On the Importance of Modelling Dynamic Demand for Competition Analysis: The case of Caffeinated Beverages

Philip G. Gayle* and Jingwen Liao+

This draft: September 22, 2020

Abstract

This paper illustrates that, compared to static discrete choice demand models, a dynamic discrete choice demand model can better capture "complementary" type consumer choice behavior among pairs of differentiated products. Measuring the competitive impacts of mergers crucially depend on both the type and strength of the relationship between products of rival firms, where sufficiently strong complementarity between products of the merging firms can result in lower price-cost markups post-merger, an unattainable outcome when relevant products are substitutes. Accordingly, hypothetical merger simulations between leading caffeinated beverage firms selling several complementary products predict lower price-cost markups on many products post-merger.

Keywords: Dynamic Demand; Demand Complements; Soda and Coffee
JEL Classification Codes: D12; L66; L13; L41

* Kansas State University, Department of Economics, 322 Waters Hall, Manhattan, KS 66506; Voice: 785-532-4581; Fax: 785-532-6919; Email: gaylep@ksu.edu; **Corresponding author**.

+ Kansas State University, Department of Economics, 327 Waters Hall, Manhattan, KS 66506; Email: jwliao@ksu.edu.

1. Introduction

The contents of many consumer goods are packaged in discrete sizes. As such, in responding to relative price changes among these products, consumers are restricted to adjust their purchase volume of the product contents based on the discrete package sizes available. Since differentiated products within a given package size category may inherently be perceived as substitutes, over time, some consumers may switch their purchases back and forth between roughly similarly priced products of the same package size.

Relative prices among differentiated products within a package size category may be different when compared to other package size categories. For example, in the case of carbonated soda products, a relatively popular package size is a 6-pack of 16.9 fl. oz. bottles, yielding total soda content of 101.4 fl. oz. A check on soda product prices in June 2020 at a Walmart retail Supercenter reveals the following for this package size: (*i*) $2.50 for Dr Pepper; (*ii*) $2.50 for Pepsi; and (*iii*) $3.33 for Coca-Cola. An alternate package size offered is 67.7 fl. oz. single bottle products with the following prices: (*i*) $1.58 for Dr Pepper; (*ii*) $1.68 for Pepsi; and (*iii*) $1.25 for Coca-Cola.

A change in the price of a given product, causing a change in relative prices among products within that package size category, may induce some consumers to switch package size categories. The demand for many of the products within the size category that experienced the relative price change may change in the same direction, i.e., a "complementary" type consumer choice behavior, when some consumers switch package size category due to the relative price change. We now provide an example.

Consider again the two package sizes of soda products mentioned above: (*i*) a package with total soda content of 101.4 fl. oz.; and (*ii*) a package with total soda content of 67.7 fl. oz. Within each package size, there are several distinct soda products distinguished by various non-price characteristics. Prior to any price change, some consumers may switch their purchases back and forth between roughly similarly priced products within the 101.4 fl. oz. package size. Now suppose the price of one product within the 101.4 fl. oz. package size increases. This price increase likely causes consumers to have lower expected current and *future* utility from the set of 101.4 fl. oz. packaged products, and the decrease in expected utility for 101.4 fl. oz. packaged products may be sufficient to induce some consumers to either choose among soda products of a smaller package size, perhaps the 67.7 fl. oz. package size, or not choose any soda product. In this case, there is a decline in demand for 101.4 fl. oz. packaged soda products associated with the increase in the price of a single soda product of this package size, yielding "complementary" type consumer choice behavior among pairs of 101.4 fl. oz. packaged soda products.

A key objective of this paper is to illustrate that while both static and dynamic discrete choice models can capture consumers' incentives for "complementary" type choice behavior, a static discrete choice model imposes a restriction on the size of such "complementary" type incentives that causes the inherent substitutability of products within a package size category to always dominate. Furthermore, we show that unlike a static discrete choice model, the forward-looking attribute of consumers in the dynamic model increases the capacity of the discrete choice model to capture "complementary" type consumer incentives such that these incentives can at times be relatively stronger than the inherent substitutability of products within a package size category, ultimately yielding "complementary" type relationships between these products. The intuition is that, unlike assumed decision-making in a static discrete choice demand model, in a dynamic discrete choice model forward-looking consumers in making their optimal purchase decision take into account how a given price change impacts *future* expected utility the consumer obtains from the relevant size category of products, which can serve as an extra stimulus for consumers to change their current purchase decision across size categories. This increased willingness of forward-looking consumers to switch their current purchase decision across size categories of products in response to price change of a given product is a key positive driver of the strength of the "complementary" type consumer incentives that can ultimately yield "complementary" type relationships between products within a given size category. It then becomes an empirical question of whether a given product pair is ultimately treated as complements or substitutes based on consumers' patterns of choice behavior.

One appeal of using discrete choice models when modeling consumer demand for differentiated products is that these models enable researchers to flexibly model taste heterogeneity across consumers, and thereby enable researchers to use such taste heterogeneity to better understand differences in consumer choice behavior and welfare impacts of various shocks and supply-side market changes [Berry (1994); Berry, Levinsohn and Pakes (1995); Crawford and

Yurukoglu (2012); Ho and Lee (2017); Crawford, Lee, Whinston and Yurukoglu (2018)]. Since traditional *static* discrete choice models of demand [1] impose the restriction that consumers' incentive for "complementary" type choice behavior are never sufficiently strong to overturn the perceived substitutability between product pairs, then within the framework of *static* discrete choice demand models any pair of distinct products are necessarily empirical substitutes. As such, a key methodological contribution of our paper is to illustrate how a *dynamic* discrete choice demand model overcomes this shortcoming of traditional static discrete choice models of demand.

Why is it important to capture whether a given product pair is ultimately treated by consumers as complements versus substitutes? Analysis of market competition, such as the competitive impacts of mergers, crucially depends on both the type and strength of the relationship between products of rival firms. In particular, the merger of firms that sell substitute products necessarily result in higher price-cost markups on these products post-merger, an anti-competitive outcome. Furthermore, the increases in price-cost markups are expected to be greater the stronger the substitutability between products of the merging firms. Conversely, a merger between firms that sell complementary products can result in lower price-cost markups on these products post-merger, a pro-competitive outcome. Furthermore, the decreases in price-cost markups are expected to be greater the stronger the complementarity between products of the merging firms. Therefore, accurate inference on the competitive effects of mergers depends crucially on the type and strength of the relationship between products of the merging firms.

Empirically, we focus on two categories of caffeinated beverages sold in supermarkets: coffee and soda. Both are popular storable caffeinated beverages. The storability characteristic of these products facilitates consumers holding inventories of them. Purchase quantities not consumed in the current period are stored at a cost for future consumption. Holding inventories also facilitates consumers being able to inter-temporally smooth consumption, and optimally choose lumpy purchases in response to temporary versus permanent price changes. As discussed in Hendel and Nevo (2006b), such dynamic considerations of consumers of storable goods provide a reason to favor using a dynamic demand model over a static demand model. A static demand model ignores consumers' inventory and forward-looking behaviors. Caffeinated beverages sold

[1] Popularly used static models of demand include, the standard conditional logit, random coefficients logit, and nested logit.

in supermarkets do satisfy this reason for using a dynamic demand model. However, our paper posits another reason for wanting to use a dynamic demand model, which is to capture potential "complementary" type consumer choice behavior.

For decades, coffee and soda producers have sought to forge cooperative partnerships. In 1994, the well-known soda giant, Pepsi Co, and the famous coffee chain, Starbucks Corp, jointly found the North American Coffee Partnership to produce ready-to-drink coffee (RTD). This partnership markets products produced by Starbucks Corp, and utilizes the distribution system of Pepsi Co. More recently, at the end of 2018, Keurig Green Mountain, a coffee producer, indeed merged with Dr Pepper Snapple Group, a soda company. The merged firm owns 31 coffee brands and 20 soda brands.

A merger between leading firms in each caffeinated beverage category can incentivize the merged firm to exercise its market power, which may substantially increase price-cost margins of the merged firm's products. Whether the mergers would be of antitrust concern greatly depends on how consumers perceive the relationship between products across caffeinated beverage categories. Informed by the discussion above, it is prudent to use a dynamic demand model, as oppose to a traditional static model of demand, to more accurately estimate the relationship between products of the merging firms, which in turn facilitates more accurate assessment of the potential merger effects.

To achieve our objectives, we adopt a dynamic model of demand proposed by Hendel and Nevo (2006b). They show that the probability of choosing a given branded product conditional on package size (quantity of product content) is independent of dynamic considerations. Therefore, they provide a tractable three-step estimation procedure which enables us to split the dynamic decision into a static brand choice conditional on content quantity, and a dynamic content quantity choice.

We estimate the dynamic demand for caffeinated beverages using scanner data from the Information Resources Inc. (IRI) marketing dataset [Bronnenberg et al. (2008)]. The data contain two sets of information. First, we observe the weekly purchasing records of a group of coffee and soda consumers over a period of two years. These data include households' demographic information as well as the price and product content quantity they purchase. In addition, the dataset includes a panel of product attributes and retail store promotional activities.

Given the structural dynamic demand parameter estimates, we first derive the price elasticities of demand. For comparison, we also estimate a traditional static conditional logit model of demand, and use the parameter estimates from this static model to compute price elasticities of demand. The results reveal that ignoring the dynamics in consumer demand can yield inaccurate estimates of the relationship between products in the same category. For example, in the case of 144 fl. oz. packaged ground coffee products, our dynamic model predicts the cross-price elasticity of Morning Joe, a Starbucks branded coffee product, with respect to branded products of The JM Smuckers is a mean -0.1, suggesting complementary type relationships. However, a pure static demand model that assumes products are substitutes by default predicts cross-price elasticities among the 144 fl. oz. packaged ground coffee products listed above to be a mean 0.01. Similarly, the dynamic model yields cross-price elasticities suggesting complementary type relationships between some pairs of soda products within a given package size, while the static model yields cross-price elasticities suggesting substitute relationships between the same pairs of soda products.

Next, we simulate two types of mergers. First, we assume two firms from the same caffeinated beverage category merge. Comparing the estimates predicted from our dynamic model to a pure static model, we show that some price-cost margins generated from the dynamic demand model are predicted to decline due to empirical complementary type relationships across products within a given package size. For example, in case of the counterfactual merger between Starbucks and The JM Smucker, two major coffee firms, the dynamic model predicts that consumers would observe price-cost markups on Starbucks products' to decrease by 8.41% on average. In contrast, the static model predicts price-cost markups on these products will increase by a mean 1.68%.

We then perform a counterfactual merger simulation between leading firms in each caffeinated beverage category, i.e. cross-category merger. Even though in this data set we do not observe empirical complementarity across coffee-soda product pairs, the merger effects predicted by the two models, dynamic versus static, differ substantially. Our simulated merger between Keurig and Dr. Pepper yield a predicted mean percentage increase in price-cost markups from the static demand model that is approximately 3 times as high as the predicted mean percentage increase in markups from the dynamic demand model.

The paper proceeds as follows. In the next section, we briefly describe related literature. In section 3 we present the demand and supply model of caffeinated beverages, discuss estimation, and formally illustrate how the dynamic demand model captures complementary type relationships across products. We describe the data in section 4. In section 5 we report and discuss parameter estimates of the static and dynamic portions of the dynamic demand model. In section 6 we report and discuss demand elasticities and hypothetical merger effects. Concluding remarks are gathered in section 7.

2. Related Literature

Our paper is most closely related to the literature on empirical dynamic demand models of consumer goods. This literature can be decomposed into two major branches: (*i*) dynamic demand models of durable goods; and (*ii*) dynamic demand models of storable goods. Recent contributions to the durable goods branch of the literature include, Melnikov (2013) who proposes a dynamic demand of durable goods that are not repurchased, and applies it to a market for computer printers. Gowrisankaran and Rysman (2012) applies the framework to the digital camcorders industry and allow consumers to repurchase. Those models have been extended and applied to answer many other research questions. For example, Lee (2013) studies vertical integration in a two-sided market assuming consumers are forward-looking, and more recently Huang (2019) estimates consumer dynamic demand involving human capital accumulation.

Our study contributes most to the branch of the literature on dynamic demand for storable goods. Contributions to this literature include Erdem, Imai and Keane (2003), who construct a dynamic model of demand that focuses on the role of price expectation on storable goods. Hendel and Nevo (2006a) finds evidence of household stockpiling behavior and suggests that a static demand model may inaccurately estimate consumers' price sensitiveness. Hendel and Nevo (2006b) proposed a dynamic discrete choice framework considering consumer inventory behavior, as well as an estimation strategy that reduces the computational burden of the dynamic demand model. Hartmann and Nair (2010) extends Hendel and Nevo (2006b) to estimate the demand for tied products. They endogenize product choice in terms of which tied products consumers have at home. Hendel and Nevo (2013) and Dubios and Magnac (2015) further incorporate within the demand for storable goods framework, firms' strategic price-setting behavior. They derive equilibrium price-setting and identify the outcomes and effects of intertemporal price discrimination. Additionally, some studies such as Wang (2015) and Osborne (2018) show the importance of considering a dynamic demand setting when measuring welfare changes.

Our study also fits into the literature on measuring the competitive effects of a horizontal merger. The majority of this literature adopts a static discrete choice model of demand. For example, Nevo (2000) uses the methodology of estimating a static structural model of demand to evaluate mergers in the ready-to-eat cereal industry. Pinkse and Slade (2004) studies the market effects resulting from mergers between brewers of beer in the UK. Other studies that use and discuss a static demand framework for measuring the market effects of horizontal mergers include, Ivaldi and Verboven (2005) and Bjornerstedt and Verboven (2016).

Our study also contributes to the extensive literature on the coffee industry, and perhaps even more extensive literature on the caffeinated soda industry. For example, Dub é (2004) develops a multiple discreteness model to estimate the demand for soda, and Dubé (2005) applies the multiple-unit purchase demand model to analyze merger cases in the soda industry, and draws antitrust policy implications. Lopez and Fanuzzi (2012) estimates a random coefficient logit model of demand for soda to simulate the effect of caloric taxes. Wang (2015) adopts a dynamic discrete choice model to estimate the soda tax effect on consumer welfare. More recently, Gayle and Indika (2020) investigate the impact on soda prices of vertical integrations in which PepsiCo and Coca-Cola each acquired their major bottlers. There are numerous studies that estimate the demand for coffee, including McManus (2007), Bonnet and Villas-Boas (2016), Villas-Boas (2007), and Gayle and Lin (2020). However, to the best of our knowledge, no study analyzed a model of demand that jointly incorporates both coffee and soda caffeinated beverages, which is necessary to study existing and potential partnerships between firms across these two product categories. As such, our study fills this gap in the literature.

3. The Model

In this section, we build a dynamic demand model based on the framework of Hendel and Nevo (2006b). A key modification of the Hendel and Nevo (2006b) framework is that we allow households to purchase, consume, and keep an inventory level for each of the two caffeinated beverage categories, while the framework in Hendel and Nevo (2006b) focus on consumer choice behavior over a single product category. In addition, we use a utility function specification that permits estimation of parameters that describe the aggregate relationship between the two caffeinated beverage categories. Finally, assuming firms compete in prices, Bertrand-Nash fashion, we recover product-level price-cost markups.

3.1 Dynamic Demand

Household h obtains a per period utility from consuming coffee, soda, and an outside option:

$$u(c_{ht}^s, c_{ht}^c, \nu_{ht}^s, \nu_{ht}^c; \gamma) + \gamma_0 O_{ht},$$

where c_{ht}^s is the amount of soda consumed by household h at time t; c_{ht}^c is the consumption of coffee; ν_{ht}^s and ν_{ht}^c are random shocks to the utility that changes the marginal utility from consumption of each beverage category; γ is a set of taste parameters in the consumption utility function; O_{ht} is the consumption of the outside option; and γ_0 the marginal utility from consuming the outside option.

There are J products in the market, where J_1 of them are soda products, and the rest $J_2 = J - J_1$ of them are coffee products. Each product is defined as a brand-package size combination. The total consumption of each category (coffee or soda) by household h in period t is $c_{ht}^s = \sum_{j \in J_1} c_{jht}^s$ and $c_{ht}^c = \sum_{j \in J_2} c_{jht}^c$. We adopt the following functional form of utility from consuming the differentiated products:

$$\begin{aligned} u(c_{ht}^s, c_{ht}^c, \nu_{ht}^s, \nu_{ht}^c) = \; & \gamma_1 (c_{ht}^s + \nu_{ht}^s) + \gamma_2 (c_{ht}^s + \nu_{ht}^s)^2 \\ & + \gamma_3 (c_{ht}^c + \nu_{ht}^c) + \gamma_4 (c_{ht}^c + \nu_{ht}^c)^2 \\ & + \gamma_5 (c_{ht}^s + \nu_{ht}^s)(c_{ht}^c + \nu_{ht}^c), \end{aligned} \tag{1}$$

where γ_1, γ_2, γ_3, γ_4 and γ_5 are parameters to be estimated. The chosen functional form of utility implies the following set of linear inverse demand functions (Martin, 2009):

$$p^s = \gamma_1 + 2\gamma_2 \, c^s + \gamma_5 c^c + \tilde{\upsilon}^s, \tag{2}$$

$$p^c = \gamma_3 + 2\gamma_4 \, c^c + \gamma_5 c^s + \tilde{\upsilon}^c, \tag{3}$$

where $\tilde{\upsilon}^s$ and $\tilde{\upsilon}^c$ are composite functions of ν_{ht}^s and ν_{ht}^c; and γ_5 is the parameter that captures the aggregate relationship between the two caffeinated beverage categories. If $\gamma_5 = 0$, the beverage categories are independent in demand; if $\gamma_5 > 0$, then coffee and soda categories are substitutes; while if $\gamma_5 < 0$, the two categories are demand complements. In addition, $\gamma_1 > 0, \gamma_3 > 0, \; \gamma_2 <$

0, and $\gamma_4 < 0$ imply diminishing marginal utilities from consumption levels and downward sloping demand curves.

Let x index package sizes, and therefore quantity content of products. As such, $x = 0, 1, 2, \dots, X$, where $x = 0$ represents a quantity content of zero, which corresponds to choosing the outside option, while X is the total number of distinct package sizes across the two beverage categories. In each period, household h enjoys indirect utility, g_{hjxt}, from purchasing one unit of product j of pacackge size x, i.e:

$$g_{hjxt} = \alpha_1 p_{jxt} + \alpha_2 A_{jxt} + \sum_l^L \alpha_{3l}\, p_{jxt} Demo_{hl} + \xi_{jxt} + \varepsilon_{hjxt}, \qquad (4)$$

where p_{jxt} is the price of product j of package size x; A_{jxt} represents promotional activities to entice product purchase; $Demo_{hl}$ is household-specific demographic variable l, which allows the model to capture heterogeneity in price sensitivities across households via the set of interaction variables $(p_{jxt}Demo_{h1}, p_{jxt}Demo_{h2}, \dots, p_{jxt}Demo_{hL})$; ξ_{jxt} is a composite of product-specific attributes observed by consumers and firms, but not by us the researchers; and ε_{hjxt} is a mean zero random preference shock. Given the standard law of demand, we expect: $\alpha_1 < 0$, and $|\alpha_1| > \left|\sum_l^L \alpha_{3l}\, Demo_{hl}\right|$ for all h.

Because the products are storable, quantity purchased does not necessarily equal to the quantity consumed, the difference is stored as inventory. Following Hendel and Nevo (2006b), the utility from consumption is not product-specific. Instead, the product-specific utility is revealed at the time of purchase. Thus, consumption is not affected by which brand is in storage.

The ability to store products allow households to smooth consumption level when the price is high. However, holding inventory is costly to the consumer. The cost the consumer incurs to hold inventory of each beverage category at time *t* is $f(i_{h,t+1}^s)$ and $f(i_{h,t+1}^c)$ for soda and coffee, respectively. Note that $i_{h,t+1}^s$ and $i_{h,t+1}^c$denote household $h's$ level of soda and coffee inventories, respectively. Households' preference for storing products depend on the marginal cost of inventory for each category. The end-of-period inventory for each category is equal to the inventory leftover at the beginning of the period, plus the purchase in period *t*, minus the level of consumption during period *t*. The cost of the inventory functions are specified as follows:

$$f(i_{h,t+1}^s) = \beta_1 i_{h,t+1}^s + \beta_2 (i_{h,t+1}^s)^2, \qquad (5)$$

$$f(i_{h,t+1}^c) = \beta_3 i_{h,t+1}^c + \beta_4 (i_{h,t+1}^c)^2, \tag{6}$$

where β_1, β_2, β_3 and β_4 are parameters to be estimated. We expect $\beta_1 > 0$, $\beta_3 > 0$, $\beta_2 > 0$ and $\beta_4 > 0$, which yield inventory cost functions that are increasing and convex.

Given all components of utility, household *h* chooses the product to purchase, and the quantity of each category to consume in order to maximize the discounted value of expected future utility. We denote household $h's$ purchase of brand *j* of package size *x* in market *m* at time *t* by $d_{hjxmt} = 1$, with $\sum_{j,x} d_{hjxmt} = 1$. The consumer's problem in period *t* can be represented as

$$\begin{aligned} V(\phi_t) \\ = \max_{\{c_{ht}^s, c_{ht}^c, d_{hjxmt}\}} \sum_{t=1}^{\infty} \delta^{t-1} \; E[\, u(c_{ht}^s, c_{ht}^c, \nu_{ht}^s, \nu_{ht}^c) - f(i_{h,t+1}^s) - f(i_{h,t+1}^c) \\ + \sum_{j,x} d_{hjxmt} \, g(p_{jxmt}, A_{jxmt}, \xi_{hjmx}, \varepsilon_{hjxmt}) \,|\phi_t] \end{aligned} \tag{7}$$

s.t. $c_{ht}^s, c_{ht}^c, i_{ht}^s, i_{ht}^c, x_{ht} \geq 0$,

$\sum_{j,x} d_{hjxmt} = 1$,

$i_{h,t+1}^s = i_{ht}^s + r_t x_{ht} - c_{ht}^s$,

$i_{h,t+1}^c = i_{ht}^c + (1 - r_t) x_{ht} - c_{ht}^c$,

where $\delta > 0$ is the discount factor for each period; ϕ_t denotes the state at time *t*; r_t is a zero-one indicator variable that equals 1 if the unit purchased/chosen is soda, 0 otherwise. The state ϕ_t consists of the current price, promotional activities, the beginning-of-period inventory for each category, and the two types of uncertainty to households: (*i*) the shock to the utility of consumption; and (*ii*) the shock to the utility of purchase. For notation simplicity, we drop the subscripts *h* and *m* in what follows.

Empirically, we follow Hendel and Nevo (2006b) and make the following three assumptions.

Assumption 1: The random shocks to consumption at each period ν_{ht}^s and ν_{ht}^c are independent and identically distributed (*i.i.d.*) across individuals and over time.

Assumption 2: Price and promotional activities follow an exogenous first-order Markov process.

Assumption 3: The random shock term in the utility of brand choice purchase follows a type I extreme value distribution, and is *i.i.d.* across individuals and over time.

These assumptions are commonly made in related literature to reduce the state space and to produce a tractable solution to the dynamic programming problem.

The third assumption allows us to derive a closed-form probability of observing each household *h*'s purchase history conditional on state variables and inventory:

$$\Pr(d_{jxt}|\phi_t) = \frac{\exp(\alpha_1 p_{jxt} + \alpha_2 A_{jxt} + \sum_l^L \alpha_{3l}\, p_{jxt} Demo_{hl} + \xi_{jxt} + M(\phi_t, j, x))}{\sum_{k,y} \exp(\alpha_1 p_{kyt} + \alpha_2 A_{kyt} + \sum_l^L \alpha_{3l}\, p_{kyt} Demo_{hl} + \xi_{kyt} + M(\phi_t, k, y))}, \tag{8}$$

where

$$M(\phi_t, j, x) = Max_c\{u(c_{ht}^s, c_{ht}^c, \nu_{ht}^s, \nu_{ht}^c) - f(i_{h,t+1}^s) - f(i_{h,t+1}^c) + \delta E(V(\phi_{t+1})|d_{jxt}, c_{ht}^s, c_{ht}^c, \phi_t)\}$$

and $E(V(\cdot))$ is the expected value of the future utility as a function of the current state and household's decision.

3.2 Demand Estimation

Following a three-step procedure proposed by Hendel and Nevo (2006b), we estimate parameters of this dynamic model by maximizing the likelihood of households' product choices. Allowing heterogeneity in brand preferences, the estimation procedure involves splitting the consumer's problem into a static brand choice and a dynamic quantity choice. As discussed in Hendel and Nevo (2006b), splitting the consumer's problem in this way greatly reduces the dimensionality of the state space when estimating the dynamic quantity, which in turn decreases the computational burden.

The probability of choosing a product *j* is effectively a joint probability of choosing a product and package size, which can be written as:

$$\Pr(d_{jt}, x_t) = \Pr(d_{jt} = 1|p_t, x_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c) \times \Pr(x_t|p_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c), \tag{9}$$

where $\Pr(d_{jt} = 1|p_t, x_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c)$ is the probability of choosing brand *j* conditional on wanting a product of package size x_t; while $\Pr(x_t|p_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c)$ is the probability of choosing package size x_t, which is effectively the quantity choice. Hendel and Nevo (2006b) prove and discuss in detail why $\Pr(d_{jt} = 1|p_t, x_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c)$ can simply be estimated using a static conditional logit prior to estimating $\Pr(x_t|p_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c)$ with the dynamic quantity choice portion of the model.

3.2.1 The static conditional logit model

Assuming ε is distributed type I extreme value, the probability of choosing product/brand *j* conditional on package size is:

$$\Pr\left(d_{jt} = 1\middle|x_t, p_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c\right) = \frac{\exp(\alpha_1 p_{jxt} + \alpha_2 A_{jxt} + \sum_l^L \alpha_{3l}\, p_{jxt} Demo_{hl} + \xi_{jxt})}{\sum_k \exp(\alpha_1 p_{kxt} + \alpha_2 A_{kxt} + \sum_l^L \alpha_{3l}\, p_{kxt} Demo_{hl} + \xi_{kxt})}, \qquad (10)$$

where the denominator sums over all available brands within package size *x* at time *t*. Hence the estimation of brand choice parameters is based on a purely static conditional logit model. Each household faces a choice set that includes all products/brands offered in the given package size, which is the size of the household actual purchases.

The product price variable is well-known to be endogenous since product attributes captured in ξ_{jxt}, such as brand quality and TV commercials are unobserved to researchers but observed by households making purchase decisions, and these attributes are correlated with product price. As such, we follow Gayle and Xie (2018), and formally account for the endogeneity of price by first specifying the following reduced-form price equation:

$$p_{jxt} = Z_{jxt}\lambda + \epsilon_{jxt}, \qquad (11)$$

where Z_{jxt} is a matrix of non-price product characteristics and a set of instrument variables that influence product price; λ is a vector of parameters associated with the variables in Z_{jxt}; and ϵ_{jxt} is assumed to be independently and identically distributed normal random price shock variable, with mean zero and standard deviation σ^p.

It is reasonable to conjecture that production marginal cost, such as electricity cost, affects product price, but is uncorrelated with the random component of purchase utility. Motivated by Villas Boas (2007), we interact the price of electricity with brand-size dummies to allow changes in electricity price to affect product-level marginal costs, and in turn product prices, differently across brands and package sizes. Thus, the interaction of electricity price with brand-size dummies is one set of instruments we include in Z_{jxt}. In addition, other non-price product characteristics are included in Z_{jxt}: (1) the caffeine content of each product we read from the package; (2) the promotional activities, such as where the product is displayed in a store and whether there is special product advertising in a store; (3) the interaction between electricity price and household income; (4) the interaction between caffeine content and whether the family has kid under 17 years old; and (5) brand-size fixed effects.

Therefore, the probability of households' purchase decisions conditional on package size is captured by the following closed-form expression:

$$Pr\left(d_{hjmt} = 1 \middle| p_{mt}, x_t, \epsilon_{jxt}\,; \alpha\right) = \frac{exp\left(\alpha_1 p_{jxmt} + \alpha_2 A_{jxmt} + \sum_l^L \alpha_{3l}\, p_{jxmt} Demo_{hl} + \frac{\epsilon_{jxmt}}{\sigma^p}\right)}{\sum_k exp\left(\alpha_1 p_{kxmt} + \alpha_2 A_{kxmt} + \sum_l^L \alpha_{3l}\, p_{kxmt} Demo_{hl} + \frac{\epsilon_{kxmt}}{\sigma^p}\right)}, \tag{12}$$

where $\frac{\epsilon_{jxmt}}{\sigma^p}$ is included in the closed-form probability expression to mitigate the endogeneity problem. The likelihood function for the static conditional logit model, taking price endogeneity into account is:

$$L(\alpha, \lambda) = \prod_h \prod_j \prod_m \prod_t d_{hjmt} \Pr\left(d_{hjmt} = 1 \middle| p_{mt}, x_t, \epsilon_{jxmt};\ \alpha\right) \times \Phi\left(\epsilon_{jxmt} | Z_{jxmt};\ \lambda\right), \tag{13}$$

where d_{hjmt} is an observed zero-one indicator variable that is equal to 1 if household h chooses product j in market m during period t, and 0 otherwise. The $\Pr\left(d_{hjmt} = 1 \middle| p_{mt}, x_t, \epsilon_{jxmt}; \alpha\right)$ component characterizes the conditional likelihood of the logit product choice probabilities. To obtain the likelihood that is unconditioned on ϵ_{jxmt}, we multiply the conditional likelihood by the probability of observing specific values of ϵ_{jxmt}, where $\epsilon_{jxmt} = p_{jxmt} - Z_{jxtm}\lambda$ based on equation (11). Since we assume that ϵ_{jxmt} follows a normal distribution with mean zero and standard deviation σ^p, then $\Phi(\cdot)$ is the normal probability density function.

A convenient feature of the likelihood function above is that it enables identifying the parameter vector λ separately from the parameter vector α. As such, these parameters can be estimated using a two-step procedure. In the first step, we estimate λ by using ordinary least square estimation of equation (11). Given $\hat{\lambda}$, we can compute $\hat{\epsilon}_{jxmt}$, $\hat{\sigma}^p$ and $\Phi\left(\hat{\epsilon}_{jxmt} | Z_{jxmt};\ \hat{\lambda}\right)$. In the second step, we can plug them into equation (13) and identity $\hat{\alpha}$, which maximizes the likelihood function $L(\alpha, \hat{\lambda})$.

3.2.2 Expected utility from purchasing size *x*

After recovering the parameters for brand preferences, we compute an inclusive value for each package size across the beverage categories, which is the expected utility a household gets from consuming the corresponding category and package size. It is given by:

$$\omega_{xt} = log\left[\sum_j exp\left(\alpha_1 p_{jxt} + \alpha_2 A_{jxt} + \sum_l^L \alpha_{3l}\, p_{jxt} Demo_{hl} + \frac{\epsilon_{jxmt}}{\sigma^p}\right)\right], \qquad (14)$$

where *j* indexes the brands in package size *x*. As discussed in Hendel and Nevo (2006b), the state variables can be compressed into the single index ω_{xt}, such that $F(\omega_t|\phi_{t-1})$ can be summarized by $F(\omega_t|\omega_{t-1})$, where $F(\cdot)$ is the cumulative probability distribution function that characterizes the state transition of ω_t.

3.2.3 The simplified dynamic problem

In the simplified dynamic problem, the consumer decides the quantity to consume given the expected utility from consuming a package size *x*. The Bellman equation associated with the simplified dynamic problem is:

$$V(\omega_{xt}, i_t^s, i_t^c, \nu_t^s, \nu_t^c) = \underset{(c^s, c^c, x)}{\text{Max}} \{u(c_t^s, c_t^c, \nu_t^s, \nu_t^c) - f\left(i_{h,t+1}^s\right) - f\left(i_{h,t+1}^c\right) + \omega_{xt} + \varepsilon_{jxt} +$$
$$\delta E[V(\omega_{t+1}, i_{t+1}^s, i_{t+1}^c, \nu_{t+1}^s, \nu_{t+1}^c)|\omega_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c]\}, \qquad (15)$$

where $V(\cdot)$ is the value function at state $(\omega_{xt}, \nu_t^s, \nu_t^c, i_t^s, i_t^c)$ and is the unique solution to the Bellman equation.

The extreme value probability distribution assumption on ε_{jxt} implies the following closed-form solution of the dynamic problem in terms of the integrated value function $\bar{V}$:

$$\bar{V} = \sigma_\varepsilon log\left[\sum_x exp\left(\frac{u(\cdot) - f^s(\cdot) - f^c(\cdot) + \omega_{xt} + \delta F_x \bar{V}}{\sigma_\varepsilon}\right)\right], \qquad (16)$$

where σ_ε is the dispersion parameter; $x = 0, 1, 2, \ldots X$, where $x = 0$ represents the case that the household chooses the outside option, which is any other option that is not one of the six sizes we consider; $\bar{V}$ is an $N \times 1$ vector of unique values that solves the fixed-point problem in equation

(16), where N is the number of unique states a household faces; and F_x is the $N \times N$ transition probability matrix conditional on choosing size *x*.

The assumed exogenous state variables in the simplified dynamic problem are: $\{\omega_{t,x\,=\,0,1,2,\ldots,X}, c_t^s, c_t^c\}$. We discretize each state variable based on 10 percentiles and obtain the possible unique combinations, N, of the state variables each household faces. In addition, we assume each exogenous state variable follows an AR (1) process, which yields the following equations:

(i) $\omega_{yt} = \rho_y^{\omega} + \sum_{x=1}^{X} \rho_{y,x}^{\omega} \omega_{x,t-1} + \zeta_{yt}^{\omega}$, for $y = 1, 2, \ldots, X$

(ii) $c_t^s = \rho_0^s + \rho_1^s c_{t-1}^s + \zeta_t^s$

(iii) $c_t^c = \rho_0^c + \rho_1^c c_{t-1}^c + \zeta_t^c$,

where ζ_{yt}^{ω}, ζ_t^s, and ζ_t^c are assumed to be normally distributed. Therefore, we can compute the transition probability matrices for each household. For example, each entry of the transition probability matrix for package size 1 is determined by:

$$Pr\big(\omega_{1,t+1}, c_{t+1}^s, c_{t+1}^c | x_t = 1, \omega_{1t}, \omega_{2t}, \ldots, \omega_{Xt}, c_t^s, c_t^c\big) =$$

$$Pr\big(\omega_{1,t+1} \big| \omega_{1t}, \omega_{2t}, \ldots, \omega_{Xt}\big) * Pr(\, c_{t+1}^s | \, c_t^s) * Pr(\, c_{t+1}^c | \, c_t^c)$$

Next, the probability of purchasing quantity *x* in terms of the value $\bar{V}$ that satisfies the integrated value function in equation (16) is given by:

$$\Pr(x_t | \omega_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c; \gamma, \beta_1, \beta_2, \beta_3, \beta_4)$$

$$= \frac{\exp\big(\omega_{xt} + max\,\{u(c_t^s, c_t^c, \nu_t^s, \nu_t^c; \gamma) - f(i_{h,t+1}^s; \beta_1, \beta_2) - f(i_{h,t+1}^c; \beta_3, \beta_4) + \delta F_x \bar{V}\,\}\big)}{\sum_x \exp\big(\omega_{xt} + max\,\{u(c_t^s, c_t^c, \nu_t^s, \nu_t^c; \gamma) - f(i_{h,t+1}^s; \beta_1, \beta_2) - f(i_{h,t+1}^c; \beta_3, \beta_4) + \delta F_x \bar{V}\,\big)}. \quad (17)$$

Given a series of quantity choices for each household over time, the probability in equation (17) is used to construct the following likelihood function:

$$L(\gamma, \beta_1, \beta_2, \beta_3, \beta_4, \sigma_\varepsilon) = \prod_h \prod_x \prod_t D_{hxt} \Pr(x_t | \omega_t, i_t^s, i_t^c, \nu_t^s, \nu_t^c; \gamma, \beta_1, \beta_2, \beta_3, \beta_4), \quad (18)$$

where the D_{hxt} is a zero-one indicator variable that takes the value 1 when household *h* is observed choosing quantity size *x* during period *t*, but 0 otherwise. We estimate the dynamic parameters, $(\gamma, \beta_1, \beta_2, \beta_3, \beta_4, \sigma_\varepsilon)$, by maximizing the likelihood in (18) with respect to these parameters.

Following the practice in many dynamic demand estimations, a value for the time discount parameter, δ, is assumed, not estimated. We assume δ to be 0.96, which is consistent with a 4% interest rate.

The process of estimating the dynamic parameters is divided into two loops: (*i*) an outer loop; and (*ii*) an inner loop. The outer loop iterates on different values for the dynamic parameters to maximize the log likelihood, while the inner loop solves the dynamic programming problem described in equation (16) on each iteration given the associated set of dynamic parameters of the outer loop. Since there exists a fixed point that satisfies the integrated Bellman equation in (16), the optimal $\bar{V}$ can be solved in the inner loop using value function iterations.

To summarize, the dynamic model proposed by Hendel and Nevo (2006b) imposes several assumptions to enable splitting the likelihood of the dynamic choice problem into two components: (*i*) a product brand choice conditional on quantity size; and (*ii*) a quantity size choice. First, households' brand preferences are inferred at a specific point in time conditional on the quantity they decide to purchase. We then use the estimated brand preferences to summarize state variables influencing these preferences into a single index, which is the inclusive value, i.e. expected utility, associated with a given quantity size of the products. The inclusive value indexes that correspond to the available quantity sizes are brought to the dynamic portion of the problem, which focuses on households' optimal quantity size choices. In this last step, we recover households' dynamic behavior parameters by solving a simplified version of the dynamic problem based on households' observed quantity size purchase choices. Even though the assumptions needed for splitting the full dynamic decision problem are restrictive, they simplify the dimension of the state space in the dynamic problem and reduce the computational burden.

3.2.4 Dynamic Demand Responses to Price Changes

An important feature of a dynamic demand model is its ability to capture how changes in expected future utility due to price changes influence consumers' current purchase decisions. Given that consumers' decisions on the quantity size (package size) to purchase are informed by forward-looking considerations, we can derive their purchase responses to a price change, which is captured by the derivatives of probabilities of choosing a product (a combination of brand and size) with respect to a price.

First, we rewrite the choice probability in equation (9):

$$Pr(j) = Pr(j|x) \times Pr(x). \tag{19}$$

Thus, the own-price effect can be expressed as:

$$\frac{\partial Pr(j)}{\partial p_j} = \frac{\partial Pr(j|x)}{\partial p_j}\ Pr(x) + Pr(j|x)\frac{\partial Pr(x)}{\partial \omega_x}\frac{\partial \omega_x}{\partial p_j}, \tag{20}$$

where product $j \in x, p_j$ is the price of product *j*, and ω_x is the inclusive value associated with the group of products in package size *x*. From equation (10) we can compute $\frac{\partial Pr(j|x)}{\partial p_j} = (\alpha_1 + \sum_l^L \alpha_{3l}\, Demo_{hl})Pr(j|x)[1 - Pr(j|x)]$, which is expected to be negative since we expect $\alpha_1 < 0$, and $|\alpha_1| > \left|\sum_l^L \alpha_{3l}\, Demo_{hl}\right|$ and therefore $\left(\alpha_1 + \sum_l^L \alpha_{3l}\, Demo_{hl}\right) < 0$ for all *h*. Equation (14) implies $\frac{\partial \omega_x}{\partial p_j} = (\alpha_1 + \sum_l^L \alpha_{3l}\, Demo_{hl})Pr(j|x)$, which is expected to be negative since $(\alpha_1 + \sum_l^L \alpha_{3l}\, Demo_{hl}) < 0$, suggesting that when product *j* belongs to size *x* and the price of product *j* increases, the expected utility from having the option to choose among the products within package size *x* decreases. $\frac{\partial Pr(x)}{\partial \omega_x}$ is expected to be positive, suggesting households are more likely to purchase size *x* when the expected utility from consuming quantity *x* marginally increases, *ceteris paribus*. In summary, the expected sign of every component on the right-hand-side of equation (20) is determined, and they jointly reveal that the own-price effects are expected to be negative.

The cross-price effect among products from the same package size can be computed as follows:

$$\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} = \underbrace{\frac{\partial Pr(k|x)}{\partial p_j}\ Pr(x)}_{B_1} + \underbrace{Pr(k|x)\frac{\partial Pr(x)}{\partial \omega_x}\frac{\partial \omega_x}{\partial p_j}}_{B_2}, \tag{21}$$

where both *j* and *k* are products in the group of size x, and *j* ≠ *k*. Using equation (10), it can be shown that $\frac{\partial Pr(k|x)}{\partial p_j}$ in the B_1 term on the right-hand-side of equation (21) is $\frac{\partial Pr(k|x)}{\partial p_j} = -(\alpha_1 + \sum_l^L \alpha_{3l}\, Demo_{hl})Pr(k|x)Pr(j|x)$, which is positive since $\left(\alpha_1 + \sum_l^L \alpha_{3l}\, Demo_{hl}\right) < 0$. The rationale for a positive $\frac{\partial Pr(k|x)}{\partial p_j}$ is the following. Conditional on the purchase quantity *x*, i.e. package size *x*, holding everything else constant, a higher price of product *j* would increase the probability of choosing product *k*, an alternative product within package size *x*. Mathematically,

$Pr(k|x)$ is predicted by a static conditional logit model, where alternatives in the same choice set are assumed to be substitutes in a static discrete choice demand model. Therefore, the partial derivative in the B_1 term on the right-hand-side of equation (21), which is generated from the static portion of our demand model, effectively says that products within package size *x* are inherently substitutes. Since $Pr(x)$ is positive due to being a probability, then the B_1 term on the right-hand-side of equation (21) is positive, which can be re-written as:

$$B_1 = -\frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k|x)Pr(j|x)Pr(x), \tag{22}$$

where $\tilde{u}_{hj} = \alpha_1 p_j + \alpha_2 A_j + \sum_l^L \alpha_{3l}\, p_j Demo_{hl} + \xi_j$, and $\frac{\partial \tilde{u}_{hj}}{\partial p_j} = (\alpha_1 + \sum_l^L \alpha_{3l}\, Demo_{hl}) < 0$.

Now let's consider the remaining terms on the right-hand-side of equation (21), collectively labeled B_2. $Pr(k|x)$ is positive due to being a probability, while $\frac{\partial Pr(x)}{\partial \omega_x}$ is positive and $\frac{\partial \omega_x}{\partial p_j}$ is negative as discussed previously. Thus, the B_2 term on the right-hand-side of equation (21) is negative, implying that if the price of product *j* increases, which in turn decreases the expected utility, ω_x, from having the option to choose among the products within package size *x*, then the probability of choosing size *x* will in turn decline. Therefore, when the price of product *j* increases, the B_2 term in equation (21) is incentivizing the consumer to either switch to a different package size, or the outside option. Note that by the consumer switching to a different package size effectively decreases the demand for all products within package size *x* in response to an increase in the price of product *j*, yielding a "complementary" type demand effect between product *j* and other products within package size *x*.

In summary, term B_1 in equation (21) captures consumers' incentive to switch their purchase decision to alternatives within size category x in response to a change in price of product j, a product within size category x. However, term B_2 captures consumers' incentive to switch their purchase decision to alternatives within another size category other than x, or the outside option, in response to a change in price of product j.

A Closer Look at the Terms in Equation (21)

Without loss of generality let's limit the size categories to be only x and y, plus the outside option. As such, the expression for the probability of choosing quantity size category *x* is:

$$\Pr(x) = \frac{\exp\left(\omega_x + u(\cdot) - f^s(\cdot) - f^c(\cdot) + \delta F_x \bar{V}(\omega_x, \omega_y)\right)}{1 + \underbrace{\exp\left(\omega_x + u(\cdot) - f^s(\cdot) - f^c(\cdot) + \delta F_x \bar{V}(\omega_x, \omega_y)\right)}_{Q_1} + \underbrace{\exp\left(\omega_y + u(\cdot) - f^s(\cdot) - f^c(\cdot) + \delta F_y \bar{V}(\omega_x, \omega_y)\right)}_{Q_2}}, \tag{23}$$

where δ is the discount factor for each period; $\bar{V}$ represents the integrated value function that yields an $N \times 1$ vector of unique values that correspond to the N number of unique states a household faces; and F_x and F_y are $N \times N$ transition probability matrices conditional on choosing size *x* or size *y*, respectively. Using equation (23), it can be shown that,

$$\frac{\partial Pr(x)}{\partial \omega_x} = Pr(x)\left\{1 - Pr(x) + \delta\left[F_x - F_x Pr(x) - F_y Pr(y)\right]\frac{\partial \bar{V}}{\partial \omega_x}\right\}. \tag{24}$$

From equation (24), it is evident that $\frac{\partial Pr(x)}{\partial \omega_x} > 0$ yields the restriction that $1 - Pr(x) + \delta\left[F_x - F_x Pr(x) - F_y \Pr(y)\right]\frac{\partial \bar{V}}{\partial \omega_x} > 0$.

Using equation (24), and the fact that $\frac{\partial \omega_x}{\partial p_j} = \frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(j|x)$, the following expression for B_2 can be obtained:

$$B_2 = \frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k) Pr(j|x)\left\{1 - Pr(x) + \delta\left[F_x - F_x Pr(x) - F_y Pr(y)\right]\frac{\partial \bar{V}}{\partial \omega_x}\right\}, \tag{25}$$

Note that $B_2 < 0$ because $\frac{\partial \tilde{u}_{hj}}{\partial p_j} < 0$, and all other terms on the right-hand-side of equation (25) are positive. Using equations (21), (22) and (25), we obtain:

$$\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} = B_1 + B_2$$

$$= \frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k) Pr(j|x) \underbrace{\left\{\delta\left[F_x - F_x Pr(x) - F_y Pr(y)\right]\frac{\partial \bar{V}}{\partial \omega_x} - Pr(x)\right\}}_{A_1} \tag{26}$$

Since $\frac{\partial \tilde{u}_{hj}}{\partial p_j} < 0$, the sign of $B_1 + B_2$ depends on the sign of term A_1 in equation (26). If $A_1 < 0$, then $B_1 + B_2 > 0$, while if $A_1 > 0$, then $B_1 + B_2 < 0$.

The ultimate relationship between products j and k, which is captured by the sign of the partial derivative, $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x}$, crucially depends on the relative strengths of the B_1 and B_2 terms on the right-hand-side of equation (21), which we ultimately see is determined by the sign of term A_1 in equation (26). Equation (26) reveals that the forward-looking nature of consumers, which is determined by the value of δ in term A_1, plays a key role in influencing the magnitude, and sometimes the sign, of term A_1. Note that the impact on future utility resulting from the price change is captured by $\left[F_x - F_x Pr(x) - F_y Pr(y)\right]\frac{\partial \bar{V}}{\partial \omega_x}$ in term A_1, which is only allowed to influence the consumer's current choice when $\delta > 0$, i.e. when the consumer is forward-looking. At an extreme where consumers are not forward-looking and therefore do not account for impacts on future utility induced by changes in state variable(s), i.e. $\delta = 0$, then $A_1 = -Pr(x) < 0$, and

$$\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} = -\frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k)Pr(j|x)Pr(x) = -\frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k)Pr(j). \quad (27)$$

Note that the right-hand-side expression in equation (27) is positive, implying that the products are substitutes. Furthermore, the right-hand-side expression in equation (27) is identical to the cross-price effect that would be generated from a traditional static logit model of demand [see Train (2009)].[2] However, when consumers are forward-looking, i.e. $0 < \delta < 1$, then term A_1 shown in equation (26) can either be positive or negative. Table A1 in the appendix provides model estimates of term $A_1$for a few states in our data sample, clearly showing that the value of A_1 is positive at some states, but negative at some states.

In summary, the B_1 term is positive implying that the products/brands within a given package size category are inherently substitutes, while the B_2 term is negative due to incentivized consumer quantity choice behavior across package sizes that captures incentives for "complementary type" consumer choice behavior among pairs of products within a package size category. As such, if the B_2 term is sufficiently strong, then the ultimate cross-price effect between

[2] See page 58 in Chapter 3 in Train, Kenneth E. (2009), "Discrete Choice Methods with Simulation," second edition, *Cambridge University Press*.

products *j* and *k* captured by $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x}$ will be negative, suggesting that products *j* and *k* should be empirically treated as complements, an outcome in the framework that is only possible when consumers are forward-looking in their decision-making such that the impact on their future utility is taken into account when making current choices. These results are summarized in **Proposition 1**.

Proposition 1: *The dynamic demand model presented in this paper yields a cross-price effect between distinct products j and k having the same package size* x*, i.e.* $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x}$*, that can either be negative, implying the products are empirical complements, or positive, implying the products are empirical substitutes. Furthermore, whether the two products are empirical complements depends on whether incentivized changes in consumers' optimal quantity choice outweighs the perceived inherent substitutability between the products, which is only possible in the framework when consumers are forward-looking in their decision-making such that the impact on their future utility is taken into account when making current choices.*

Proof of Proposition 1: From equation (21), $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} = \underbrace{\frac{\partial Pr(k|x)}{\partial p_j}\, Pr(x)}_{B_1} + \underbrace{Pr(k|x)\frac{\partial Pr(x)}{\partial \omega_x}\frac{\partial \omega_x}{\partial p_j}}_{B_2}$, where $B_1 > 0$ and $B_2 < 0$. From discussions in the text, $B_1 > 0$ due to the percieved inherent substitutability between the products, while $B_2 < 0$ due to incentivized changes in consumers' optimal quantity choice. First, it is straightforward to see that, $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} < 0$ if $|B_2| > |B_1|$, but $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} > 0$ if $|B_2| < |B_1|$. In addition, equation (26) establishes that $B_1 + B_2 = \frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k)Pr(j|x)\underbrace{\left\{\delta\left[F_x - F_x Pr(x) - F_y Pr(y)\right]\frac{\partial \bar{V}}{\partial \omega_x} - Pr(x)\right\}}_{A_1}$. Since $\frac{\partial \tilde{u}_{hj}}{\partial p_j} < 0$, the sign of $B_1 + B_2$ depends on the sign of term A_1. If $A_1 < 0$, then $B_1 + B_2 > 0$, while if $A_1 > 0$, then $B_1 + B_2 < 0$. It is evident that A_1 is always negative when $\delta = 0$. As such, A_1 can only be positive when $\delta > 0$, which is necessary for expected changes in future utility, $\left[F_x - F_x Pr(x) - F_y Pr(y)\right]\frac{\partial \bar{V}}{\partial \omega_x}$, to influence current consumer choices, i.e. when consumers are forward-looking in their decision-making. *QED.*

A Static Nested Logit Model

We now show that a static nested logit model does capture the "complementary type" incentives of consumers described above, however the static nested logit model imposes a restriction on the size of such "complementary type" incentives that causes the inherent substitutability of products within a size category to always dominate. Consider the indirect utility consumer h obtains from purchasing product j being specified as follows:

$$U_{hj} = \bar{u}_j(p_j) + \rho\zeta_{hx} + (1-\rho)\varepsilon_{hj} \tag{28}$$

where product groups, i.e. nests, are indexed by x, and therefore product groups here correspond to package size categories. In equation (28), $\bar{u}_j$, which is a function of the price of product j, is the mean utility across consumers; ζ_{hx} is a random component of utility that is common to all products in group x; and ε_{hj} is a mean-zero random component of utility that is specific to product j. The law of demand requires that $\frac{\partial \bar{u}_j}{\partial p_j} < 0$. The parameter, ρ lies between zero and one, $0 \leq \rho < 1$, and measures the correlation of the consumers' utility across products belonging to the same group.

Let there be $\mathcal{G}_x$ products in group x. The well-known formula for the probability of choosing product j conditional on choosing group x is [see Berry (1994)]:

$$Pr(j|x) = \frac{e^{\frac{\bar{u}_j}{(1-\rho)}}}{D_x} \tag{29}$$

where $D_x = \sum_{j \in \mathcal{G}_x} e^{\frac{\bar{u}_j}{(1-\rho)}}$. Furthermore, the well-known formula for the probability of choosing group x is:

$$Pr(x) = \frac{D_x^{(1-\rho)}}{1+\sum_{x=1}^{X} D_x^{(1-\rho)}} \tag{30}$$

Therefore, the unconditional probability of choosing product j is:

$$Pr(j) = Pr(j|x) \times Pr(x) = \frac{e^{\frac{\bar{u}_j}{(1-\rho)}}}{D_x} \times \frac{D_x^{(1-\rho)}}{1+\sum_{x=1}^{X} D_x^{(1-\rho)}} \tag{31}$$

The cross-price effect among products from the same group can be derived using:

$$\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} = \underbrace{\frac{\partial Pr(k|x)}{\partial p_j} Pr(x)}_{B_1^{NL}} + \underbrace{Pr(k|x)\frac{\partial Pr(x)}{\partial \bar{u}_j}\frac{\partial \bar{u}_j}{\partial p_j}}_{B_2^{NL}} \tag{32}$$

Note that equations (32) and (21) are very similar in structure. Using equations (28) through (31), it is straightforward to show that:

$$B_2^{NL} = \frac{\partial \bar{u}_j}{\partial p_j} Pr(k) Pr(j|x) \underbrace{[1 - Pr(x)]}_{C_1^{NL}} \tag{33}$$

Term B_2^{NL} captures consumers' incentive to switch their purchase decision to alternatives within package size categories other than x, or the outside option, in response to a change in price of product j, a product within size category x. Since $\frac{\partial \bar{u}_j}{\partial p_j} < 0$ and $Pr(k)$, $Pr(j|x)$, and $Pr(x)$ are all probabilities lying between zero and one, then $B_2^{NL} < 0$. In other words, $B_2^{NL} < 0$ reveals that the static nested logit model does capture incentives for "complementary type" consumer choice behavior among pairs of products within a package size category.

It is instructive to compare, B_2 in equation (25) and B_2^{NL} in equation (33). B_2 in equation (25) is the following:

$$B_2 = \frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k) Pr(j|x) \left\{ 1 \underbrace{- Pr(x) + \delta \left[F_x - F_x Pr(x) - F_y Pr(y) \right] \frac{\partial \bar{V}}{\partial \omega_x}}_{A_1} \right\}$$

which can be rewritten as:

$$B_2 = \frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k) Pr(j|x) \underbrace{\{1 + A_1\}}_{C_1} \tag{34}$$

Recall from discussions above, an ultimate complementary relationship between products k and j requires $A_1 > 0$, which would yield $C_1 > 1$. Note that for the nested logit model, equation (33) reveals that we must have $C_1^{NL} \leq 1$. In other words, the nested logit model imposes an upper bound on the size of B_2^{NL}, which is less than the attainable size of B_2 in the dynamic model that assumes consumers are forward-looking. Furthermore, it is indeed the forward-looking feature of the dynamic model that allows the possibility of $C_1 > 1$ in equation (34). It is only when consumers are forward-looking we have $\delta > 0$, which allows current purchase decisions to be influenced by changes in expected future utility resulting from the price change, captured by $\left[F_x - F_x Pr(x) - F_y Pr(y) \right] \frac{\partial \bar{V}}{\partial \omega_x}$ in term A_1. Note that when consumers are not forward-looking, then $\delta = 0$, $A_1 =$

$-Pr(x)$, and $C_1 \leq 1$, equivalent to the size restriction imposed by the static nested logit model. In fact, when $\delta = 0$, the expressions for B_2 and B_2^{NL} are similar.

For completeness, using equations (28) through (31), it is straightforward to show that:

$$B_1^{NL} = -\frac{1}{(1-\rho)}\frac{\partial \bar{u}_j}{\partial p_j} Pr(k|x)Pr(j|x)Pr(x) \tag{35}$$

where B_1^{NL} captures consumers' incentive to switch their purchase decision to alternatives within size category x in response to the change in price of product j. Using equations (32), (33) and (35), it is straightforward to show that for the static nested logit model, the following is one way to express the within-size category cross-price effect:

$$\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} = \frac{\partial \bar{u}_j}{\partial p_j} Pr(k)Pr(j|x)\underbrace{\left\{-\frac{\rho}{(1-\rho)} - Pr(x)\right\}}_{A_1^{NL}} \tag{36}$$

which can be rewritten as

$$\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} = -\frac{\partial \bar{u}_j}{\partial p_j} Pr(k)Pr(j|x)\left\{\frac{\rho}{(1-\rho)} + Pr(x)\right\} \tag{37}$$

Since $\frac{\partial \bar{u}_j}{\partial p_j} < 0$ and $0 \leq \rho < 1$, while $Pr(k)$, $Pr(j|x)$, and $Pr(x)$ are all probabilities, then $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x} = B_1^{NL} + B_2^{NL} > 0$. Therefore, the ultimate model-predicted relationship between products within size category group x are necessarily substitutes in the traditional static nested logit model because this static model imposes a restriction on the size of "complementary type" consumer incentives, captured by B_2^{NL}, that causes the inherent substitutability of products within a size category, captured by B_1^{NL}, to always dominate.

In summary, unlike a static nested logit model, the forward-looking attribute of consumers in the dynamic model increases the capacity of the discrete choice model to capture "complementary type" consumer incentives such that these incentives can at times be relatively stronger than the inherent substitutability of products within a size category, ultimately yielding "complementary type" relationships between these products. Unlike a static discrete choice model, in a dynamic discrete choice model forward-looking consumers in making their optimal purchase decision take into account how a given price change impacts future expected utility the consumer obtains from the relevant size category of products, which can serve as an extra stimulus for consumers to change their current purchase decision across size categories. This increased willingness of forward-looking consumers to switch their current purchase decision across size categories of products in response to price change of a given product is a key positive driver of the strength of the "complementary type" consumer incentives that can ultimately yield "complementary type" relationships between products within a given size category.

Products from Different Package size Categories

When products are from different package size categories, i.e. $j \in x, k \in y$, and $x \neq y$, the cross-size cross-price effect is captured by the following:

$$\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x,k\epsilon y} = Pr(k|y)\frac{\partial \Pr(y)}{\partial \omega_x}\frac{\partial \omega_x}{\partial p_j}. \tag{38}$$

Recall that $\frac{\partial \omega_x}{\partial p_j}$ is negative as discussed above. However, the sign of $\frac{\partial Pr(y)}{\partial \omega_x}$ is ambiguous, and depends on how consumers perceive the relationship between the relevant size category pair. The partial derivative $\frac{\partial Pr(y)}{\partial \omega_x}$ measures the marginal change in the probability of the consumer choosing size category y due to a change in the expected utility from having the option to choose among the products within size category x. As such, $\frac{\partial Pr(y)}{\partial \omega_x} > 0$ is interpreted as the consumer perceiving the size categories x and y as complements, while $\frac{\partial Pr(y)}{\partial \omega_x} < 0$ is interpreted as the consumer perceiving the size categories x and y as substitutes.

To better illustrate the ambiguity in the sign of $\frac{\partial Pr(y)}{\partial \omega_x}$, we again rewrite the expression for the probability of choosing a quantity size category, and limit the size categories to be only x and y, plus the outside option. As such, the expression for the probability of choosing quantity size category y is:

$$\Pr(y) = \frac{\exp\left(\omega_y + u(\cdot) - f^s(\cdot) - f^c(\cdot) + \delta F_y \bar{V}\left(\omega_x, \omega_y\right)\right)}{1 + \underbrace{\exp\left(\omega_x + u(\cdot) - f^s(\cdot) - f^c(\cdot) + \delta F_x \bar{V}\left(\omega_x, \omega_y\right)\right)}_{Q_1} + \underbrace{\exp\left(\omega_y + u(\cdot) - f^s(\cdot) - f^c(\cdot) + \delta F_y \bar{V}\left(\omega_x, \omega_y\right)\right)}_{Q_2}}.$$

(39)

Using equation (39), **Lemma 1** establishes that it is possible that the expected utility from having the option to choose among the products within size category *x* has a positive marginal impact on the probability of choosing size category *y*, and this positive marginal impact is only possible when consumers are forward-looking in their decision-making. As such, a positive shock to the expected utility of purchasing from size category *x* can result in a higher probability of purchasing from size category y.

Lemma 1: *The marginal effect of a change in the probability of choosing size category y due to a change in the expected utility from size category x is positive, .i.e.* $\frac{\partial Pr(y)}{\partial \omega_x} > 0$, *when the ratio of partials,* $\frac{\partial Q_2 / \partial \omega_x}{\partial Q_1 / \partial \omega_x}$, *is sufficiently large, otherwise* $\frac{\partial Pr(y)}{\partial \omega_x} < 0$. *Furthermore,* $\frac{\partial Pr(y)}{\partial \omega_x} > 0$ *is only possible when* $\delta > 0$, *i.e. when consumers are forward-looking in their decision-making.*

Proof of Lemma 1: Let equation (39) be represented more compactly as, $\Pr(y) = \frac{Q_2(\omega_x)}{1+Q_1(\omega_x)+Q_2(\omega_x)}$. Therefore, $\frac{\partial Pr(y)}{\partial \omega_x} = \frac{[1+Q_1+Q_2]\frac{\partial Q_2}{\partial \omega_x} - Q_2\left[\frac{\partial Q_1}{\partial \omega_x}+\frac{\partial Q_2}{\partial \omega_x}\right]}{[1+Q_1+Q_2]^2}$, and $\frac{\partial Pr(y)}{\partial \omega_x} > 0$ if $[1+Q_1+Q_2]\frac{\partial Q_2}{\partial \omega_x} - Q_2\left[\frac{\partial Q_1}{\partial \omega_x} + \frac{\partial Q_2}{\partial \omega_x}\right] > 0$, which can be rearranged to yield $\frac{\partial Q_2 / \partial \omega_x}{\partial Q_1 / \partial \omega_x} > \frac{Q_2}{1+Q_1}$. As such, $\frac{\partial Pr(y)}{\partial \omega_x} > 0$ when $\frac{\partial Q_2 / \partial \omega_x}{\partial Q_1 / \partial \omega_x} > \frac{Q_2}{1+Q_1}$, otherwise $\frac{\partial Pr(y)}{\partial \omega_x} < 0$. Note that $\frac{\partial Q_2}{\partial \omega_x} = 0$ when $\delta = 0$ since Q_2 is only a function of ω_x when $\delta > 0$. If $\delta = 0$, then $\frac{\partial Q_2}{\partial \omega_x} = 0$ and the numerator of $\frac{\partial Pr(y)}{\partial \omega_x}$ is $-Q_2\left[\frac{\partial Q_1}{\partial \omega_x}\right]$. In addition, we know that $Q_2 > 0$ and $\frac{\partial Q_1}{\partial \omega_x} > 0$, which implies that $-Q_2\left[\frac{\partial Q_1}{\partial \omega_x}\right] < 0$. Therefore, when $\delta = 0$, we must have $\frac{\partial Pr(y)}{\partial \omega_x} < 0$. *QED.*

A shock to the expected utility from size *x* directly influences term Q_1 in equation (39) through ω_x. However, note that because of the forward-looking behavior of consumers captured by the dynamic demand model via terms $\delta F_x \bar{V}(\omega_x, \omega_y)$ and $\delta F_y \bar{V}(\omega_x, \omega_y)$, a shock to expected utility ω_x indirectly influences terms Q_1 and Q_2 through $\delta F_x \bar{V}(\omega_x, \omega_y)$ and $\delta F_y \bar{V}(\omega_x, \omega_y)$, respectively. Furthermore, term Q_2 is also in the numerator of equation (39). It is important to note that the possibility of $\frac{\partial Pr(y)}{\partial \omega_x} > 0$, formally established in **Lemma 1**, crucially relies on consumers being forward-looking when making their purchase decisions, i.e. the presence of terms $\delta F_x \bar{V}\left(\omega_x, \omega_y\right)$ and $\delta F_y \bar{V}\left(\omega_x, \omega_y\right)$ in equation (39), which are terms that are absent from traditional static discrete choice demand models.

In the case where $\frac{\partial Pr(y)}{\partial \omega_x}$ is positive, then $\frac{\partial Pr(y)}{\partial \omega_x}\frac{\partial \omega_x}{\partial p_j}$ is negative, and equation (38) yields a negative cross-price effect between products *j* and *k*, rendering them empirical complements. The intuition is that a decrease in the price of product *j*, a product within size category *x*, increases the expected utility from size category *x*, which, somewhat tantamount to a positive income effect, may be sufficient to induce the consumer to sometimes purchase product *k*, a product from another size category. Since a fall in the price of product *j* resulted in an increase in the demand for product *k*, then these cross-size category products are empirical complements.

On the other hand, when $\frac{\partial Pr(y)}{\partial \omega_x}$ is negative, then $\frac{\partial Pr(y)}{\partial \omega_x}\frac{\partial \omega_x}{\partial p_j}$ is positive, and equation (38) yields a positive cross-price effect between products *j* and *k*, rendering them empirical substitutes. The intuition is that a decrease in the price of product *j*, a product within size category *x*, increases the expected utility from size category *x*, which may be sufficient to lure consumers away from purchasing product *k*, a product from another size category. Since a fall in the price of product *j* resulted in a decrease in the demand for product *k*, then these cross-size category products are empirical substitutes.

A Closer Look at the Terms in Equation (38)

Using equation (39), it can be shown that,

$$\frac{\partial Pr(y)}{\partial \omega_x} = Pr(y)\underbrace{\left\{\delta\left[F_y - Pr(y)F_y - Pr(x)F_x\right]\frac{\partial \bar{V}}{\partial \omega_x} - Pr(x)\right\}}_{D_1}. \qquad (40)$$

Furthermore, using equation (40) along with the fact that $\frac{\partial \omega_x}{\partial p_j} = \frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(j|x)$, we can rewrite equation (38) as:

$$\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} = \frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(j|x)Pr(k)\underbrace{\left\{\delta\left[F_y - Pr(y)F_y - Pr(x)F_x\right]\frac{\partial \bar{V}}{\partial \omega_x} - Pr(x)\right\}}_{D_1}. \qquad (41)$$

Since $\frac{\partial \tilde{u}_{hj}}{\partial p_j} < 0$, the sign of $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y}$ depends on the sign of term D_1 in equation (41). If $D_1 < 0$, then $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} > 0$ while if $D_1 > 0$, then $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} < 0$.

The ultimate relationship between products j and k, which is captured by the sign of the partial derivative, $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{(j,k)\epsilon x}$, crucially depends on the sign of $\frac{\partial Pr(y)}{\partial \omega_x}$ on the right-hand-side of equation (38), which we ultimately see is determined by the sign of term D_1 in equations (40) and (41). Equation (41) reveals that the forward-looking nature of consumers, which depends on the value of δ in term D_1, plays a key role in influencing the magnitude and sign of term D_1. At an extreme where consumers are not forward-looking, i.e. $\delta = 0$, then $D_1 = -Pr(x) < 0$ and $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} = -\frac{\partial \tilde{u}_{hj}}{\partial p_j} Pr(k)Pr(j) > 0$, like in equation (27) above [see page 58 in Train (2009)]. Therefore, when consumers are not forward-looking, the products are necessarily substitutes. As such, within the framework here, the outcome of an empirical complementary relationship between two products from different size categories is only possible when consumers are forward-looking in their decision-making, i.e. when $\delta > 0$. **Proposition 2** summarizes the key results on the model-predicted relationship between cross-size category product pairs.

Proposition 2: *The dynamic demand model presented in this paper yields a cross-price effect,* $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y}$*, between distinct products j and k that are from different package size categories, x and y respectively, which can either be negative, implying the products are empirical complements, or positive, implying the products are empirical substitutes. Specifically, the sign of* $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y}$ *depends on the sign of* $\frac{\partial Pr(y)}{\partial \omega_x}$ *as follows:*

(i) *If* $\frac{\partial Pr(y)}{\partial \omega_x} > 0$*, then* $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} < 0$*, implying that when consumers' perceive the relevant size category pair as complements, they will also perceive the relevant cross-size cross-product pairs as complements; and*

(ii) *If* $\frac{\partial Pr(y)}{\partial \omega_x} < 0$*, then* $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} > 0$*, implying that when consumers' perceive the relevant size category pair as substitutes, they will also perceive the relevant cross-size cross-product pairs as substitutes.*

Furthermore, $\frac{\partial Pr(y)}{\partial \omega_x} > 0$ is only possible when $\delta > 0$, i.e. when consumers are forward-looking in their decision-making.

Proof of Proposition 2: From equation (38), $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} = Pr(k|x)\frac{\partial Pr(y)}{\partial \omega_x}\frac{\partial \omega_x}{\partial p_j}$. From derivations and discussions in the text we know that $\frac{\partial \omega_x}{\partial p_j} < 0$ and $Pr(k|x) > 0$. Therefore, based on equation (38) and *Lemma 1*, $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} > 0$ when $\frac{\partial Pr(y)}{\partial \omega_x} < 0$, and $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} < 0$ when $\frac{\partial Pr(y)}{\partial \omega_x} > 0$. Furthermore, equations (40) and (41) establish that $\frac{\partial Pr(y)}{\partial \omega_x} = Pr(y)\underbrace{\left\{\delta\left[F_y - Pr(y)F_y - Pr(x)F_x\right]\frac{\partial \bar{V}}{\partial \omega_x} - Pr(x)\right\}}_{D_1}$ and $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} = \frac{\partial \tilde{u}_{ij}}{\partial p_j}Pr(j|x)Pr(k)\underbrace{\left\{\delta\left[F_y - Pr(y)F_y - Pr(x)F_x\right]\frac{\partial \bar{V}}{\partial \omega_x} - Pr(x)\right\}}_{D_1}$. Since $\frac{\partial \tilde{u}_{ij}}{\partial p_j} < 0$, the sign of $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y}$ depends on the sign of term D_1. If $D_1 < 0$, then $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} > 0$, while if $D_1 > 0$, then $\left.\frac{\partial Pr(k)}{\partial p_j}\right|_{j\epsilon x, k\epsilon y} < 0$. It is evident that D_1 is always negative when $\delta = 0$. As such, D_1 can only be positive when $\delta > 0$, i.e. when consumers are forward-looking in their decision-making. *QED.*

3.3 Supply

To model the supply of coffee and soda products, we assume firms simultaneously choose prices for differentiated products in a Bertrand-Nash competition setting. Let each firm, indexed by *f*, produce a set of products, denoted by F_f. Hence, in a market, firm *f* decides market prices for the set of products in F_f by solving its variable profit maximization problem:

$$\max_{p_j \forall j \in F_f} \left[\textstyle\sum_{j\in F_f}(p_j - mc_j)\, q_j\right],$$

where p_j is the price of product *j*, mc_j is the marginal cost firm *f* incurs by offering product *j*, and q_j is the quantity sold. Assuming the existence of a pure-strategy static Bertrand–Nash price equilibrium with strictly positive prices, then the price, p_j, for any product *j* satisfies the following first-order condition:

$$q_j + \textstyle\sum_{k\in F_f}(p_k - mc_k)\frac{\partial q_k}{\partial p_j} = 0, \text{ for all } j = 1,2,\dots,J. \qquad (42)$$

The market-clearing condition is that product quantity sold equals its market demand D_j. A product's market demand is equal to the probability of consumers choosing product j, $Pr(j)$, multiplied by the market population of consumers, *Pop*. As such, in equilibrium we have:

$$q_j \ = \ D_j \ = \ Pr(j) \times Pop \tag{43}$$

Therefore, the first-order conditions in equation (42) can be re-written as:

$$Pr(j) + \sum_{k \in F_f} (p_k \ - \ mc_k) \frac{\partial Pr(k)}{\partial p_j} = 0, \tag{44}$$

where we calculate $\frac{\partial Pr(k)}{\partial p_j}$ as discussed previously when we laid out the demand model. If $j = k$, then $\frac{\partial Pr(k)}{\partial p_j}$ captures the own-price effect, as shown in equation (20). If $j \neq k$, but $j, k \in x$, then $\frac{\partial Pr(k)}{\partial p_j}$ captures the within-size cross-price effect as in equation (21). The sign of the within-size cross-price effect tells us how consumers perceive the relationship between products k and *j*. If $j \neq k$, but $j \in x$, $k \in y$, and $x \neq y$, i.e. products k and *j* belong to different size categories, then $\frac{\partial Pr(k)}{\partial p_j}$ captures the cross-size cross-price effect as in equation (38), and depends on the relationship between the two size categories of products k and *j*, respectively.

Thus, the markup for product *j* can be calculated as

$$\text{Markup}(j) = p_j - mc_j = -\frac{p_j}{e_{jj}} - \sum_{k \neq j} (p_k - mc_k) \frac{q_k e_{jk}}{q_j e_{jj}}, \tag{45}$$

where $e_{jj} = \frac{\partial Pr(j)}{\partial p_j} \frac{p_j}{Pr(j)}$ is the own-price elasticity for product *j*; and $e_{jk} = \frac{\partial Pr(k)}{\partial p_j} \frac{p_j}{Pr(k)}$ is the cross-price elasticity of product *k* with product *j*.

The Impact of a Merger on Markups

Now suppose firm f acquires firm *g*, but prior to the merger firm *g* only produces product *r*. Post-merger, the merged firm maximizes the joint profit across all its products, which yields the following markup equation for product *j*:

$$Markup^{post}(j) = -\frac{p_j}{e_{jj}} - \sum_{k\neq j}(p_k - mc_k)\frac{q_k e_{jk}}{q_j e_{jj}} \underbrace{-\frac{(p_r - mc_r)q_r e_{jr}}{q_j e_{jj}}}_{Merger\ effect} \qquad (46)$$

Comparing equations (45) and (46), we see the merger effect on markup is $Merger\ effect = -\frac{(p_r - mc_r)q_r e_{jr}}{q_j e_{jj}}$. Given that own-price elasticity, e_{jj}, is negative, and the merger effect expression has a negative sign multiplying the ratio, then the sign of the merger effect term depends on the sign of the cross-elasticity term, e_{jr}. Therefore, the sign of the merger effect term depends on the relationship between products j and r. Recall that equations (26) and (41) revealed that forward-looking consumers could perceive some product pairs as complements. If products j and r are complements, then e_{jr} is negative yielding a negative merger effect term, and therefore a fall in the markup of product j due to the merger. On the other hand, if products j and r are substitutes, then e_{jr} is positive yielding a positive merger effect term, and therefore an increase in the markup of product j due to the merger.

4. Data

The data used in this study are from the IRI marketing data set [Bronnenberg et al. (2008)], which include two panels: (*i*) household-level purchasing records; and (*ii*) store-level scanner information on product sales. In the household panel, we observe the household ID, the week and store of purchase, and the product (brand-size) purchased from 2011 to 2012. The household-level data covers two separate states: Massachusetts and Wisconsin. Based on the week, store, and product ID, we collect the price and other product characteristics information such as caffeine content and promotional activities from the store-level data. While the identity of each retail store is concealed in the data, it is reported that any given store in the data is either a supermarket or drug store.

Our sample is restricted to households that meet the following criteria. First, we keep only households who made one shopping trip per week and purchased one product on a given trip. Secondly, we restrict our focus to the three most popular package sizes in each beverage category based on the quantity sold. The six (6) size categories are: (*i*) 144 fl. oz. packaged soda; (*ii*) 2 liter (67.6 fl. oz.) packaged soda; (*iii*) 20 fl. oz. single-bottle packaged soda; (*iv*) 144 fl. oz. packaged coffee; (*v*) 407 fl. oz. packaged coffee; and (*vi*) 288 fl. oz. packaged coffee. To keep consistent quantity measurement across categories, we transform the coffee package size into fl. oz. according to the brewing instruction on the package. Additionally, households who made less than 6 trips per year were dropped.

In total, the sample in our study includes the purchasing behavior of 27 households in 2012. The demographic distribution of these households is summarized in Table 1. As shown in Table 1, our sample covers households whose annual income range from less than ten thousand dollars to greater than a hundred thousand dollars. We expect higher-income families are likely to be less sensitive to price change. As such, the wide range in income levels of our households should help with identifying potential differences in price sensitivities associated with consumer heterogeneity with respect to income. Table 1 also reports data on the distribution of our households who have kids under the age of 17 years old. Whether a family has kids under 17 years old may affect its preference towards soda, which contains less caffeine than coffee.

Table 1. Household Demographic Information

Household income category (Pre-tax per year)	Number of households
$0 to $9999	2
$10,000 to 11,999	1
$12,000 to $ 14,999	1
$15,000 to $19,999	2
$20,000 to $24,999	2
$25,000 to $34,000	5
$45,000 to $54,000	1
$55,000 to $64,999	3
$65,000 to $74,000	1
$75,000 to $99,999	6
$100,000 or greater	3
Household with kids	
0 (No kid under 17)	25
1 (with kids under 17)	2

Table 2 reports the summary statistics of households shopping trips. On average, the households in our sample consume three different brands over the 53 weeks in 2012. On average, each household concentrate their purchases on two sizes. Over the 53 weeks, an average household makes ten store visits, where on each visit the household purchased either a coffee or soda product.

Table 2. Summary statistics: information on household shopping trips

	Mean	Std. Dev.	Min	Max
Number of brands	2.937	1.281	1	5
Number of sizes	2.255	0.868	1	4
Number of visits	9.541	4.169	6	19

Table 3 presents product information. With few exceptions, it tends to be the case that larger package size products are priced lower per fl. oz.. Coffee products are higher caffeinated than soda products. The data records two types of promotion activities at retail stores captured by the following two zero-one indicator variables, respectively: (*i*) *Feature*; and (*ii*) *Display*. The *Feature* variable captures whether the store advertised the product during a given week, while the *Display* variable captures whether the product was displayed in the lobby or at the end-aisle during a given week. The data in Table 3 reveals that there are more promotional activities on 144 fl. oz. packaged soda and 407 fl. oz. packaged coffee.

On each shopping trip, a household decides on which brand to choose in a choice set. A choice set is defined as a group of brands offering the same package size that are available at the store in a given week. Table 3 shows that, on average, soda has more brands available in choice sets of any given package size compared to coffee. For example, there are roughly 50 soda brands offered in the 144 fl. oz. package size, while 12 coffee brands offered in the 144 fl. oz. package size.

Table 3. Summary Statistics: product information					
	Obs	Mean	Std. Dev.	Min	Max
Panel 1: Soda 144 fl. oz.					
Number of distinct brands/products available in a given choice set for the given package size. [a]	3,366	49.56	10.51	4	64
Price (cents per fl. oz.)	3,366	2.98	0.69	1.18	13.24
Caffeine (mg per fl. oz.)	3,366	1.59	1.80	0	5.67
Feature	3,366	0.24	0.43	0	1
Display	3,366	0.20	0.40	0	1
Panel 2: Soda 67.6 fl. oz. (2L)					
Number of distinct brands/products available in a given choice set for the given package size.	1,272	25.66	3.29	8	30
Price (cents per fl. oz.)	1,272	2.29	0.46	1.19	2.96
Caffeine (mg per fl. oz.)	1,272	1.65	1.87	0	5.67
Feature	1,272	0.19	0.39	0	1
Display	1,272	0.13	0.34	0	1
Panel 3: Soda 20 fl. oz. (Single-bottle)					
Number of distinct brands/products available in a given choice set for the given package size.	345	18.77	5.01	11	25
Price (cents per fl. oz.)	345	8.06	0.75	3.95	9.45
Caffeine (mg per fl. oz.)	345	2.21	1.86	0	5.70
Feature	345	0	0	0	0
Display	345	0.01	0.08	0	1

Notes: [a] A choice set is the combination of package size, store, and week.

Table 3. Continues					
	Obs	Mean	Std. Dev.	Min	Max
Panel 4: Coffee 144 fl. oz.					
Number of distinct brands/products available in a given choice set for the given package size.	496	12.15	2.75	7	18
Price (cents per fl. oz.)	496	5.67	1.14	2.20	8.05
Caffeine (mg per fl. oz.)	496	11.39	2.27	0	11.84
Feature	496	0.08	0.27	0	1
Display	496	0.06	0.25	0	1
Panel 5: Coffee 407 fl. oz.					
Number of distinct brands/products available in a given choice set for the given package size.	31	2.74	0.68	2	4
Price (cents per fl. oz.)	31	2.65	0.55	1.96	3.44
Caffeine (mg per fl. oz.)	31	11.84	0.00	11.84	11.84
Feature	31	0.13	0.34	0	1
Display	31	0.13	0.34	0	1
Panel 6: Coffee 288 fl. oz.					
Number of distinct brands/products available in a given choice set for the given package size.	287	12.80	3.41	8	20
Price (cents per fl. oz.)	287	5.57	1.24	2.08	9.02
Caffeine (mg per fl. oz.)	287	11.67	1.39	0	11.84
Feature	287	0.09	0.29	0	1
Display	287	0.06	0.23	0	1

Notes: [a] A choice set is the combination of package size, store, and week.

Like in Hendel and Nevo (2006b), we do not observe data on consumption, and therefore approximate households’ beverage-specific consumption levels by assuming that households engage in consumption smoothing over time. We use each consumer’s purchase history (throughout the year 2011) prior to our estimation period (the year 2012) to approximate per-period smooth consumption levels of the beverage categories adopted by each consumer. The approximation is made by dividing each consumers' total purchased quantities, by product category, throughout the year prior to the estimation period by the number of weeks in the year, yielding smooth weekly beverage category-specific consumption levels for households. We assume that consumers follow their same consumption smoothing behavior throughout the estimation sample periods as we approximated for them during the prior periods. The pre-data period contains the purchase behavior of all households used for estimating the parameters of the demand model. Additionally, we allow households to experience random shocks to consumption in each beverage category in each period, which we assume follows a standard normal distribution. As such, each household's beverage-specific consumption level varies from period to period around its approximated smoothed mean level due to the random shocks.

The last piece of information we need for the estimation is the beverage-specific inventory levels, which are not observed by us the researchers. However, if we know the beginning-of-period inventory level in the initial week, once we have the consumption level and purchase decision, the end-of-period inventories can be calculated following the simple accounting rules of the model. In this study, we assume that households attempt to maintain an initial inventory for each category equal to their pre-data period historical maximum package size purchased of each category.

5. Parameter Estimates

5.1 Consumers' Brand Choice Preference Conditional on Quantity/Package Size

We begin this section by discussing the parameter estimates of the conditional logit model, which characterizes households brand choices conditional on beverage category and package size. As previously discussed, the relevant parameters are estimated using Maximum Likelihood estimation based on the likelihood function described in equation (13). The columns in Table 4 are distinguished based on whether the endogeneity of the price variable is taken into account during estimation. We report parameter estimates when the endogeneity of price is not accounted for only to facilitate readers getting a sense of the importance of accounting for the endogeneity of price. However, the remainder of the discussion focuses on the parameter estimates in column 2, which accounts for the endogeneity of price in estimation.

Consistent with economic theory, the parameter estimate on price is negative and statistically significant at conventional levels of statistical significance. All other factors held constant, households enjoy higher utility with lower product prices.

While the coefficient estimate on caffeine content is positive, it is not statistically significant at conventional levels of statistical significance. As such, the evidence is not consistent with a clear preference for products having higher levels of caffeine.

The parameter estimates on the *Display* and *Feature* variables are each positive and statistically significant at conventional levels of statistical significance. The evidence therefore suggests that store-level promotional activities captured by these variables are effective in incentivizing consumers to make product purchases.

The parameter estimate on the interaction term between price and consumer income is positive and statistically significant at conventional levels of statistical significance. Consistent with economic intuition, households with higher income are less price-sensitive compare to households with lower income.

The coefficient estimate on caffeine interacting with the *Kid* variable is positive, but not statistically significant at conventional levels of statistical significance. We are cautious about not reading too much into the statistical insignificance of this parameter estimate since there are only two households in our sample that have kids under 17 years old. In other words, the statistical insignificance could be a result of too little variation in the *Kid* variable.

Table 4. Brand Choice Conditional on Size

	Estimation Does not Account for Endogeneity of Price	Estimation Accounts for Endogeneity of Price
	1	2
Price (cents per fl. oz.)	-0.7954*** (0.1200)	-1.4846*** (0.1047)
Caffeine (mg per fl. oz.)	-0.1096*** (0.0343)	0.0315 (0.0347)
Feature (=1 if product is featured during the given week)	0.7987*** (0.2518)	0.7861*** (0.2501)
Display (=1 if product is specially displayed during the given week)	0.4440* (0.2295)	0.4181* (0.2303)
Price × Income	0.0448*** (0.0094)	0.0994*** (0.0093)
Caffeine × Kid	0.0967 (0.16290)	0.0483 (0.1639)
Brand-Size fixed effects	Yes	Yes
Log likelihood	498.56	8723.94

Notes: Parameters estimated using maximum likelihood. Standard errors are in parentheses. *** indicates statistical significance at 1%. * indicates statistical significance at 10%.

5.2 Consumers' Consumption Utility and Inventory Cost Function Parameters

The parameters of the utility from consumption and inventory cost parameters are presented in Table 5. All parameter estimates are statistically significant at conventional levels of statistical significance. The sign pattern of the inventory cost function parameter estimates suggest that inventory storage cost incurred by consumers is increasing and convex for each beverage category.

The parameter estimates in the utility from consumption function show the following sign pattern, $\gamma_1 > 0$, $\gamma_3 > 0$, $\gamma_2 < 0$, and $\gamma_4 < 0$, which imply diminishing marginal utilities and downward sloping beverage-specific demand curves, a result consistent with standard demand theory. Importantly, γ_5 is positive, indicating that on average households perceive aggregate consumptions of coffee and soda as complements rather than substitutes.

6. Using the Estimated Model for Market Analysis

We begin this section with an empirical investigation of the model- predicted relationships between pairs of products with the same package size, and between pairs of products with different package size categories. First, we focus on the within size relationships and discuss the dynamic demand elasticities, specifically how sensitive the dynamic decision on size is, and how price elastic the demand is. Subsequent to assessing the own-price and cross-price elasticities within size categories, we perform hypothetical within-beverage-category merger analyses to reveal how within-size product markups are predicted to change based on the choice behavior of forward-looking households.

For comparison, we redo the analysis based on the estimates from a pure static conditional logit model. In the pure static model, households are not assumed to be forward-looking, and therefore do not hold inventory stocks of products. Comparing the results from the two models informs us of the importance of considering the dynamics when estimating consumer demand. Finally, we simulate mergers between coffee and soda firms that are motivated by real cooperation between these firms, and analyze the competition effects of the mergers.

Table 5. Estimates of Dynamic Parameters		
Cost to Consumer of holding Soda Inventory	Parameter	Estimates
Linear	β_1	0.0008*** (0.00005)
Quadratic	β_2	0.0004*** (0.000008)
Cost to Consumer of holding Coffee Inventory		
Linear	β_3	0.0005*** (0.000085)
Quadratic	β_4	0.00003*** (0.000001)
Utility from Consumption		
Consumption of Soda		
Linear	γ_1	0.0065*** (0.00037)
Quadratic	γ_2	-0.0086*** (0.00039)
Consumption of Coffee		
Linear	γ_3	0.0089*** (0.00045)
Quadratic	γ_4	-0.0033*** (0.00044)
Interaction of the two Beverage-specific Consumption levels	γ_5	0.004397*** (0.00040)
	σ_ε	0.231*** (0.00065)
Log likelihood		83.38

Notes: Standard errors are in parentheses. The standard errors are computed using a bootstrapping method. The bootstrapping method involves taking normal random draws of the beverage-specific consumption shocks across consumers in order to generate new sets of beverage-specific consumption series. The model is then re-estimated for each generated beverage-specific set of consumption series, yielding a corresponding set of parameter estimates. We use this process to generate 25 new sets of parameter estimates, and use these to compute the bootstrap standard errors. *** indicates statistical significance at 1%.

6.1 Size Category Choice Elasticities

We first summarize the predicted results of $\frac{\partial Pr(y)}{\partial \omega_x}$ and present them as elasticities in Table 6. We selected four representative households that have different demand responses to a change in the expected utility of purchasing from a given size category. For each household, each entry in the table shows the percentage change in probability of choosing the specified size if the expected utility of purchasing size 1 (144 fl. oz. packaged soda) increases by 1%.

The elasticity estimates in the table reveal that all own effects are positive, i.e. $\frac{\partial Pr(x_1)}{\partial \omega_{x_1}} > 0$ for all households, suggesting that when households have higher expected utility from size 1, the probability of them purchasing from size 1 increases. In particular, the own size category elasticity estimates in the table reveal that a 1% increase in the expected utility associated with the product options in size 1 will result in an increase in the probabilities that households 1, 2, 3, and 4 purchase size 1 by 4.63%, 5.34%, 5.66%, and 5.56%, respectively. However, consistent with the possibilities we previously laid out in **Lemma 1**, it is evident that the sign of the size category cross effects vary across households.

The first household in Table 6 considers all other five size categories as substitutes to size1 since the size category cross elasticities are negative. In particular, for household 1 we see that a 1% increase in his expected utility associated with the product options in size 1 results in a decrease in the probability of this household choosing sizes 2, 3, 4, 5 and 6 by 2.50%, 2.24%, 2.40%, 2.05% and 2.28%, respectively. On the contrary, household 2 perceives all other five sizes as complements to size 1. Household 3 considers the two liter packaged size soda as a substitute size to 144 fl. oz. packaged soda products, but single-bottled 20 fl. oz. soda products as a complementary size to 144 fl. oz. packaged soda products. In addition, household 3 considers all three size categories of coffee products as substitutes for 144 fl. oz. packaged soda products.

If household 4's expected utility from purchasing 144 fl. oz. soda products increases by 1%, this increases his probability of choosing two liter soda products, 20 fl. oz. single-bottle soda products, and 407 fl. oz. coffee products by 0.031%, 0.119%, and 0.029%, respectively. As such, household 4 perceives two liter soda products, 20 fl. oz. single-bottle soda products, and 407 fl. oz. coffee products as complementary to 144 fl. oz. soda products. However, if household 4's expected utility from purchasing 144 fl. oz. soda products increases by 1%, this decreases his probability of choosing 144 fl. oz. and 288 fl. oz. coffee products by 0.106% and 0.039% , respectively. As such, household 4 perceives 144 fl. oz. and 288 fl. oz. coffee products as substitutes for 144 fl. oz. soda products.

Table 6. Size Category Choice Elasticities						
House-hold ID	%ΔProb (Soda144)/ %Δω_1 Size 1	%ΔProb(Soda 2L)/ %Δω_1 Size 2	%ΔProb(Soda 20)/ %Δω_1 Size 3	%ΔProb(Coffee 144)/% Δω_1 Size 4	%ΔProb(Coffee 407)/ %Δω_1 Size 5	%ΔProb(Coffee 288)/ %Δω_1 Size 6
1	4.625	-2.503	-2.243	-2.397	-2.054	-2.284
2	5.336	0.231	0.358	0.212	0.382	0.279
3	5.655	-0.012	0.001	-0.043	-0.020	-0.031
4	5.556	0.031	0.119	-0.106	0.029	-0.039

6.2 Product-level Price Elasticities and Merger Simulations

6.2.1 Soda Products

Some estimates of product-level own-price and cross-price elasticities of soda products are presented in Table 7. In light of these elasticity estimates, we perform a hypothetical merger between the two firms that produce the example products. The effects from merger between two soda firms are presented in Table 8. For comparison, we estimate a purely static demand model, and use it for computing demand elasticities for the same example products, and for simulating the same merger between soda firms. Results based on the purely static demand model are reported in Panel b of Table 7 and Table 8 respectively. All soda products in Table 7 and Table 8 belong to the 144 fl. oz. package size.

7 UP and A&W are soda brands owned by Dr. Pepper, while Pepsi ONE is produced by Pepsi Co. As shown in Table 7, the estimated own-price elasticity generated by the dynamic demand model for 7 UP is -2.076, meaning that a 1% increase in its price will result in a 2.076% decrease in its quantity demanded. A similar interpretation applies to the own-price elasticity estimates for the other products reported in Table 7, each estimate suggesting that forward-looking consumers have price elastic demand for soda products. Interestingly, and confirming arguments in Hendel and Nevo (2006b), the own-price elasticity estimates for these products generated from a purely static conditional logit demand model are smaller in absolute terms, suggesting that ignoring the forward-looking behavior of consumers who have the option to hold product inventories can result in less elastic demand estimates.

The mean cross-price elasticities between 7 UP and other Dr. Pepper and Pepsi products generated from the dynamic demand model are negative, suggesting complementary relationships between these products. For example, the cross-price elasticity between Dr. Pepper's 144 fl. oz.

packaged 7 UP product and PepsiCo's branded 144 fl. oz. packaged products in the market is a mean -0.0004, revealing that a 1% increase in price of this package size 7 UP product results in a decrease in quantity demand for PepsiCo's branded products of the same package size by a mean 0.0004%. In contrast, Dr. Pepper's 144 fl. oz. A&W product is revealed to be a substitute for other Dr. Pepper and PepsiCo products within this package size as indicated by the positive average cross-price elasticity estimates. In the case of PepsiCo's PEPSI ONE and PEPSI 144 fl. oz. products, evidence of both the existence of substitutes and complements can be seen in Panel a. of the table. As expected, the cross-elasticity estimates generated by the purely static conditional logit demand model are all positive, suggesting that the relevant products are all substitutes.

Table 7. Own-price & cross-price elasticities of 144 fl. oz. package size soda products in example market						
Panel a. Dynamic Demand Model						
	Dr. Pepper (7 UP)	Dr. Pepper (A&W)	Pepsi Co (PEPSI ONE)	Pepsi Co (PEPSI)	Coca Cola (SPRITE)	Coca Cola (Diet CHERRY COKE)
Own-price elasticities	-2.07577	-1.84181	-1.88234	-1.95972	-1.24878	-1.25814
Average Cross-price elasticity						
Other Dr. Pepper products	-0.00040	0.02359	-0.00034	0.01854	0.14543	0.03089
Other PepsiCo. Products	-0.00038	0.02326	-0.00036	0.01866	0.14711	0.02979
All other products in this market	-0.00043	0.02338	-0.00039	0.01886	0.15051	0.03014
Panel b. Static Demand Model						
Own-price elasticities	-1.47275	-1.45407	-1.34178	-1.3914	-0.9039	-0.9127
Average Cross-price elasticity						
Other Dr. Pepper products	6.84E-13	0.01829	1.76E-13	0.01564	0.09800	0.01651
Other PepsiCo. Products	6.90E-13	0.01843	1.78E-13	0.01579	0.09929	0.01674
All other products in this market	7.01E-13	0.01875	1.82E-13	0.01608	0.10191	0.01720

As previously discussed, and shown in equation (36), the effects on product markups resulting from a merger depend on the relationship between products of the firms that merge. Table 8 summarizes the predicted changes in product markups resulting from a hypothetical merger between Dr. Pepper and Pepsi Co. in the example market used for the information reported in Table 7. Panel a and Panel b in Table 8 show the results generated from a dynamic demand model and a static demand model, respectively.

On average, the dynamic model predicts that the hypothetical merger will result in a mean decrease of 5.3% in product markups, on net, an overall pro-competitive outcome. However, the hypothetical merger do have very different predicted markup effects on Dr. Pepper's products compared to PepsiCo. products. The price-cost markups of Dr. Pepper products are predicted to increase by a mean 56.66%, while the price-cost markups of PepsiCo. products are predicted to decrease by a mean 92.8%. In stark contrast, the static model predicts product markups will increase by a mean 85.1%, with markups on Dr. Pepper products predicted to increase by a mean 73.4%, and markups on PepsiCo. products predicted to increase by a mean 100.6%, pure anti-competitive effects, which are substantial. These contrasting predicted merger effects across the dynamic and static demand models highlight the importance of considering the dynamics, and assuming that households are forward-looking in their consumer choice behavior.

Table 8. Predicted Changes in Product Markups if Pepsi & Dr. Pepper Merged (Example market)

Panel a. Dynamic Model					
	Distinct Products	Mean	Std. Dev.	Min	Max
Merged firms	41	-5.3%	74.6%	-100.0%	58.0%
Dr. Pepper	24	56.7%	0.7%	54.5%	58.0%
Pepsi Co	17	-92.8%	3.7%	-100.0%	-89.1%
Panel b. Static Model					
	Distinct Products	Mean	Std. Dev.	Min	Max
Merged firms	42	85.1%	13.7%	73.0%	101.0%
Dr. Pepper	24	73.4%	0.6%	73.0%	75.4%
Pepsi Co	18	100.6%	0.4%	100.2%	101.0%

Note: In the dynamic model, one product has negative markup, and so excluded from the calculation.

6.2.2 Coffee Products

The qualitative results described above for soda products also apply to coffee products. Table 9 presents the own-price and cross-price elasticities of two coffee products within the 144 fl. oz. package size. The coffee brands are, Morning Joe, a brand of Starbucks, and Dunkin Donuts, a brand of The J M Smucker Co. Regarding own-price elasticities, it is apparent that coffee products are more elastic than soda products. For example, if the price of Starbucks Morning Joe 144 fl. oz. package size product increases by 1%, its quantity demand decreases by 3.18%, suggesting households are more likely to switch to other alternatives in this market. However, as we saw for soda products, it is indeed also the case for coffee products that own-price elasticity estimates generated from the purely static conditional logit demand model are smaller in absolute terms compared to the corresponding estimates generated from the dynamic demand model, suggesting again that ignoring the forward-looking behavior of consumers who have the option to hold product inventories can result in less elastic demand estimates.

Now considering the cross-price elasticity estimates for coffee products reported in Table 9. Estimates in the table reveal that a 1% price increase for Morning Joe's 144 fl. oz. package size product is predicted to increase the quantity demand of other 144 fl. oz. package size Starbucks products by a mean 0.07%; but decrease the demand for THE J M SMUCKER CO 144 fl. oz. package size coffee products by a mean 0.1%. These cross-price elasticity estimates from the dynamic demand model suggest consumers perceive Morning Joe as substitutes to other Starbucks coffee products, but complements to THE J M SMUCKER CO coffee products. In contrast, cross-price elasticity estimates generated from the static model in Panel b. suggest that all coffee products in this market are substitutes with respect to either Morning Joe or Dunkin Donuts coffee products.

Table 9. Own-price & Cross-price elasticities of 144 fl. oz. package size coffee products in example market

Panel a. Dynamic Demand Model		
	Starbucks (Morning JOE)	THE J M SMUCKER CO (DUNKIN DONUTS)
Own price elasticities	-3.18232	-2.61718
Average cross-price elasticities		
Other Starbucks products	0.07004	0.16233
Other J M SMUCKER products	-0.09955	-0.07128
All other products in this market	0.02383	0.03285
Panel b. Static Demand Model		
	Starbucks (Morning JOE)	THE J M SMUCKER CO (DUNKIN DONUTS)
Own price elasticities	-1.14247	-1.55461
Average cross-price elasticities		
Other Starbucks products	0.01025	0.01096
Other JM products	0.01036	0.01099
All other products in this market	0.01047	0.01101

Table 10 reports the predicted changes in product markups due to a hypothetical merger between Starbucks and The JM Smucker. The dynamic and static models generate very contrasting predicted effects of this hypothetical merger. The dynamic model predicts a rise in product markups by a mean 6.66%, with markups on Starbucks products predicted to fall by a mean 8.41%, while markups on The JM Smucker products predicted to increase by a mean 21.74%. However, the static model predicts an increase in markups on all coffee products, with a mean increase of 1.45%. Again, the message is clear from the predicted merger results, a failure to consider the potential complementary relationships between some pairs of products will yield misleading predictions of the merger effects.

Table 10. Change in markup if Starbucks & The JM Smucker merge (Example market)

Panel a. Dynamic Model					
	Distinct Products	Mean	Std. Dev.	Min	Max
Merged firms	4	6.663%	18.499%	-16.064%	22.106%
Starbucks	2	-8.414%	10.819%	-16.064%	-0.763%
The JM Smucker	2	21.741%	0.517%	21.375%	22.106%
Panel b. Static Model					
	Distinct Products	Mean	Std. Dev.	Min	Max
Merged firms	4	1.453%	0.263%	1.224%	1.682%
Starbucks	2	1.681%	0.001%	1.680%	1.682%
The JM Smucker	2	1.225%	0.002%	1.224%	1.226%

6.2.3 Counterfactual Merger between Soda and Coffee Firms

Using both the dynamic demand and static demand models, we simulate a hypothetical merger between Pepsi Co. and Starbucks, two separate firms that have cooperative arrangements since 1994. Regarding the merger between Keurig and Dr. Pepper that actually occurred in 2018, under the limitation of the timespan of our data, we are not able to analyze the actual merger directly from our data. Instead, we conduct a counterfactual experiment that investigates what if the merger happened during the time frame of our data.

The predicted markup changes from the two counterfactual experiments are reported in Table 11 and Table 12. In these cases, we do not observe complementary relationships between the products of the firms that hypothetically merge. As such, we expect the models will predict that the mergers increase product markups. However, it is likely that the magnitudes of the predicted markup changes will differ across the dynamic and static demand models since they generate different estimates of own-price and cross-price elasticities.

Table 11 presents the markup changes from the Keurig – Dr. Pepper merger. Our dynamic model predicts the markups of products from the merged firm will increase by a mean 1.51%. On the other hand, the static demand model predicts product markups will increase by a mean 4.76%.

Similar results and findings, reported in Table 12, are confirmed by the second hypothetical merger case. If the Pepsi company and its business partner, Starbucks coffee company, decide to merge, the dynamic model predicts that markups on their products will increase by a mean 1.45%. However, the static demand model predicts that markups on their products will increase by a mean 5.36%.

Table 11. Change in markup if Dr. Pepper and Keurig merge					
Panel a. Dynamic Model					
	Distinct Products	Mean	Std. Dev.	Min	Max
Merged firms	64	1.51%	4.34%	0	22.67%
Dr. Pepper	55	0.90%	2.18%	0	8.15%
Keurig	9	5.20%	9.91%	0	22.67%
Panel b. Static Model					
	Distinct Products	Mean	Std. Dev.	Min	Max
Merged firms	64	4.76%	7.90%	0.43%	44.99%
Dr. Pepper	55	2.36%	1.23%	0.43%	4.23%
Keurig	9	19.41%	14.14%	0.98%	44.99%

In summary, even though the dynamic demand model has not revealed complementary relationships between the products of the coffee and soda firms used for performing hypothetical merger simulations, there still exists substantial differences in the magnitudes of the predicted merger effects across the static and dynamic demand models. In case of the Keurig-Dr. Pepper hypothetical merger, which eventually occurred in 2018, the predicted mean percentage increase in price-cost markups from the static demand model is approximately 3 times as high as the predicted mean percentage increase in markups from the dynamic demand model. In case of the Pepsi-Starbucks hypothetical merger, the predicted mean percentage increase in price-cost markups from the static demand model is almost 4 times as high as the predicted mean percentage increase in markups from the dynamic demand model. However, these differences in the predicted merger effects across the models are not surprising since, for many soda and coffee products, we found clear evidence that own-price elasticity estimates generated from the purely static conditional logit demand model are smaller in absolute terms compared to the corresponding estimates generated from the dynamic demand model. A key takeaway message here is therefore clear: Ignoring the forward-looking behavior of consumers who have the option to hold product inventories can result in less elastic demand estimates, which in turn yields misleading predictions of merger effects.

Table 12. Change in markups if Pepsi & Starbucks merge					
Panel a. Dynamic Model					
	Distinct Products	Mean	Std. Dev.	Min	Max
Merged firms	153	1.45%	3.54%	0	25.83%
Pepsi Co	128	1.02%	1.88%	0	9.05%
Starbucks	25	3.62%	7.40%	0	25.83%
Panel b. Static Model					
	Distinct Products	Mean	Std. Dev.	Min	Max
Merged firms	156	5.36%	9.62%	0.29%	58.03%
Pepsi Co	131	2.13%	2.33%	0.29%	9.06%
Starbucks	25	22.27%	14.61%	0.87%	58.03%
Note: In the dynamic model, three products have negative markup, so excluded from the calculation.					

7. Conclusion

A key objective of this paper is to illustrate that, compared to static discrete choice demand models, a dynamic discrete choice demand model can better capture "complementary" type consumer choice behavior among pairs of differentiated products. Furthermore, we show that such "complementary" type consumer choice behavior captured by a dynamic demand model only renders a product pair as "empirical complements" if this aspect of their choice behavior is sufficiently strong to overturn the inherent substitutability consumers perceive between the pair of product alternatives in their choice set. It then becomes an empirical question of whether a given product pair is ultimately treated as complements or substitutes based on consumers' patterns of choice behavior.

While both static and dynamic discrete choice models can capture consumers' incentives for "complementary" type choice behavior, a static discrete choice model imposes a restriction on the size of such "complementary" type incentives that causes the inherent substitutability of products to always dominate. We show that, unlike a static discrete choice model, the forward-looking attribute of consumers in the dynamic model increases the capacity of the discrete choice model to capture "complementary" type consumer incentives such that these incentives can at times be relatively stronger than the inherent substitutability of products, ultimately yielding empirical "complementary" type relationships between these products. As such, a key methodological contribution of our paper is to illustrate how a dynamic demand model can be used to overcome this shortcoming of traditional static discrete choice models of demand.

Accurately estimating the relationship between products is crucially important for the analysis of market competition. In particular, measuring the competitive impacts of mergers crucially depend on both the type and strength of the relationship between products of rival firms, where strong complementarity between products of the merging firms can result in lower price-cost markups post-merger, which is an unattainable outcome when the products are instead substitutes. To illustrate the implications of using a dynamic demand model compared to a static demand model, we estimated each model on a sample of consumers who make purchase decisions on a menu of soda and coffee products over a two year period. We use each estimated model to simulate the effects on price-cost markups from hypothetical mergers between soda firms, between coffee firms, and between soda and coffee firms, and document substantial differences in predicted merger effects across these models.

Using the dynamic demand model, we show evidence of empirical complementary relationships between pairs of soda products, and pairs of coffee products, even though most product pairs are found to be empirical substitutes.

Regarding the simulated effects from a hypothetical merger between Dr. Pepper and PepsiCo, the dynamic model predicts a mean decrease of 5.3% in product markups, on net, an overall pro-competitive outcome. In stark contrast, the static model predicts product markups will increase by a mean 85.1%, with markups on Dr. Pepper products predicted to increase by a mean 73.4%, and markups on PepsiCo products predicted to increase by a mean 100.6%, pure anti-competitive effects, which are substantial.

Regarding the simulated effects from a hypothetical merger between Starbucks and The JM Smucker Co., the dynamic model predicts a rise in product markups by a mean 6.66%, with markups on Starbucks products predicted to fall by a mean 8.41%, while markups on The JM

Smucker products predicted to increase by a mean 21.74%. However, the static model predicts an increase in markups on all coffee products, with a mean increase of 1.45%.

Regarding the simulated effects from a hypothetical merger between soda and coffee firms, for the Keurig-Dr. Pepper hypothetical merger, which eventually occurred in 2018, the predicted mean percentage increase in price-cost markups from the static demand model is approximately 3 times as high as the predicted mean percentage increase in markups from the dynamic demand model. In case of the Pepsi-Starbucks hypothetical merger, the predicted mean percentage increase in price-cost markups from the static demand model is almost 4 times as high as the predicted mean percentage increase in markups from the dynamic demand model. These differences in the predicted merger effects across the models are not surprising since, for many soda and coffee products, we found clear evidence that own-price elasticity estimates generated from the purely static conditional logit demand model are smaller in absolute terms compared to the corresponding estimates generated from the dynamic demand model.

In summary, a key message from this research is that ignoring the forward-looking behavior of consumers who have the option to hold product inventories can result in inaccurate estimates of demand elasticities, which in turn yields misleading predictions of merger effects. But is there an even broader takeaway for analyzing certain types of mergers?

To the best of our knowledge, this study is the first to analyze a model of demand that jointly incorporates both coffee and soda caffeinated beverages, which is necessary to study existing and potential partnerships between firms across these two product categories. As such, our study fills this gap in the literature. However, a broader takeaway is that future research may use the demand framework in this paper to explore consumers' choice behavior across products of different categories, which will better facilitate studying partnerships of varying degrees between firms often delineate as being in different industries. The demand framework presented in this paper is particularly useful for this endeavor since we have shown how the model can be used to flexibly measure the strength of both substitute and complementary relationships between products of firms often delineate as being in different industries.

Appendix

Table A1: Model Estimates of Term A_1 in Equation (26) for a few States in our Data Sample

Values of Term A_1	-0.029	-0.032	-0.247	-0.198	0.038	0.007	0.026	0.031

References

Berry, Steven T. 1994. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*, Vol. 25, No. 2, Summer 1994, pp. 242-262.

Berry, Steven T., James Levinsohn and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica*, Vol. 63, No. 4, July 1995, pp. 841-890.

Björnerstedt, Jonas and Frank Verboven. 2016. "Does Merger Simulation Work? Evidence from the Swedish Analgesics Market." *American Economic Journal: Applied Economics* 8 (3): 125-164.

Bonnet, Céline and Sofia Berto Villas-Boas. 2016. "An Analysis of Asymmetric Consumer Price Responses and Asymmetric Cost Pass-through in the French Coffee Market." *European Review of Agricultural Economics* 43 (5): 781-804.

Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela. 2008. "Database paper—The IRI marketing data set." *Marketing Science, 27*(4), pp.745-748.

Crawford, Gregory and Ali Yurukoglu. 2012. "The Welfare Effects of Bundeling in Multichannel Television Markets," *American Economic Review*, 102 (2): 643-685.

Crawford, Gregory, Robin Lee, Michael Whinston and Ali Yurukoglu. 2018. "The Welfare Effects of Vertical Integration in Multichannel Television Markets," *Econometrica*, Vol. 86, No. 3, May 2018: 891- 954.

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

Dubé, Jean-Pierre. 2005. "Product Differentiation and Mergers in the Carbonated Soft Drink Industry." *Journal of Economics & Management Strategy* 14 (4): 879-904.

Dubois, Pierre and Thierry Magnac. 2015. "Consumer Demand with Unobserved Stockpiling and Intertemporal Price Discrimination." Working Paper, Toulouse School of Economics.

Erdem, Tülin, Susumu Imai and Michael P. Keane. 2003. "Brand and Quantity Choice Dynamics Under Price Uncertainty." *Quantitative Marketing and Economics* 1 (1): 5-64.

Gayle, Philip G. and Indika, Nuwan. 2020. "Does vertical integration in the U.S. Carbonated Soft Drink industry lead to anticompetitive effects?" Manuscript, *Kansas State University*.

Gayle, Philip G. and Lin, Ying. 2020. “How Much Do Consumers Value Single-cup Coffee Brew Technology? Assessing Market Impacts of Single-cup Brew Technology on the US Brew-At-Home Coffee Market,” Manuscript, *Kansas State University*.

Gayle, Philip G. and Xie, Xin. 2018. "Entry Deterrence and Strategic Alliances." *Economic Inquiry* 56 (3): 1898-1924.

Gowrisankaran, Gautam and Marc Rysman. 2012. "Dynamics of Consumer Demand for New Durable Goods." *Journal of Political Economy* 120 (6): 1173-1219.

Hartmann, Wesley R. and Harikesh S. Nair. 2010. "Retail Competition and the Dynamics of Demand for Tied Goods." *Marketing Science* 29 (2): 366-386.

Hendel, Igal and Aviv Nevo. 2006a. "Sales and Consumer Inventory." *The RAND Journal of Economics* 37(3): 543-561

Hendel, Igal and Aviv Nevo. 2006b. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica* 74 (6): 1637-1673.

Hendel, Igal and Aviv Nevo. 2013. "Intertemporal Price Discrimination in Storable Goods Markets." *The American Economic Review* 103 (7): 2722-2751.

Ho, Kate and Robin Lee. 2017. “Insurer Competition in Health Care Markets,” *Econometrica*, Vol. 85, No. 2, March 2017: 379-417.

Huang, Yufeng. 2019. "Learning by Doing and the Demand for Advanced Products." *Marketing Science* 38 (1): 107-128.

Ivaldi, Marc and Frank Verboven. 2005. "Quantifying the Effects from Horizontal Mergers in European Competition Policy." *International Journal of Industrial Organization* 23 (9-10): 669-691.

Lee, Robin S. 2013. "Vertical Integration and Exclusivity in Platform and Two-Sided Markets." *The American Economic Review* 103 (7): 2960-3000.

Lopez, Rigoberto A. and Kristen L. Fantuzzi. 2012. "Demand for Carbonated Soft Drinks: Implications for Obesity Policy." *Applied Economics* 44 (22): 2859-2865.

Martin, Stephen. 2009. "Microfoundations for the Linear Demand Product Differentiation Model, with Applications," Purdue University Economics Working Papers 1221, Department of Economics, Purdue University.

McManus, Brian. 2007. "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee." *RAND Journal of Economics* 38 (2): 512-532.

Melnikov, Oleg. 2013. "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market." *Economic Inquiry* 51 (2): 1277-1298.

Nevo, Aviv. 2000. "Mergers with Differentiated Products: The Case of Ready-to-Eat Cereal." The RAND Journal of Economics 31 (3): 395-421

Osborne, Matthew. 2018. "Approximating the Cost-of-Living Index for a Storable Good." *American Economic Journal. Microeconomics* 10 (2): 286-314.

Pinkse, Joris and Margaret E. Slade. "Mergers, Brand Competition, and the Price of a Pint." *European Economics Review* 48 (2004): 617-643.

Train, Kenneth E. 2009. "Discrete Choice Methods with Simulation," Second edition, *Cambridge University Press*.

Villas-Boas, Sofia Berto. 2007. "Using Retail Data for Upstream Merger Analysis." *Journal of Competition Law and Economics* 3 (4): 689-715.

Wang, Emily Yucai. 2015. "The Impact of Soda Taxes on Consumer Welfare: Implications of Storability and Taste Heterogeneity." *The RAND Journal of Economics* 46 (2): 409-441.