both more susceptible to other brands' improvements and more detrimental to other brands when it improves its own attributes.

The negative marginal effect of a brand subcategory-count on the probability of purchase can be understood as a more "noisy" learning process on that brand's quality for the consumer, holding constant other facts that also affect this learning, such as the number of purchases in this brand made in the past, and the size of the brand in terms of unique UPC's it carries. This reduced form result can be interpreted in the context of the structural model as follows. Learning is more "noisy" when the brand portfolio is spread across more product categories because each category c has a different variance $\sigma^2_{\delta c}$ in the disturbance term ξ_{ijct} to the experienced quality. In general, more unique values of $\sigma^2_{\delta c}$ brings a higher chance of larger variance, and the noisier signals from these categories bring less information on the "true" quality to the brand. Of course, it is possible for a brand to concentrate on a single product category yet have a large disturbance variance in that category, or for a brand to be spread across many categories, all having small (but different) disturbance variance. In this arguably extreme example, learning will still be more efficient in the second brand with smaller disturbance variances. The negative marginal effect from the reduced form results suggest that this situation is rare and does not dominate the overall negative effect from multiple categories mentioned above.

2.5 Discussion

This paper presents a dynamic consumer learning model with special emphasis on learning about a brand's unobserved quality through purchases of this brand's products from multiple categories. It models the consumer learning process explicitly using Bayesian updating mechanism, appealing to the theory of conjugate prior distributions to aid estimation. Its reduced form regression result indicates that a more widely spread brand portfolio, in the sense of the brand carrying products in more snack categories, has a negative effect on purchase probability, controlling for other factors likely to affect the consumer's brand choice. Together with the hypothesized structural model of consumer learning, this result suggests that a more widely spread brand portfolio has a negative effect on consumer learning, compared with a similarly sized brand with a similarly experienced consumer, but a brand portfolio that is more concentrated in fewer product categories. My structural model hypothesizes that a consumer's belief in the quality of the brand converges more slowly to the "true" quality when the brand covers many snack categories because each category has its own variance to the consumer's perception error. Observations with a larger perception error contains less information. The reduced form result suggests that a brand covering more snack categories does have a higher chance of noisier signals overall.

This paper's exclusive focus on Bayesian consumer learning may be insufficient in explaining or predicting firm behavior. It models the consumer's decision on the demand side only, as a response to supply-side factors. The brand portfolios observed in the dataset are assumed to be exogenously given, in both the structural model and the reduced form regressions. Any attempt to control for the potential endogeneity of firms' choice in their portfolios is complicated by the non-linearity of the ASC logit model, where the usual instrumental variable techniques cannot be applied directly. There is also the commonly encountered difficulty in finding a good instrumental variable. A deeper, potential weakness in this model is that it does not endogenize the brands' decision on their portfolio choice. A general equilibrium model that endogenizes some aspect of firm behavior—price, various measures of portfolio choice, or both—will allow the researcher to investigate how a wide range of exogenous factors affect a firm's choice on portfolio components, which is our main variable of interest, either statically or dynamically. It is essential to endogenize portfolio choice if one wants to investigate how a brand's portfolio changes in an equilibrium; exogenous changes in portfolio imposed on the demand system may never be realized in an equilibrium, thus the out-of-sample prediction may be of little practical relevance. An example of a situation where a full (instead of partial) equilibrium study is more appropriate includes an antitrust policy change that limits or promotes multi-market contact between consumer product brands, such as policies on bundling or exclusive dealing.

Table 14: Cross-brand Marginal Effects of Panelist's Purchase History with Brand

	Bachman	Cape Cod	Cottage Fries	Gibbles	Herrs	Kettle	Lays	Pringles	Ruffles	State Line	Utz	Wachusett	Wise
Bachman	$5.69e^{-3}$	$-2.95e^{-4}$	$-2.7e^{-5}$	$-1.8e^{-5}$	$-4.6e^{-5}$	$-8e^{-5}$	$-3.41e^{-3}$	$-4.92e^{-4}$	$-3.19e^{-4}$	$-5e^{-5}$	$-3.7e^{-4}$	$-3.5e^{-5}$	$-5.42e^{-4}$
Cape Cod	$-2.95e^{-4}$	0.178	$-9.08e^{-4}$	$-5.93e^{-4}$	$-1.53e^{-3}$	$-2.64e^{-3}$	−0.113	−0.0163	−0.0105	$-1.66e^{-3}$	−0.0122	$-1.14e^{-3}$	−0.0179
Cottage Fries	$-2.7e^{-5}$	$-9.08e^{-4}$	0.0174	$-5.5e^{-5}$	$-1.42e^{-4}$	$-2.46e^{-4}$	−0.0105	$-1.51e^{-3}$	$-9.82e^{-4}$	$-1.55e^{-4}$	$-1.14e^{-3}$	$-1.06e^{-4}$	$-1.67e^{-3}$
Gibbles	$-1.8e^{-5}$	$-5.93e^{-4}$	$-5.5e^{-5}$	0.0114	$-9.3e^{-5}$	$-1.61e^{-4}$	$-6.85e^{-3}$	$-9.88e^{-4}$	$-6.41e^{-4}$	$-1.01e^{-4}$	$-7.43e^{-4}$	$-6.9e^{-5}$	$-1.09e^{-3}$
Herrs	$-4.6e^{-5}$	$-1.53e^{-3}$	$-1.42e^{-4}$	$-9.3e^{-5}$	0.0292	$-4.14e^{-4}$	−0.0176	$-2.55e^{-3}$	$-1.65e^{-3}$	$-2.6e^{-4}$	$-1.91e^{-3}$	$-1.79e^{-4}$	$-2.81e^{-3}$
Kettle	$-8e^{-5}$	$-2.64e^{-3}$	$-2.46e^{-4}$	$-1.61e^{-4}$	$-4.14e^{-4}$	0.0502	−0.0305	$-4.40e^{-3}$	$-2.86e^{-3}$	$-4.5e^{-4}$	$-3.31e^{-3}$	$-3.09e^{-4}$	$-4.85e^{-3}$
Lays	$-3.41e^{-3}$	−0.113	−0.0105	$-6.85e^{-3}$	−0.0176	−0.03055	0.872	−0.188	−0.122	−0.0192	−0.141	−0.0132	−0.207
Pringles	$-4.92e^{-4}$	−0.0163	$-1.51e^{-3}$	$-9.88e^{-4}$	$-2.55e^{-3}$	$-4.40e^{-3}$	−0.188	0.287	−0.0176	$-2.77e^{-3}$	−0.0204	$-1.905e^{-3}$	−0.0299
Ruffles	$-3.19e^{-4}$	−0.0105	$-9.82e^{-4}$	$-6.41e^{-4}$	$-1.65e^{-3}$	$-2.86e^{-3}$	−0.122	−0.0176	0.192	$-1.80e^{-3}$	−0.0132	$-1.24e^{-3}$	−0.0194
State Line	$-5e^{-5}$	$-1.66e^{-3}$	$-1.55e^{-4}$	$-1.01e^{-4}$	$-2.6e^{-4}$	$-4.5e^{-4}$	−0.0192	$-2.77e^{-3}$	$-1.80e^{-3}$	0.0318	$-2.08e^{-3}$	$-1.95e^{-4}$	$-3.06e^{-3}$
Utz	$-3.7e^{-4}$	−0.0122	$-1.14e^{-3}$	$-7.43e^{-4}$	$-1.91e^{-3}$	$-3.31e^{-3}$	−0.141	−0.0204	−0.0132	$-2.08e^{-3}$	0.220	$-1.43e^{-3}$	−0.0224
Wachusett	$-3.5e^{-5}$	$-1.14e^{-3}$	$-1.06e^{-4}$	$-6.9e^{-5}$	$-1.79e^{-4}$	$-3.09e^{-4}$	−0.0132	$-1.91e^{-3}$	$-1.24e^{-3}$	$-1.95e^{-4}$	$-1.43e^{-3}$	0.0219	$-2.10e^{-3}$
Wise	$-5.42e^{-4}$	−0.0179	$-1.67e^{-3}$	$-1.09e^{-3}$	$-2.81e^{-3}$	$-2.85e^{-3}$	−0.207	−0.0299	−0.0194	$-3.06e^{-3}$	−0.0224	$-2.10e^{-3}$	0.313

Notes: Tabulated from ASC logit regression coefficients from model (4) of table 13. Marginal effects are evaluated at the variables' mean values. All estimates are statistically significant at the 1% level.

Table 15: Cross-brand Marginal Effects of Brand's Subcategory-Count

	Bachman	Cape Cod	Cottage Fries	Gibbles	Herrs	Kettle	Lays	Pringles	Ruffles	State Line	Utz	Wachusett	Wise
Bachman	$-8.9e^{-5}$	$4.6e^{-6}$	$4.3e^{-7}$	$2.8e^{-7}$	$7.2e^{-7}$	$1.3e^{-6}$	$5.3e^{-5}$	$7.7e^{-6}$	$5.0e^{-6}$	$7.9e^{-7}$	$5.8e^{-6}$	$5.4e^{-7}$	$8.5e^{-6}$
Cape Cod	$4.6e^{-6}$	$-2.789e^{-3}$	$1.4e^{-5}$	$9.3e^{-6}$	$2.4e^{-5}$	$4.1e^{-5}$	$1.762e^{-3}$	$2.54e^{-4}$	$1.65e^{-4}$	$2.6e^{-5}$	$1.91e^{-4}$	$1.8e^{-5}$	$2.8e^{-4}$
Cottage Fries	$4.3e^{-7}$	$1.4e^{-5}$	$-2.73e^{-4}$	$8.6e^{-7}$	$2.2e^{-6}$	$3.8e^{-6}$	$1.64e^{-4}$	$2.4e^{-5}$	$1.5e^{-5}$	$2.4e^{-6}$	$1.8e^{-5}$	$1.7e^{-6}$	$2.6e^{-5}$
Gibbles	$2.8e^{-7}$	$9.3e^{-6}$	$8.6e^{-7}$	$-1.78e^{-4}$	$1.5e^{-6}$	$2.5e^{-6}$	$1.07e^{-4}$	$1.5e^{-5}$	$1e^{-5}$	$1.6e^{-6}$	$1.2e^{-5}$	$1.1e^{-6}$	$1.7e^{-5}$
Herrs	$7.2e^{-7}$	$2.4e^{-5}$	$2.2e^{-6}$	$1.5e^{-6}$	$-4.57e^{-4}$	$6.5e^{-6}$	$2.76e^{-4}$	$4e^{-5}$	$2.6e^{-5}$	$4.1e^{-6}$	$3e^{-5}$	$2.8e^{-6}$	$4.4e^{-5}$
Kettle	$1.3e^{-6}$	$4.1e^{-5}$	$3.8e^{-6}$	$2.5e^{-6}$	$6.5e^{-6}$	$-7.85e^{-4}$	$4.77e^{-4}$	$6.9e^{-5}$	$4.5e^{-5}$	$7.0e^{-6}$	$5.2e^{-5}$	$4.8e^{-6}$	$7.6e^{-5}$
Lays	$5.3e^{-5}$	$1.762e^{-3}$	$1.64e^{-4}$	$1.07e^{-4}$	$2.76e^{-4}$	$4.77e^{-4}$	-0.0136	$2.94e^{-3}$	$1.906e^{-3}$	$3e^{-4}$	$2.208e^{-3}$	$2.06e^{-4}$	$3.237e^{-3}$
Pringles	$7.7e^{-6}$	$2.54e^{-4}$	$2.4e^{-5}$	$1.5e^{-5}$	$4e^{-5}$	$6.9e^{-5}$	$2.937e^{-3}$	$-4.481e^{-3}$	$2.75e^{-4}$	$4.3e^{-5}$	$3.19e^{-4}$	$3e^{-5}$	$4.67e^{-4}$
Ruffles	$5e^{-6}$	$1.65e^{-4}$	$1.5e^{-5}$	$1e^{-5}$	$2.6e^{-5}$	$4.5e^{-5}$	$1.906e^{-3}$	$2.75e^{-4}$	$-3.004e^{-3}$	$2.8e^{-5}$	$2.07e^{-4}$	$1.9e^{-5}$	$3.03e^{-4}$
State Line	$7.9e^{-7}$	$2.6e^{-5}$	$2.4e^{-6}$	$1.6e^{-6}$	$4.1e^{-6}$	$7.0e^{-6}$	$3e^{-4}$	$4.3e^{-5}$	$2.8e^{-5}$	$-4.97e^{-4}$	$3.3e^{-5}$	$3.0e^{-6}$	$4.8e^{-5}$
Utz	$5.8e^{-6}$	$1.91e^{-4}$	$1.8e^{-5}$	$1.2e^{-5}$	$3e^{-5}$	$5.2e^{-5}$	$2.208e^{-3}$	$3.19e^{-4}$	$2.07e^{-4}$	$3.3e^{-5}$	$-3.447e^{-3}$	$2.2e^{-5}$	$3.51e^{-4}$
Wachusett	$5.4e^{-7}$	$1.8e^{-5}$	$1.7e^{-6}$	$1.1e^{-6}$	$2.8e^{-6}$	$4.8e^{-6}$	$2.06e^{-4}$	$3e^{-5}$	$1.9e^{-5}$	$3e^{-6}$	$2.2e^{-5}$	$-3.43e^{-4}$	$3.3e^{-5}$
Wise	$8.5e^{-6}$	$2.8e^{-4}$	$2.6e^{-5}$	$1.7e^{-5}$	$4.4e^{-5}$	$7.6e^{-5}$	$3.237e^{-3}$	$4.67e^{-4}$	$3.03e^{-4}$	$4.8e^{-5}$	$3.51e^{-4}$	$3.3e^{-5}$	$-4.89e^{-3}$

Notes: Tabulated from ASC logit regression coefficients from model (4) of table 13. Marginal effects are evaluated at the variables' mean values. All estimates are statistically significant at the 5% level.

References

Ackerberg, D. A. (2003): "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, 44(3), 1007–1040.

Benkard, C. L., A. Bodoh-Creed, and J. Lazarev (2009): "Simulating the Dynamic Effects of Horizontal Mergers: US Airlines," Discussion paper.

Berry, S., M. Carnall, and P. Spiller (2006): "Airline Hubs: Costs, Markups, and the Implications of Costumer Heterogeneity," *Advances in Airline Economics, Competition Policy and Antitrust*, 1.

Berry, S., and P. Jia (2008): "Tracing the Woes: An Empirical Analysis of the Airline Industry," .

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Bulow, J. I., J. D. Geanakoplos, and P. D. Klemperer (1985): "Multimarket Oligopoly: Strategic Substitutes and Complements," *The Journal of Political Economy*, 93(3), 488–511.

Competition Commission and Office of Fair Trading (2010): "Merger Assessment Guidelines," `http://www.competition-commission.org.uk/about_us/our_organisation/workstreams/analysis/pdf/100916_merger_assessment_guidelines.pdf`.

Dacin, P. A., and D. C. Smith (1994): "The Effect of Brand Portfolio Characteristics on Consumer Evaluations of Brand Extensions," *Journal of Marketing Research*, pp. 229–242.

DeGroot, M. H. (1970): *Optimal Statistical Decisions*. New York: McGraw-Hill.

Department of Justice and Federal Trade Commission (2010a): "Hart-Scott-Rodino Annual Report," `http://www.ftc.gov/os/2011/02/1101hsrreport.pdf`.

Department of Justice and Federal Trade Commission (2010b): "Horizontal Merger Guidelines," `http://www.justice.gov/atr/public/guidelines/hmg-2010.html`.

Erdem, T., S. Imai, and M. Keane (2003): "Brand and Quantity Choice Dynamics Under Price Uncertainty," *Quantitative Marketing and Economics*, 1(1), 5–64.

Erdem, T., and M. P. Keane (1996): "Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15(1), 1–20.

Farrell, J., and C. Shapiro (1990): "Horizontal Mergers: An Equilibrium Analysis," *The American Economic Review*, 80(1), 107–126.

——— (2010): "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition," *The BE Journal of Theoretical Economics*, 10(1), 9.

Guadagni, P. M., and J. D. C. Little (1983): "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, pp. 203–238.

Hausman, J., G. Leonard, and J. D. Zona (1994): "Competitive Analysis with Differenciated Products," *Annales d'Economie et de Statistique*, pp. 159–180.

Hausman, J., S. Moresi, and M. Rainey (2010): "Unilateral Effects of Mergers with General Linear Demand," *Economics Letters*.

Hendel, I., and A. Nevo (2006a): "Measuring the Implications of Sales and Consumer Stockpiling," *Econometrica*, 74(6), 1637–73.

——— (2006b): "Sales and Consumer Inventory," *The RAND Journal of Economics*, 37(3), 543–561.

Jaffe, S., and E. G. Weyl (2010): "The First-Order Approach to Merger Analysis," *working paper*.

McDonald, J. B. (2005): "Antitrust for Airlines," `http://www.justice.gov/atr/public/speeches/217987.htm`.

McFadden, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontier in Economics, ed. P. Zarembka*, pp. 105–142, New York: Academic Press.

Meyvis, T., and C. Janiszewski (2004): "When Are Broader Brands Stronger Brands? An Accessibility Perspective on the Success of Brand Extensions," *Journal of Consumer Research*, 31(2), 346–357.

Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics and Management Strategy*, 9(4), 513–548.

Pakes, A. (2010): "Upward Pricing Pressure Screens in the New Merger Guidelines: Some Pro's and Con's," `http://www.economics.harvard.edu/faculty/pakes/files/SDGcomp.pdf`.

Ryan, S. P., and C. Tucker (2011): "Heterogeneity and the Dynamics of Technology Adoption," *NBER working paper*.

Schmalensee, R. (2009): "Should New Merger Guidelines Give UPP Market Definition?," *CPI Antitrust Chronicle*.

Shapiro, C. (1996): "Mergers with Differentiated Products," *Antitrust*, 10(2), 23–30.

——— (2010): "Unilateral Effects Calculations," Discussion paper, Working Paper, University of California, Berkeley.

Simons, J. J., and M. B. Coate (2010): "Upward Pressure on Price Analysis: Issues and Implications for Merger Policy," *European Competition Journal*, 6(2), 377–396.

Varma, G. D. (2009): "Will Use of the Upward Pricing Pressure Test Lead to an Increase in the Level of Merger Enforcement?," *Antitrust*, 24(1), 27–32.

Walters, C. (2007): "Diversions Ahead! Approximating Diversion Ratios for Retail Chain Mergers," `http://www.competition-commission.org.uk/our_role/analysis/diversion_ratios.pdf`.

Werden, G. J. (1996): "A Robust Test for Consumer Welfare Enhancing Mergers Among Sellers of Differentiated Products," *The Journal of Industrial Economics*, 44(4), 409–413.

Weyl, E. G., and M. Fabinger (2009): "Pass-through as an Economic Tool," http://home.uchicago.edu/weyl/Pass-through_10_09.pdf.

Yang, B., and A. Ching (2009): "Dynamics of Consumer Adoption of Financial Innovation: The Case of ATM Cards," *working paper*.

Appendix A Data and variable construction

The DB1B (Airline Origin and Destination Survey) dataset is publicly available from BTS RITA (Bureau of Transportation Statistics, Research and Innovative Technology Administration). It is a ten percent random sample of airline tickets from reporting airlines, collected by the Office of Airline Information. The dataset is in quarterly frequency and is available on three levels of aggregation: coupon, market, and ticket. For each itinerary, I obtain the itinerary fare from the ticket dataset, the destination from the market dataset, and the fare class information from the coupon dataset.

I use itineraries of eight pre-merger quarters, from 2003Q3 to 2005Q2, for my demand estimation. I select itineraries by the following criteria: I keep itineraries between a list of 75 major airports as defined by Benkard, Bodoh-Creed, and Lazarev (2009) in their Appendix A. I keep roundtrip tickets that involve at most one change of flight between its origin and destination airports. I drop itineraries whose dollar credibility indicator (provided by the ticket dataset) is negative. I drop itineraries with fare equal to or less than US$10 or with number of passengers larger than 100. I drop itineraries that involves *any* leg that is not in the economy class, where the fare class is defined by the airline.[12] I then aggregate the dataset of itineraries into the airline–origin airport–destination airport–quarter level. Finally, I drop any airline–origin airport–destination airport–quarter tuple with less than ten total passengers.

A smaller subset of the sample is used to estimate the BCS random coefficient model due to computational time. This subset consists of observations from a 50% sample of markets where either America West or US Airways operate in the 2005Q2 quarter. A comparison of the simple nested logit estimates generated from the complete dataset and this subset in table 16 indicates that this subset is a representative sample.

[12]Southwest Airlines, JetBlue Airways, and Sun Country Airlines code all their itineraries as "First Class" in the coupon dataset. I override their definition and assume all their tickets to be economy.

Table 16: Simple nested logit demand estimates

	Complete dataset	BCS subset
Itinerary fare[a]	-0.00498	-0.00499
	$(1.021e^{-4})$	$(7.452e^{-4})$
Direct flight	0.0150	0.0145
	$(8.18e^{-5})$	$(3.98e^{-4})$
Distance	$1.938e^{-4}$	$1.24e^{-4}$
	$(4.25e^{-6})$	$(2.46e^{-5})$
Psngr. share at origin airport	0.0119	0.0118
	$(1.783e^{-4})$	$(9.697e^{-4})$
Psngr. share at destination airport	0.00463	0.00519
	$(1.675e^{-4})$	$(8.264e^{-4})$
constant	-9.017	-8.573
	(0.0257)	(0.169)
Nested logit parameter (λ)	0.493	0.534
	(0.00213)	(0.00964)
First stage R^2	0.3160	0.3258
N	124852	5207
Average own-price elasticity	-3.108	-2.767
Average cross-price elasticity	0.384	0.180

Note: Standard errors are reported in parentheses. The dependent variable is $ln(s_{jt}) - ln(s_0)$. All variables are statistically significant at the 1% level.
[a] A Hausman instrument of prices is used to control for the endogenous itinerary fare by two stage least squares (2SLS). This instrument is defined by the simple average of itinerary fares, in 2003 dollars per mile, over all *other* observations in the same market. The unit of dollar *per mile* is used when constructing the Hausman instrument to allow for fair comparison between itinerary fares of different route lengths. Appendix B shows the first stage regression results of the endogenous variable on the exogenous and instrumental variables.

Appendix B First stage regression results

A Hausman instrument of prices is used to control for the endogenous itinerary fare. This instrument is defined by the simple average of itinerary fares, in 2003 dollars per mile, over all *other* observations in the same market. The unit of dollar *per mile* is used when constructing the Hausman instrument to allow for fair comparison between itinerary fares of different route lengths. I show in table 17 that the Hausman instrument has explanatory power on itinerary fare by the first stage regression in 2SLS, which regresses the endogenous variable (itinerary fare) on the set of exogenous and instrumental variables. All coefficients are statistically significant.

Table 17: First stage regressions of demand estimates

	Complete dataset	BCS subset[a]
Hausman instrument on price	280.024	185.93
	(3.038)	(13.098)
Direct flight	0.352	0.293
	(0.00707)	(0.0293)
Distance	0.0431	0.0351
	$(1.667e^{-4})$	$(7.538e^{-4})$
Psngr. share at origin airport	0.499	0.596
	(0.0181)	(0.0786)
Psngr. share at destination airport	0.513	0.570
	(0.0185)	(0.0777)
constant	169.794	176.643
	(1.184)	(4.984)
$\ln(s_{j\|N})$[b]	13.378	9.196
	(0.180)	(0.705)
First stage R^2	0.378	0.326
N	124852	5207

Note: Standard errors are reported in parentheses.

[a] A smaller subset of the sample is used to estimate the BCS random coefficient model due to computational time. This subset consists of observations from a 50% sample of markets where either America West or US Airways operate in the 2005Q2 quarter. A comparison of the simple nested logit estimates generated from the complete dataset and this subset indicates that this subset is a representative sample.

[b] The log of market share, conditional on being in the nest (N) of "inside" goods, is used to recover the nested logit parameter λ in the 2SLS nested logit regressions.

Appendix C Demand Estimation and Identification

My estimation procedure follows that of Nevo (2000) closely. One simplification is that there is no need for numeric integration over my random coefficients, because their discrete bimodal distribution yields an exact functional form. As a second modification, I follow Berry and Jia (2008) in using a contraction map over ξ_{jt} instead of δ_{jtr}, since the latter would involve one more variable γ, thus complicating the minimization routine. This is because ξ_{jt} is common across types, while δ_{jtr} is not. Once the demand unobservables ξ_{jt} are recovered, they are interacted with the exogenous and instrumental variables to form the GMM objective function. The parameters $(\alpha_r, \beta_r, \gamma, \lambda)_{r=1,2}$ are then estimated by a Nelder-Mead simplex search routine.

To address the issue of price endogeneity, I compute a Hausman instrument for prices, which is defined as the passenger-weighted average fare over competing products in each market. The average fare is computed in dollar-per-mile for a fair comparison between itinerary fares from routes of different lengths. I treat all other airline-route characteristics as exogenously given. Thus, my set of instrumental variables thus consists of the Hausman instrument on price and all exogenous product characteristics. The exogeneity assumption is admittedly a crude for the airport presence variables, since airport presence is computed from the total number of passengers on routes involving that airport, and the number of passengers on each of these routes is given by the endogenous market shares. One justification for this exogeneity assumption on airport presence is that the (endogenous) market shares of *other* routes involving the airport of interest is arguably unaffected by the endogenous prices and market shares of the *current* route of interest. Lastly, I have chosen not to use an instrument in the style of Berry, Levinsohn, and Pakes (1995), which is defined as the average fare (or other product characteristics) over all other products produced by the same firm in other markets. The BLP instrument is more appropriate in a case where markets are defined by geographic boundaries, and where a firm operates in most (if not all) of these geographic boundaries, such as the markets for ready-to-eat cereals. Since the total number of routes is large and an airline's presence among these routes is extremely varied, a BLP instrument computed for the airline industry will critically depend on each airline's portfolio of routes, each of which has its own unique competitive characteristics. This variable is unlikely to be a good indication of the cost of the particular route of interest.

Here I provide an intuitive argument on the identification of the random coefficient model, which includes the nested logit parameter λ, the type-specific taste coefficients $(\alpha_r, \beta_r)_{r=1,2}$, and the mixture variable γ that gives the population percentage of type I consumers.. Firstly, λ is identified from the variation in the aggregate share of all "inside" goods as the number of products changes between markets. λ is bounded between 0 and 1. In the first extreme where $\lambda = 0$, the aggregate share of all inside goods $\frac{D_t^\lambda}{1+D_t^\lambda}$ reduces to a constant even when the number of inside goods varies. In the second extreme where $\lambda = 1$, the nested logit model simplifies to the simple logit model, where market share is given by the ratio $\exp(\delta_j)$ and the inclusive value $(1+\sum_k \exp(\delta_k))$, and the inclusive value certainly depends on the total number of available products. In a simple case where all products have the same $\delta_j = \delta$, the aggregate market share of all inside goods will be $\frac{J\exp(\delta)}{1+J\exp(\delta)}$, where J is the total number of inside goods. Secondly, the type-specific taste coefficients are identified from the substitution patterns as the portfolio of products varies across markets. The overall cross-price elasticity between two products is given by a convex combination, weighted by γ and $(1-\gamma)$, of the cross-price elasticity of each customer type, which in turn are functions of (α_r, β_r) for each type r. The parameters $(\alpha_r, \beta_r)_{r=1,2}$ and γ are then chosen such that the convex combination of elasticities is closest to the substitution pattern observed in the data.

Appendix D Recovered unobserved product characteristic

The unobserved product characteristic (ξ_j) is recovered for each product during the estimation of the BCS random coefficient demand model using a contraction map. It is then interacted with instrumental variables to form moment conditions for the estimation of BCS demand parameters. Table 18 summarizes the recovered product characteristic ξ_j for major airlines. It accounts for many unobserved attributes, such as ticket restrictions, departure and arrival times, and flight frequencies. A value of ξ_j is recovered for each airline–route–quarter observation. The table includes both a simple mean across all routes and a passenger-weighted mean for comparison. While an airline may have stronger incentive to improve its "quality" on its major routes where most of its customers travel, only a weak correlation of 0.328 is observed in this sample between ξ_j and number of passengers. The positive correlation between the two nonetheless explains the mostly positive values among the weighted averages. United, Continental, and America West rank the highest in both columns, while Southwest ranks the lowest in both. Among our two merging airlines, America West has a higher average than US Airways in both columns. To put the magnitudes of ξ_j in perspective, according to the price coefficients for the BCS model estimated above, an increase in ξ_j of magnitude 1 is equivalent to a decrease in price of \$129.70 for the type 1 consumer

Table 18: AVERAGE UNOBSERVED PRODUCT CHARACTERISTIC (ξ_j) FOR MAJOR AIRLINES FROM THE ESTIMATED BCS MODEL

Airline	Mean	Psngr-weighted mean	Min	Max	N
Delta (DL)	0.0484	0.615	−2.434	2.374	651
American (AA)	−0.909	0.575	−2.540	2.069	570
Northwest (NW)	−0.0502	0.639	−3.535	1.902	538
US Airways (US)	−0.112	0.312	−6.952	2.597	527
United (UA)	0.410	1.026	−1.840	2.607	519
Southwest (WN)	−0.446	−0.178	−2.699	1.589	518
America West (HP)	0.279	0.770	−2.078	2.742	478
Continental (CO)	0.204	1.634	−1.636	3.566	299
AirTran (FL)	−0.0364	0.257	−1.770	2.182	209
Frontier (F9)	0.0491	0.746	−1.774	2.166	196
Atlantic Coast (DH)	−0.000607	−0.270	−1.982	2.274	129
All	−0.00674	0.549	−6.952	3.566	5207

Note: The unobserved product characteristic ξ_j is allowed to vary across airlines and routes, but is held constant at merger simulation.

and \$854.70 for the type 2 consumer. Thus, the average magnitudes in table 18 are within reasonable range.

I also show in table 19 with two simple OLS regressions that the recovered ξ_j variable is indeed independent to all the instrumental variables used, but is predictably, positively correlated with itinerary fare. This can be seen from the regression results that none of the variables of regression (1) is statistically significant, while that of regression (2) is. Lastly, I show in figure 4 that the distribution of ξ_j is visibly very close to normal, with a mean of -0.00674 and a standard deviation of 0.834. It also has a roughly bell-shaped distribution over observations for each airline.

Table 19: Regressions of the recovered unobserved product characteristic (ξ_j)

	(1)	(2)
Hausman instrument on price	0.161	–
	(0.161)	–
Direct flight	$-06.21e^{-5}$	–
	$(3.446e^{-4})$	–
Distance	$4.79e^{-6}$	–
	$(9.28e^{-6})$	–
Passenger-share at origin airport	$5.273e^{-4}$	–
	$(9.383e^{-4})$	–
Passenger-share at destination airport	$2.459e^{-4}$	–
	$(9.471e^{-4})$	–
Itinerary fare	–	0.00433
	–	$(1.271e^{-4})$
constant	-0.0578	-1.377
	(0.0541)	(0.0415)
N	5207	5207
R^2	0.0003	0.183

Note: Standard errors are reported in parentheses.

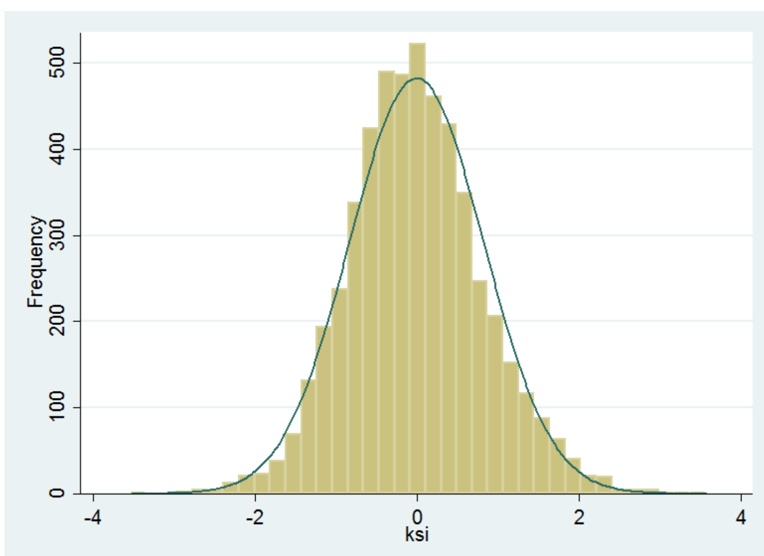

Figure 4: Histogram showing the distribution of the recovered unobserved product characteristic (ξ_j) with a best-fitted normal density.