Mergers Between Multi-Product Firms With Endogenous Variety: Theory and an Application to the RTE-Cereal Industry*

Francisco Garrido†

Georgetown University

January 2, 2020

Abstract

When multi-product firms endogenously choose their product variety, mergers can have welfare effects that go well beyond their immediate impact through prices. The structural models needed to quantify these effects, however, quickly become computationally intractable as the number of products grows. In this paper, I propose a dynamic structural model of multi-product firms and endogenous product variety that bypasses this dimensionality problem. I show that, under a nested logit demand assumption, firms introduce/remove products in a pre-determined order. As a result, the number of products each firm supplies is a sufficient statistic for the identity of the products themselves and, therefore, for all market outcomes and firms' continuation strategies. I use this result to argue that the strategies and states that need to be considered to solve the model are a small fraction of the whole strategy set and state space, making it possible to estimate the model using standard techniques in structural econometrics. I apply the model to the RTE-Cereal Industry and use it to simulate a hypothetical merger between two large players. Results show that ignoring endogenous product variety leads to a 30% overestimation of the post-merger number of products and a 19% underestimation of consumer welfare loss.

Keywords: Product Choice, Merger Simulation, Dynamic Discrete Games, Structural Estimation.

JEL Classification: D22, D43, L11, L13, L66

*I would like to thank my advisors John Rust, Nathan Miller and Roger Lagunoff for constant guidance and encouragement. All errors are my own. Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

†e-mail: fg359@georgetown.edu

1 Introduction

When firms adjust their product variety to the post-merger market structure, mergers between multi-product firms can have welfare effects beyond their immediate impact through increased prices. Recognizing this fact, the Department of Justice (DOJ) has opposed multiple mergers citing concerns of “fewer choices for consumers,” as in the Danone/WhiteWave Foods case in the raw organic milk and organic yogurt markets,[1] the ABI/SABMiller case in the beer industry,[2] and the AT&T/T-Mobile case in the telecommunications sector.[3] Ideally, merger evaluations should quantify the welfare importance of this effect. Doing so, however, may be extremely challenging, as the structural models needed to do this quickly become computationally intractable as the number of products grows.

This paper proposes and estimates a dynamic model of multi-product firms with endogenous product variety and price competition. The model allows firms to choose from a large pool of horizontally differentiated products, while remaining computationally tractable and can distinguish short-term welfare effects of mergers that operate through increased prices, from long-term effects that work through endogenous changes in product variety. Furthermore, the model is flexible in that, unlike to most research estimating structural games, it is not tailored to a specific industry; thus, it can accommodate different markets with relative ease. I apply the model to the Ready-To-Eat(RTE) cereal industry and use it to examine a hypothetical merger between two large players. I find that the merged entity removes roughly 30% of its products, which results in predicted markups 2% lower for the merged entity and 1% higher for the non-merged firms, as compared to a case where firms do not change their product choices post-merger. All in all, ignoring product variety endogeneity leads to a 19% underestimation of consumer welfare loss.

This paper makes theoretical, methodological and empirical contributions. The theoretical contribution comes from proposing a tractable model of multi-product firms with endogenous product variety, that can be used to study the interplay between market structure and product variety, and how these affect social welfare. The model can be thought of as a generalization of Anderson and de Palma (1992) who study a product variety under a nested logit demand model with symmetric firms and products. The main methodological contribution comes from stating a novel result that bypasses the dimensionality problem that afflicts structural models of endogenous product variety. Under this result, both the state space and strategy sets grow linearly, rather than exponentially, in the number of products available to each firm. Finally, this paper contributes to the empirical literature on merger evaluation, by providing a framework under which post-merger adjustments to product variety can be accounted for in welfare calculations, and showing in an application to the RTE-Cereal industry that this effects are of sizable economic importance.

To do this, I build on the theoretical contribution of Nocke and Schutz (2018), who study a static oligopolistic pricing model with exogenous product variety. They show that, under nested

[1] Federal Trade Commision and Department of Justice (2017)
[2] Federal Trade Commision and Department of Justice (2015)
[3] Federal Trade Commision and Department of Justice (2011)

logit demand, firms' profits depend on their products through a one-dimensional index, i.e., *product type*, that summarizes its demand and marginal cost. I show that the product type can be interpreted as a profitability index, so that if a product has a larger type than another, then the former will always contribute more to profits, regardless of what other products there are in the market.

Using this intuition, I argue that firms introduce/remove products in a predetermined order, from most to least profitable. Thus, in equilibrium, a firm's *bundle size*, i.e., the number of products it supplies, is a sufficient statistic for the composition of its *product bundle*, i.e., the set of products that a firm supplies. This result allows me to model firms as directly choosing -and keeping track of- bundle sizes rather than product bundles, thereby reducing strategies and states that need to be considered to solve the model, to a small subset of the whole strategy sets and state space. With this result, the standard toolkit for dynamic structural estimation tools can be used to solve and estimate the model, such as the ones proposed by Rust (1987), Aguirregabiria and Mira (2007), or Bajari et al. (2007).

The Nested Logit assumption, may seem costly, as it imposes significant restrictions on substitution patterns, and rules out more flexible demand models such as the popular random coefficients demand model proposed by Berry et al. (1995). This shortcoming, however, can be mitigated by imposing intricate nesting structures. In particular, this framework can accommodate any nesting structure in which all products of a firm lie within the same nest. Such structures can capture between-firm substitution patterns that deviate significantly from the ones imposed by the well-known Independence of Irrelevant Alternatives (IIA) property.[4]

The RTE-cereal industry is an excellent candidate on which to apply this model for several reasons. First, because it is a concentrated industry in which horizontal product differentiation is a prominent feature, with four longstanding dominant players who jointly represent around 75% of total sales. The second reason is that many researchers have looked at this industry (such as Nevo (2001), Backus et al. (2018)), and have found that, in line with my modeling assumptions, a static Nash-Bertrand equilibrium gives a reasonable approximation to pricing game. Third, because there is considerable variation in product bundles between markets and over time, and this helps identify the parameters of the dynamic models. And finally, because it highlights the empirical scope of my model since firms in this industry have many more products than other structural models of product choice can manage.

The past couple of decades have seen substantial growth in the applications of structural models in IO. This agenda, however, has focused mainly on low dimensional environments, such as the static entry models (Ciliberto et al. (2018), Li et al. (2018)), entry/exit models (Aguirregabiria and Ho (2010), Benkard et al. (2014)), offshoring (Igami (2015)), product introduction and exit (Hitsch (2006)), and dynamic pricing (Williams (2017), Rust et al. (2018)), to name a few. In contrast, applications to environments with endogenous product variety, where both the strategy sets and the state space grow exponentially with the number of products, are much less common.

The few examples of structural models studying with endogenous product variety, invariably

[4] See McFadden et al. (1973)

highlight the importance of bringing this dimension into the market analysis. Draganska et al. (2009), for example, document the importance of the strategic incentives faced by duopolistic ice-cream manufacturers when making product line decisions for the vanilla-flavored ice cream.[5] Nosko (2010) shows, in a study of the CPU market, that multi-product firms can use quality to "isolate competition to an undesirable area of the product space". Wollmann (2018) uses a model where firms endogenously choose how many of three predetermined product types to produce. He uses his model to study the effectiveness of the truck bailouts during the great recession, finding that a world without bailouts would have seen little disruption to the trucking industry, as surviving firms would have stepped in to take the place of the ones that went bankrupt.

In a dynamic setting, both Sweeting (2013) and Jeziorski (2014) propose models of product re-positioning for the radio broadcasting industry. The former studies the effect of mergers on radio positioning, finding that merged firms make their stations more differentiated to reduce cannibalization. Variety, however, does not increase because the re-positioned products are now closer to those of competing firms. The latter estimates cost efficiencies from mergers in a model where firms endogenously choose radio formats.

Closer to this paper, Hitsch (2006) proposes an empirical model of product launch and exit for Bayesian producers in the presence of demand uncertainty. He finds that, for the RTE cereal industry, the upside risk of introducing new products is significant; hence, producers benefit from experimentation, i.e., launching many new products, even if many have a low expected profit. The model in Hitsch (2006) is, in a sense, on the opposing extreme of my model. On the one hand, one could argue that his model is more realistic in that it deals with demand uncertainty, experimentation, and learning. On the other hand, one could argue that my model is more realistic because firms optimize the entirety of their bundles, rather than only one product, and consider the strategic effect of their actions on their competitor's bundles, not just their prices.

Other studies using dynamic structural models, working on issues unrelated to multi-product firms, have also bypassed the curse of dimensionality using some aggregation of the state space. This paper differs from them in that the dimensionality reduction is micro-founded, as opposed to a modeling assumption, and is therefore robust to policy changes and informed in estimation by the micro-data (in this case, demand and marginal cost estimations). Benkard et al. (2015) and Ifrach and Weintraub (2016), for example, propose equilibrium concepts that assume that all players keep track of the state of large firms, but only some descriptive moments of the state of fringe firms. Closer to this paper, both Hendel and Nevo (2006), in the context of consumers' storable good inventory problem, and Barwick et al. (2019), studying the effect of China's industrial policy on the market structure of the shipbuilding industry, assume that firms only keep track of a one-dimensional logit denominator index. In both cases, the assumption is that conditional on the logit denominator other relevant exogenous variables are serially uncorrelated. Finally, Nevo and Rossi (2008) also take advantage of the aggregative property of the logit demand to reduce the computational burden

[5] The model in Draganska et al. (2009) is duopolistic in that only two firms are active during the product selection stage of the game. During the pricing stage, however, they consider the full extent of product variety observed in the market.

of the model proposed by Pakes and McGuire (1994). They focus on investment in product quality, with fixed product bundles.

As in Wollmann (2018) and Mazzeo et al. (2018), I assume that firms choose their product bundles from a pre-existing *product menu*. Unlike them, however, I focus on a dynamic model, which allows me to estimate how long firms will take to make their extensive margin adjustments. This issue has shown to be particularly important, as the adjustment can take a considerable amount of time. In an entry-exit model for the ready-mix concrete industry, for example, Collard-Wexler (2013) finds that after a merger from duopoly to monopoly, new entry takes between nine and ten years, a substantial period for the temporary monopolist to exploit its market power. Similar findings, though less extreme, are documented by Benkard et al. (2014).

In this line, I find that the merged entity takes roughly two and a half years to removes 30% of its products. During this time, the merged entity continuously remove products from its bundle, first at a rapid rate, and later at a slower pace. Non-merged firms, in contrast, do not appear to respond over the extensive margin, but they continuously increase prices as the merged firm cuts back on its product bundle.

The dynamic nature of the model also allows me to identify strategic incentives that go beyond what a static model could produce. For example, I run counterfactuals experiments to examine the preemptive product introduction hypothesis first proposed by Schmalensee (1978) for the RTE-cereal industry. In that paper, he argued that the dominant firms in this market expanded their product offering beyond what is statically profitable in order to foreclose product introduction by smaller competitors. I test this by simulating an *open-loop equilibrium*[6] in which small firms commit to a fixed path of product bundles. In an open-loop equilibrium, large firms cannot influence the behavior of smaller competitors; thus, if the deterrence motive is an important determinant of their behavior, we should see significantly fewer products from large firms under the open-loop equilibrium. Simulation results do not seem to support this hypothesis.[7]

This paper also differs from previous work on structural games in that the model proposed is less idiosyncratic, and therefore can be easily transferred to other industries. Additionally, the model can be adjusted to accommodate product development, innovation, or firm entry and exit, and can be used to study interesting economic topics like the effect of market structure on the creation of new products. Furthermore, the model can naturally account for cost savings arising from mergers, in the spirit of Perry and Porter (1985) and McAfee and Williams (1992), making it possible to include an educated guess of the potential efficiency gains from mergers in counterfactual simulations.

To avoid the complications that multiple equilibria impose on estimation and counterfactual simulation, I follow Igami and Uetake (2019) in assuming that firms alternate moves and that the game has a finite horizon. Both in their paper and this one, alternate moves and finite time

[6]See Fudenberg and Tirole (1991).

[7]It should be noted that the context in which Schmalensee (1978) suggested this hypothesis was different from the current state of the market. The same four firms that dominated the market then do so today. But today there is significant product offering by medium and small-sized competitors (roughly half of the products *are not* offered by one of the large firms), and the introduction of store-brand varieties has increased competition in the market.

greatly simplify the estimation procedure as they guarantee equilibrium uniqueness and make pseudo maximum likelihood estimation feasible through a Nested Fixed Point (NFXP) algorithm (Rust (1987)). The alternate move assumption is relatively innocuous in this setting. The deterrent effect of introducing one product earlier than my competitors is relatively mild, more so in a game with long time-horizon, since the advantage of having the first move dissipates as the game progresses.

The paper is organized as follows. In section 2, I introduce the pricing model, which serves as the stage game of the dynamic model. In section 3, I extend this model by endogenizing the supply of varieties in a two-stage context. Here I derive the aggregative result that allows me to collapse the state and strategy spaces. In section 4, I introduce the dynamic game that I will be taking to the data, and show that the aggregative result from section 3 extends to a dynamic setting. Section 5 discusses the data. Sections 6 and 7 discuss parametrization, introduce the econometric assumptions of the model, and present the estimation results. Section 8 presents the results of counterfactual experiments. And section 9 concludes.

2 Pricing Game

The model consists of a static game of price competition, embedded in a dynamic discrete-choice game of product variety choice. The pricing model, which takes products as given, is based on Nocke and Schutz (2018). Throughout the paper, I refer to the set of products a firm supplies as its *product bundle*, and to the number of products it supplies as its *bundle size*. I begin presenting the static pricing game, for exogenous product bundles, and characterize the optimal markup and the Nash-Bertrand equilibrium conditions. The main result of this section is that, in the pricing equilibrium, any product bundle can be represented by a one-dimensional index that summarizes its competitive and welfare properties.

2.1 Differentiated Product Pricing Game - Common Markup

Each market k, at each time q, is populated by continuum of potential consumers each of which buys at most one product (during the rest of this section I omit both the k and q indices, noting that all parameters may be market-time dependent). Each firm $f = 1, \dots, F$ supplies a predetermined *product bundle*, N^f, and no product is supplied by more than one firm, i.e. $N^f \cap N^g = \varnothing$ for all $f \neq g$. Firm f's' *bundle size* is denoted by $|N^f|$, and with some abuse of notation I use F to represent both the set and the number of firms.

Consumers behave according to a discrete choice[8] random utility model with Generalized Extreme

[8]In Nocke and Schutz (2018), consumer behavior is described by a discrete-continuous choice process, rather than the discrete choice framework I use here. All the results in this paper, however, hold for the discrete-continuous case. I use the discrete choice framework because it is the default demand model in Industrial Organization, and it better aligns with my empirical application. Furthermore, results in Nocke and Schutz (2018) extend to the whole family of demand functions that exhibit Independence of Irrelevant Alternatives. How far the results of this paper generalize to a broader family of demand schedules is important work that is left for future research.

Value (GEV) taste shocks. By choosing product $j \in N^f$, consumer i gets a utility of:

$$u_{ij}^f = \alpha_j^f - \lambda p_j^f + \nu_{ij}^f \quad (1)$$

where α_j^f is a (possibly market and time specific) scalar controlling the mean utility generated by product j, λ is a positive price sensitivity parameter, and ν_{ij}^f is an idiosyncratic GEV distributed taste shock. As is standard in discrete choice models, the value of the outside option, $f = 0$, is normalized to $u_i^0 = \nu_i^0$, and the scale parameter of the GEV shock is normalized to 1.

Firm f produces product $j \in N^f$ at a constant marginal cost, c_j^f, and, for reasons that will be clear later, chooses markups $(m_j^f)_{j \in N^f}$, rather than prices, to maximize profits.[9] Using the identity $p_j^f \equiv c_j^f + m_j^f$, I rewrite the consumer's utility as:

$$u_{ij}^f = \tau_j^f - \lambda m_j^f + \nu_{ij}^f \quad (2)$$

where

$$\tau_j^f = \alpha_j^f - \lambda c_j^f \quad (3)$$

is the *product type* of $j \in N^f$, which measures the surplus that j generates when it is priced at marginal cost, i.e., $m_j^f = 0$.

Alternatives are classified into a three-level nesting structure, as described in figure 1. The first nest represents the choice between buying a product and the outside option, i.e., leaving the market. The second level consists of F nests, one for each firm, where nest f contains all of f's products. Finally, the third level represents the choice between all of f's products. Parameters $\rho \in [0, 1]$ and $\gamma^f \in [0, \rho]$ for all $f \in F$, represent the nesting parameters of the first level nest and each second level nest respectively. If $\rho = 0$, there is no extreme value shock in the upper level and all consumers choose whichever alternative, buy or not buy, provides the larger expected value. If $\rho = 1$, the upper-level nest reduces to a standard multinomial logit.

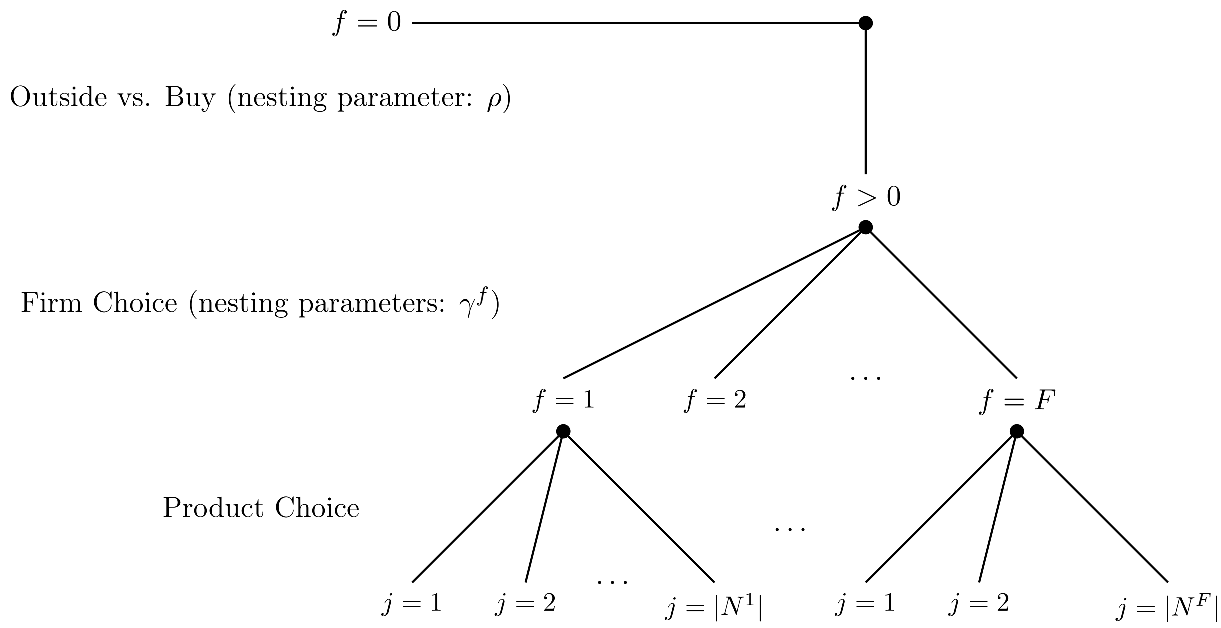
Similarly, if $\gamma^f = 0$, consumers buy from the firm that provides largest expected value, whereas if $\gamma^f = 1$ substitution between f's products and its competitors' nests reduces to a multinomial logit. For notational ease, in what follows, I suppress the super-index f, noting that each j belongs to one, and only one, N^f.

By the nested logit assumption, the market share of product j can be decomposed into three probabilities, each of which is the result of a multinomial logit decision process by consumers. First, the probability of choosing product $j \in N^f$ conditional on buying from N^f:

$$s_{j|f} = \frac{e^{(\tau_j - \lambda m_j)/\gamma^f}}{e^{h^f}} \quad (4)$$

[9]Note that that the choosing markups, $(m_j^f)_{j \in N^f}$, or prices, $(p_j^f)_{j \in N^f}$ are equivalent, as the two are tied by the identity: $p_j^f \equiv c_j^f + m_j^f$.

Figure 1: Nesting Structure of Consumer Choice



where h^f is the *firm inclusive value* of firm f, which can be roughly interpreted as the mean utility of randomly choosing a product produced by f, and it is equal to:

$$h^f \equiv \log\left(\sum_{j\in N^f} e^{(\tau_j-\lambda m_j)/\gamma^f}\right) \tag{5}$$

Second, the probability that the consumer endorses firm f conditional on actual buying a product:

$$s^{f|F} = \frac{e^{\frac{\gamma^f}{\rho}h^f}}{e^H} \tag{6}$$

where H is the *market inclusive value*, which can be roughly interpreted as the utility a consumer can expect from randomly buying a product, i.e., without knowledge of her taste shocks, and is equal to:

$$H \equiv \log\left(\sum_{f\in F} e^{\frac{\gamma^f}{\rho}h^f}\right) \tag{7}$$

And third, the probability of buying a good at all, i.e., not choosing the outside option:

$$s^F = \frac{e^{\rho H}}{1+e^{\rho H}} \tag{8}$$

Putting these together, good j's market share is:

$$s_j\left(m_j, h^f, H\right) = s_{j|f} \times s^{f|F} \times s^F \tag{9}$$

Note that both h^f and H depend on markups and, therefore, are endogenously determined in equilibrium. Also, note that s_j depends on the markups of other products only through h^f and H. Finally, consumer surplus corresponds to McFadden's Social Surplus Function, which is deeply linked to H through expression (10) below:

$$CS(H) \equiv \lambda^{-1}(\varphi + \log(1 + e^{\rho H})) \tag{10}$$

where $\varphi \approx 0.577$ is the Euler–Mascheroni constant.

Firms choose a markup schedule, $(m_j)_{j\in N^f}$, to maximize variable profits:

$$\sum_{j\in N^f} m_j \times s_j(m_j, h^f, H) \tag{11}$$

Under nested logit demand and constant marginal cost it can be shown that producers will charge a *common absolute markup* for all their products, i.e. $m^f \equiv p_j - c_j = p_l - c_l$ for all $j, l \in N^f$. To see why this is the case, consider a two product monopolist example with $\gamma^f = \rho = \lambda = 1$, where each product has a product type of $\tau_j = \alpha_j - c_j$. In this setting, demand for product $j = 1, 2$ is:

$$s_j = \frac{e^{\tau_j - m_j}}{1 + e^{\tau_1 - m_1} + e^{\tau_2 - m_2}}$$

The monopolist chooses markups to maximize profits $m_1 s_1 + m_2 s_2$. Taking the first-order condition with respect to m_1:

$$s_1 - m_1 s_1(1 - s_1) + m_2 s_2 s_1 = 0$$

Rearranging terms gives:

$$m_1 = 1 + m_1 s_1 + m_2 s_2$$

from which the common markup result immediately follows.

Back to the general case, there are three immediate consequences from the common markup result. First, firm's profit maximization problem can be expressed as a single dimensional optimization problem with respect to the markup, m^f. Second, since all products earn firms the same markup, firms do not care which particular products they sell, only the aggregate quantity. In other words, $s_{j|f}$ gets "integrated out" of the firms profits, as the sum of $s^f_{j|f}$ over all $j \in N^f$ is by definition equal to 1. And third, the firm inclusive value defined in (5) becomes:

$$h^f = \tau^{N^f} - \frac{\lambda}{\gamma^f} m^f \tag{12}$$

where τ^{N^f} is the *bundle type*, which, like the product type τ_j, can be interpreted as the surplus generated by bundle N^f when prices are equal to marginal cost, $m^f = 0$, and corresponds to:

$$\tau^{N^f} = \log \left(\sum_{j \in N^f} e^{\tau_j / \gamma^f} \right) \tag{13}$$

Note how that in equation (12) the firm inclusive value under common markup in (12), resembles non-random component of the utility function in (2). In (12), τ^{N^f} represents the surplus that a consumer can expect to get by choosing a product of firm f at random when $m^f = 0$, which is then adjusted by the f's markup multiplied by the price sensitivity λ/γ^f. With these elements in hand, I write firms profits as:

$$\pi^f(m^f, H; N^f) \quad = \quad m^f \underbrace{\frac{e^{(\gamma^f \tau^f - \lambda m^f)/\rho}}{e^H}}_{s^{f|F}} \underbrace{\frac{e^{\rho H}}{1 + e^{\rho H}}}_{s^F} \tag{14}$$

where $s^{f|F}$ is the share of firm f among all product bought, and s^F is share of customers who buy any product, i.e., do not choose the outside option, as defined in (6) and (8). The optimal markup that results from the first-order condition, $m^f(H)$, is defined implicitly as the solution to:

$$\frac{\lambda}{\rho} m^f = 1 + \frac{\lambda}{\rho} m^f \underbrace{\frac{e^{(\gamma^f \tau^{N^f} - \lambda m^f)/\rho}}{e^H}}_{s^f} \underbrace{\left(1 - \frac{\rho}{1 + e^{\rho H}}\right)}_{1 - \rho s^0} \tag{15}$$

where $s^0 = 1 - s^F$, is the share of the outside option. Note that, for a fixed H, the left-hand side is a 45 degree line with respect to $\frac{\lambda}{\rho} m^f$, whereas the right-hand side is strictly decreasing for $\frac{\lambda}{\rho} m^f > 1$, hence, the solution to (15), $m^f(H)$, is unique.[10]

The solution to (15), $m^f(H)$, can be interpreted as f's best response to H, which summarizes the f's competitors' behavior. Given a profile of optimal markup functions, $(m^f(H))_{f \in F}$, the Nash-Bertrand equilibrium is pinned down by the equilibrium market inclusive value, which under common markup is:

$$H = \log \left(\sum_{f \in F} e^{(\gamma^f \tau^{N^f} - \lambda m^f)/\rho} \right) \tag{16}$$

Intuitively, in equilibrium, H has to induce markups that generate an inclusive value of H, as per expression (16). In this line, H, has a dual interpretation: it is a measure of the consumer welfare generated by the market (as per equation (10)), and a measure of its competitiveness. On the supply end of the model, equation (15), a higher H implies more competition and hence lower markups. On the demand side, a higher H implies a more attractive market for consumers and a lower share of the outside option, i.e. higher $s^F = e^{\rho H}/(1 + e^{\rho H})$.

[10]For more details see Appendix A

With these elements, I define the Nash-Bertrand equilibrium of this game in terms of equations (15), which represents profit-maximizing markup given H, and equation (16).

Definition 2.1 (Nash-Bertrand Equilibrium)**.** A Nash-Bertrand Equilibrium for this game is a tuple of markup profile and market inclusive value, $((m^f)_{f\in F}, H)$, such that:

1. Given H, m^f solves (15) for all $f \in F$.

2. H is consistent with the markup profile $(m^f)_{f\in F}$, as per equation (15).

Note that the equilibrium H is pinned down by a single equation obtained by plugging the optimal markups $(m^f(H))_{f\in F}$, which solve (15), into the market inclusive value definition in (16). This is:

$$H = \log\left(\sum_{f\in F} e^{(\gamma^f \tau^{N^f} - \lambda m^f(H))\rho}\right) \tag{17}$$

The left-hand side of this equation corresponds to the 45 degree line in with respect to H. As for the right-hand side, it can be shown that it is a strictly positive and increasing function of H, with a slope in the interval $(0,1)$.[11] Thus, equation (17) has a unique solution, implying that the game has a unique equilibrium.

The common markup property and the aggregative nature of the game through the inclusive value, greatly reduce the computational burden of solving the Nash-Bertrand equilibrium. Additionally, as I will discuss later, this simplifies the estimation of the dynamic model, and the calculation of counterfactual experiments. In particular, regarding the Nash-Bertrand equilibrium, the firm's problem is reduced from finding a multi-dimensional optimal pricing schedule, $(p_j)_{j\in N^f}$, to finding a one-dimensional markup, m^f. In consequence, the problem of finding the Nash Bertrand equilibrium is reduced from a system of $\sum_{f\in F}|N^f|$ simultaneous equations, to one of $F+1$ equations (one markup per firm plus the equilibrium market inclusive value). Furthermore, since the equations that determine the equilibrium interact only through equation (17), the cheapest approach, computationally, to finding the equilibrium is to guess a value for H, solve the F one-dimensional equations that determine each optimal markup, $m^f(H)$, use this markups to get a new guess for equation H as per equation (16), and repeat till convergence.

Finally, note that the equilibrium inclusive value depends on each N^f only through the one-dimensional bundle type, τ^{N^f}. Furthermore, given H, f's optimal markup does not depend on its competitor's bundles and depends on its own product bundle only through τ^{N^f} too. Similar conclusions hold for market shares, profits, and consumer surplus. As a result, comparative statics exercises in which a firm unilaterally changes its product bundle are equivalent to the study of changes in τ^{N^f} on equilibrium outcomes. In section 3, I take advantage of this observation to show that f's profits are strictly increasing in τ^{N^f}, and narrow down the set of bundles that firms will play in equilibrium.

[11] See appendix B

3 Two-stage Game

Here I extend the analysis by endogenizing firms' product bundles. To do this, I propose a two-stage game of product choice, which I extend to a dynamic context in section 4. The goal of this section is to show that, under mild additional assumptions, each firm has a unique strictly dominant bundle for each bundle size, so, without loss of generality, firm's strategy space can be reduced to the space of bundle sizes, and that the bundle size is a sufficient statistic for the composition of the bundle.

3.1 Setup, Sub-Game Perfect Equilibrium, and Dominant Strategies

Consider a two-stage game of product choice similar to the one considered in Anderson and de Palma (1992). In the first stage, firms simultaneously choose a product bundle N^f, from a countable set $\mathcal{N}^f$, and pay a fixed cost:

$$FC^f(N^f) > 0 \tag{18}$$

for all $N^f \subset \mathcal{N}^f$.

After choosing N^f, firms observe the *bundle profile* $(N^f)_{f \in F}$ and compete in prices as described in section 2. Without loss of generality, I restrict my attention to second stage pricing strategies that satisfy the common markup assumption. With these elements in place, I can define the strategy sets and payoffs of this game and the equilibrium concept.

Definition 3.1 (Two-Stage Game, Strategies, and Payoffs)**.** A strategy for firm f in this game is a tuple $(N^f, m^f(N))$, representing a product bundle for the first stage, N^f, and a common markup, m^f, for every possible bundle profile $N = (N^f, N^{-f})$, i.e., every possible first stage outcome.

The payoff generated by strategy profile $(N^f, m^f(N))_{f \in F}$ is

$$\pi^f(m^f(N), H; N) - FC^f(N^f) \tag{19}$$

where π^f and H are as defined in (14) and (16) respectively.

Definition 3.2 (Sub-game Perfect Equilibrium)**.** A sub-game perfect equilibrium for this game is a strategy profile and a market inclusive value, $\big((N^f, m^f(N))_{f \in F}, H(N)\big)$ such that:

- The tuple $(m^f(N)_{f \in F}, H(N))$ is an equilibrium of the Nash-Bertrand pricing game, as per definition 2.1, for all $N = (N^f, N^{-f})$.
- N^f maximizes:
$$\pi^f(m^f(N), H(N); N) - FC^f(N^f)$$
given N^{-f}, for all $f \in F$.

A brute force search of the equilibria of this game requires solving all possible second stage pricing games, of which there are $2^{\sum_f |\mathcal{N}^f|}$, computing profits, and going over each of them looking for mutually optimal first stage bundles. Even for a modest number of firms and products, this can be computationally infeasible. To bypass this problem, I take advantage of the particular structure of the second stage game to show that, under mild additional assumptions, for any firm f and bundle size $n^f \in \mathbb{N}$, there is a bundle that strictly dominates all other bundles of size n^f. By focusing on the dominant bundles, I effectively reduce each firm's strategy set from $2^{|\mathcal{N}^f|}$ to $|\mathcal{N}^f|$ alternatives.

To this end, note that all second stage functions and equilibrium quantities, i.e., H, CS, and $(m^f, s^f, \pi^f)_{f\in F}$, depend on N^f only through τ^{N^f}. Thus, the impact of a unilateral expansion of firm f's product bundle, is equivalent to the impact of an increase in τ^{N^f}. As common sense would suggest, such a change benefits firm f by increasing demand for its products and its markups, and it harms f's competitors by reducing demand for their products and their markups. Furthermore, it raises consumer welfare. These intuitions are formalized in proposition 1, where the subindex 0 (resp. 1) denotes the equilibrium value of the corresponding variable before (after) the introduction of the new product.

Proposition 1. *Say f unilaterally introduces product j' to the market so that its product bundle goes from N^f to $N^f \cup \{j'\}$. The following holds:*

a) The equilibrium inclusive value and consumer surplus increase: $H_1 > H_0$, $CS_1 > CS_0$.

b) f's market share, markup, and profits increase: $s_1^f > s_0^f$, $s_1^{f|F} > s_0^{f|F}$, $m_1^f > m_0^f$, and $\pi_1^f > \pi_0^f$.

c) f's competitors' market shares, markups and profits decrease: $s_1^{g|F} < s_0^{g|F}$, $m_1^g < m_0^g$, $\pi_1^g < \pi_0^g$ for all $g \neq f$.

Proof. See appendix C. □

By proposition 1, f's equilibrium variable profits are increasing in τ^{N^f}. It stands to reason that for a given bundle size n^f, the bundle that maximizes f's variable profits is the one that maximizes τ^{N^f}, which in this context is the one composed by the n products with largest τ_j's. I denote these bundles by N_n^f, and refer to it as *anonymous bundles* because, as I will argue below, conditional on the bundle size the identity of the products in this type of bundles is irrelevant.

If, in addition, FC^f depends on N^f only through its size, then all bundles of size n generate the same fixed cost. These two observations imply that the anonymous bundle N_n^f, strictly dominates all other bundles of size n.

Assumption 1, below, states the conditions under which the strict dominance result holds, and definition 3.3 provides a formal definition of anonymous bundle.

Assumption 1. *Assume the following:*

a) FC^f depends on N^f only through $|N^f|$, i.e., for any two bundles N^f and $\tilde{N}^f$ such that. $|N^f| = |\tilde{N}^f|$, we have $FC^f(N^f) = FC(\tilde{N}^f)$.

b) *Products can be ranked by their type, τ_j, in strictly decreasing order,*[12] *with the subindex j representing j's position in the rank, i.e.*

$$\tau_1 > \tau_2 > \tau_3 \dots$$

Definition 3.3 (Anonymous Bundle)**.** An anonymous bundle of size n for firm f is the bundle composed by the n products with largest product type. This is:

$$N_n^f \equiv \{1, 2, 3 \dots, n\} \tag{20}$$

Proposition 2, below, formalizes the intuition that conditional on a bundle size, the anonymous bundle strictly dominates all other bundles.

Proposition 2. *Assume 1 holds, then:*

a) *For every firm f, the anonymous bundle N_n^f stricly dominates all other bundles of size n. This is:*

$$\pi^f\left(m^f, H; (N_n^f, N^{-f})\right) - FC^f(N_n^f) > \pi^f\left(m^f, H; (N^f, N^{-f})\right) - FC^f(N^f)$$

for all N^{-f}, and all $N^f \neq N_n^f$ such that $|N^f| = n$. (Dependence of m^f and H on bundles N_n^f, N^f, and N^{-f} is omitted for brevity).

b) *For every firm f, and for any N^{-f}, the anonymous bundle N_n^f generates the largest consumer surplus out of all bundles of size n. This is:*

$$CS\left(N_n^f, N^{-f}\right) > CS\left(N^f, N^{-f}\right)$$

for all N^{-f}, all $N^f \neq N_n^f$ such that $|N^f| = n$.

Proof. See appendix D. □

In essence, this result says that when deciding between two products j and l, if $\tau_j > \tau_l$ a firm will prefer to introduce j before l, regardless what other products there are in the market. Hence, we can interpret τ_j as a bundle-independent measure the profitability of product j. This result has two significant consequences. First, without loss of generality, when searching for an equilibrium, I can focus exclusively on the space of anonymous bundles $(N_n^f)_{n=1,2,\dots,|\mathcal{N}^f|}$. Second, in equilibrium, firm f's bundle size is a sufficient statistic for the composition of the bundle itself.

[12] There are two implicit assumptions embedded here. First, that there are no two products j and k. such that $\tau_j = \tau_k$. This rules out situations in which firms are indifferent between two products, and serves to guarantee that best responses are singleton.

Second, there is an assumption that the notion that the notion of "the product with the next highest τ_j" is well defined. In particular, it rules out situations where there exist a subsequence of products, $\kappa(j)$, such that $\tau_{\kappa(j)} < \bar{\tau}$ for all j and $\tau_{\kappa(j)} \to \bar{\tau}$. In such situation, standing at $\bar{\tau}$ there would be no next highest product.

The two stage game is likely to generate multiple equilibria. To see this, consider a two-firm, $f = 1, 2$ one product per firm game where π^M is the profit either firm gets from being a monopolist (the only firm that introduces its product), $\pi^D < \pi^M$ is the profit from being a duopolist (both introduce their product), fc is the cost of introducing a product, and the profit of not introducing a product is normalized to 0. This is a version of the entry model studied by Tamer (2003), and it is easy to that if the market can support either firm independently but not both, i.e., $\pi^D < fc < \pi^M$, the game has two pure strategy equilibria, one where $f = 1$ enters and $f = 2$ does not, and one where $f = 1$ does not enter and $f = 2$ does.

The dynamic model I present in section 4 differs from this one in that firms receive a bundle size specific profit shock which is privately observed. It should be noted, however, that introducing such random shocks to this two-stage game is trivial, and that proposition 2 will still hold in this environment, as the presence of size specific random shock does not change the comparison of bundles of the same size. Furthermore, introducing the random shock will not alleviate the multiple equilibrium problem.

3.2 Two Firm Example Numerical Example

To build some intuition before moving to the dynamic game, I present a simple two firm example of the two-stage game. The properties of the product choice game are similar to those of Cournot competition in various respects. First, the number of products, like the Cournot quantities, are strategic substitutes. Second, an increase in market size leads to an outward shift in the best responses, i.e. greater markets have more product varieties. Third, the strategic effect is less than one-for-one, i.e., the slope of the best response is (weakly) smaller than 1.

Each firm has a product menu of 20 products, i.e. $|\mathcal{N}^f| = 20$. The price sensitivity is. $\lambda = 1$ and all products have the same production cost $c_j^0 = c_j^1 = 0$. Furthermore, the demand parameter α_j^f is described by:

$$\alpha_j^0 = \log\left(\frac{0.2 - 0.01\kappa(j-1)}{5000}\right)$$

$$\alpha_j^1 = \log\left(\frac{0.16 - 0.005(j-1)}{5000}\right)$$

where j is the product index, and parameter κ controls the speed with which the demand for firm 0's marginal variety drops (a larger κ implying a faster drop). Both firms have a linear fixed cost of $FC^f(n^f) = 0.001n^f$, where n^f is the number of products that f introduces to the market.

I set $\gamma^0 = \gamma^1 = 1$, and $\rho = 0.1$ to match the results from demand estimation from section 6. By proposition 2, firms will choose anonymous bundles, so bundle types are:

$$N_n^f = \{1, 2, 3 \ldots, n\}$$

so that the bundle types are:

$$\tau^{n^0} = \sum_{j=1}^{n^1} e^{(\alpha_j - c_j)/\rho} = \sum_{j=1}^{n^1} \frac{0.2 - 0.01\kappa(j-1)}{1000}$$

$$\tau^{n^1} = \sum_{j=1}^{n^0} e^{(\alpha_j - c_j)/\rho} = \sum_{j=1}^{n} \frac{0.16 - 0.005(j-1)}{1000}$$

Note that $\tau^{n^0} > \tau^{n^1}$ for. low values of n, and the inequality is flipped for large values. Firm 0 chooses n^0 to solve:

$$\max_{n^0} \mathcal{D} \times \pi^0(n^0, n^1) - 0.001n^0$$

Firm 1 solves an analogous problem.

I vary the parameters in red, κ and $\mathcal{D}$, to study their effects on profits and variety. I begin by setting $\kappa = 1$ and $\mathcal{D} = 1$, and numerically solved the pricing game for all possible combinations of anonymous bundle, (N_n^1, N_n^2), of which there are 400, using equations (15) and (16). Figure 2 plots, on the left π^0 as a function of n^0 for different values of n^1. This panel shows the strategic substitutability between n^0 and n^1, i.e., for higher values of n^1 the optimal n^0 decreases, as shown by the gray line that collects the optimum for different values of n^1. The panel on the right presents the best responses which have a slope weakly smaller than 1. As in the Cournot case with linear demand, the Nash Equilibrium is an attraction point, so that iteration over the best responses is guaranteed to converge to it.

Figure 2: π^0 vs. n^0 for levels of n^1. Best Responses.

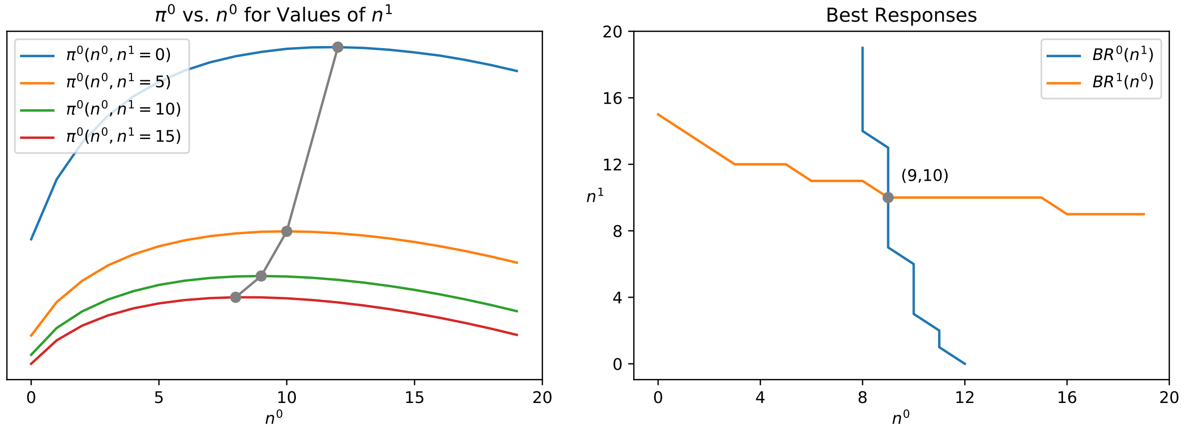
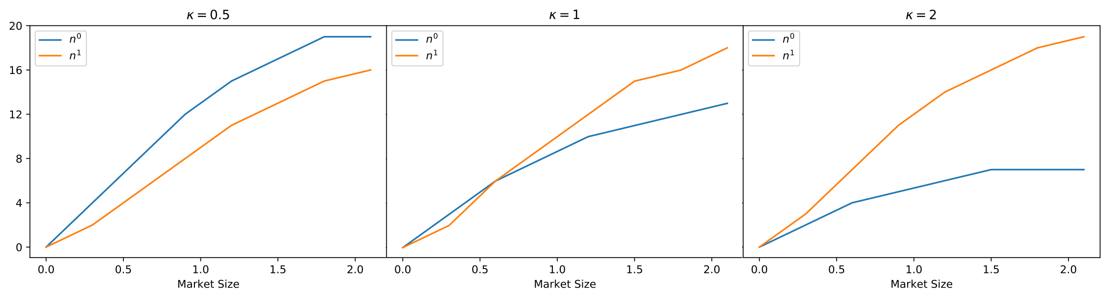


Figure 3 presents the equilibrium combinations of (n^0, n^1) for different values of $\mathcal{D}$ in the interval $[0, 2]$. Each corresponds to a different value of $\kappa = 0.5, 1, 2$, meaning a different rate with which product types of firm 0 drop. A larger κ implies a faster rate of decline of τ^{n^0}, so the profitability of the marginal variety decreases at a faster rate.

As argued above, the equilibrium number of varieties are increasing in market size. When $\kappa = 1$, for small $\mathcal{D}$, both firms introduce few products. Firm 0 introduces more products than firm 1, because for low n, $\tau^{n_0} > \tau^{n_1}$, i.e., the marginal variety of 0 is greater more profitable than the marginal variety of 1. This relation is reversed for large markets because the profitability of firm 0's products decreases faster than those of firm 1.

Figure 3: Equilibrium (n^0, n^1) vs. Market Size



Finally, note that higher values of κ imply that the products of firm 0 are less profitable, so as κ grows firm 0 tends to introduce fewer products and firm 1 to introduce more. This is represented by a rightward shift of n^0, and a leftward shift of n^1.

4 Dynamic Game

In this section I present the dynamic-discrete choice game. I begin by presenting the firm's problem and specifying the strategies and equilibrium concept I focus on.

The dynamic's model are generated firms expectations regarding the evolution of demand and marginal cost, firms' beliefs about their competitors future behavior, adjustment costs that influence the firms' cost-benefit calculations, and, as is standard in the literature, a private choice specific shock that rationalizes disagreements between the data and the model.

4.1 Firm's Problem

As before, there is a fixed set of firms indexed by $f = 1, \ldots, F$. Time is discrete and indexed by $q = 0, 1, 2, \ldots, Q$, with $Q \leq \infty$, and firms choose product bundles one period in advance, from a time invatiant product menu $\mathcal{N}^f$, with products indexed by $j = 1, \ldots, |\mathcal{N}^f|$.

Timing is as follows. At the start of every date, firms observe all product bundles $N_q = (N_q^f, N_q^{-f})$ and the vector of product types,

$$\tau_q = (\tau_{jq}^f)_{\forall j \in \mathcal{N}^f, \forall f \in F} \tag{21}$$

which lives in a finite set $\mathcal{T}$. Then, firms pay a fixed cost $FC^f(N_q^f)$, and engage in Nash-Bertrand price competition as described in section 2, which results in a variable profit of $\pi^f(N_q^f, N_q^{-f}, \tau_q)$.

This notation is meant to highlight that profits depend on product bundles, but that the profitability of a bundle depends on the underlying product types which can change over time.

After price competition, firms observe their private choice specific shocks $\xi_t^f(N_{q+1}^f)$, and choose next period's product bundles, $N_{q+1}^f \in \mathcal{N}^f$. Finally, by adjusting product assortment from N_q^f to N_{q+1}^f firms incur in an adjustment cost of $AC^f(N_{q+1}^f, N_q^f)$.

Note that vector τ_q defined in (21) collects the product type of all products, regardless of whether they are in the market or not. This is different from $\tau^{N_q} = (\tau^{N_q^f}, \tau^{N_q^{-f}})$ defined in section 2, which represents the vector of bundle types for products that firms actually supply, i.e.,

$$\tau^{N_q^f} = \log\left(\sum_{j\in N_q^f} e^{\tau_{jq}^f/\gamma^f}\right)$$

for each f.

I collect the publicly observable state in:

$$x_q \equiv (N_q^f, N_q^{-f}, \tau_q) \tag{22}$$

and, with these elements, I write f's non-random stage payoff as:

$$\Pi^f(N_{q+1}^f, x_q) \equiv \pi^f(x_q) - FC^f(N_q^f) - AC^f(N_{q+1}^f, N_q^f) \tag{23}$$

So, the stage payoff is simply the market profits, minus fixed costs and adjustment cost, and it depends today's decision, N_{q+1}^f, only through the adjustment cost. The rest of the effect of N_{q+1}^f comes from the continuation value, as shown below.

Given an initial state, x_q, and a state contingent sequence $(N_i^{-f})_{i=q+1,\ldots,Q}$, f chooses a contingent sequence of product bundles to maximize its discounted profits at time q. This is:

$$\begin{aligned} V_q^f(x_q, \xi_q^f) &= \max_{(N_q^f)_{q+1,\ldots,Q}} E\left[\sum_{i=q}^{Q-1} \beta^{i-q}\Big(\Pi^f(N_{i+1}^f, x_i) + \theta_{scale}^f \xi_i^f(N_{i+1}^f)\Big)\right. \\ &+ \left. \beta^{Q-q} V_Q^f(x_Q, \xi_Q)\Big| x_q, \xi_q^f\right] \end{aligned} \tag{24}$$

where the expectation is taken with respect to $(x_i, \xi_i^f)_{i=q+1,\ldots,Q}$. Parameter β is the common discount factor, V_Q^f is the terminal value, which is set to zero if $Q = \infty$, and θ_{scale}^f is the scale parameter of the choice specific profit shock.

Let the Markov transition kernel $P_q^f(N_{q+1}^{-f}, \tau_{q+1}, \xi_{q+1}^f | N_{q+1}^f, x_q, \xi_q^f)$ represent f's beliefs about the evolution of the market conditional on the current state x_q, its private shock ξ_q^f, and its choice, N_{q+1}^f. Function P_q^f subsumes f's beliefs about the evolution of demand and marginal costs, embedded in τ_q, its belief about the evolution of the private shocks, and its beliefs about the behavior of its competitors. With this, I define the expected value function, EV, as the value of playing optimally starting next period conditional on state x_q and on choosing N^f_{q+1} today. This is:

$$EV^f_{q+1}(N^f_{q+1}, x_q, \xi^f_q) \equiv \int_{N^{-f}_{q+1}, \tau_{q+1}\xi^f_{q+1}} V^f_{q+1}(N^f_{q+1}, N^{-f}_{q+1}, \tau_{q+1}, \xi^f_{q+1}) \times dP^f_q(N^{-f}_{q+1}, \tau_{q+1}, \xi^f_{q+1}|N^f_{q+1}, x_q, \xi^f_q) \tag{25}$$

Then, by the principle of optimality, f's Bellman Equation is:

$$V^f_q(x_q, \xi^f_q) = \max_{N^f_{q+1}} \Pi^f(N^f_{q+1}, x_q) + \xi^f_q(N^f_{q+1}) + \beta EV^f_{q+1}(N^f_{q+1}, x_q, \xi^f_q) \tag{26}$$

These are the main building blocks of the dynamic model. The rest of this section aims to provide assumptions under which (a version of) the strict dominance result of section 3 carries to the dynamic game. A step in this direction is to ensure that bundles of the same size are fungible from a cost perspective. This is guaranteed by the following assumption.

Assumption 2. *Assume the following holds:*

a) The fixed cost FC^f depends on N^f only through its size. For any N^f and $\tilde{N}^f$ such that $|N^f| = |\tilde{N}^f|$, $FC^f(N^f) = FC^f(\tilde{N}^f)$.

b) The adjustment cost AC^f depends only on the difference in sizes of N^f_q and N^f_{q+1}. For any N^f and $\tilde{N}^f$, $AC^f(N^f_q, N^f_{q+1}) = AC^f(|N^f_{q+1}| - |\tilde{N}^f_q|)$.

c) If N^f and $\tilde{N}^f$ have the same size, then: $\xi^f_t(N^f) = \xi^f_t(\tilde{N}^f)$. In other words, the random shock is specific to a particular bundle size, not to a particular bundle.

Together, assumptions a), b) and c) imply that, from a cost perspective, bundles of the same size are interchangeable. In consequence, bundle size is a sufficient statistic for firms' costs. Assumption c), which may appear somewhat ad-hoc, is quite natural given assumption b). Given that AC^f is independent of the composition of the bundles, it seems natural for the adjustment cost shocks, ξ^f_t, to be independent of the composition of the bundle too.

These assumptions play a fundamental role in the process of reducing firms' strategy sets, and the state space of the game. Without them bundles of the same size would yield different costs, so even if bundles can be aggregated on the demand side, firms would have to compare all possible bundles, and keep track of the bundles of their competitors, to make their own choices of N^f and forecast the evolution of the market.

4.2 Strategies and Equilibrium Concept

At this point, all cost elements of a firm's period payoffs depend only on the size of N^f, while market profits depend on the aggregate bundle N, only though the F dimensional vector $\tau^{N_q} = (\tau^{N^f_q})_{f \in F}$. In what follows, I show that under certain assumptions, the result of section 3 can be extended to this dynamic context, i.e., firms introduce products from highest to lowest τ_{jq}. To do this, I present the definition of a Markov Strategy in the spirit of Maskin and Tirole (2001), and that of Anonymous Markov Strategies, the class of Markov Strategies on which I will focus.

Definition 4.1. A Markov Strategy for firm f is a series of functions, $\sigma^f = (\sigma_q^f)_{\forall q}$, each mapping from f's the current state (x_q, ξ_q^f) to its next period's product bundle, N_{q+1}^f. This is:

$$\sigma_q^f : \mathcal{N}^f \times \left(\underset{g \neq f}{\times} \mathcal{N}^g \right) \times \mathcal{T} \times \mathbb{R}^{|\mathcal{N}^f|} \to \mathcal{N}^f$$

Definition 4.2. An Anonymous Markov Strategy (AMS) for firm f is a Markov Strategy satisfying:

a) Actions depend only on the size of the product bundles, i.e. if $|N_q^f| = |\tilde{N}_q^f|$ for all $f \in F$:

$$\sigma_q^f(\underbrace{N_q^f, N_q^{-f}, \tau_q}_{x_q}, \xi_q^f) = \sigma_q^f(\underbrace{\tilde{N}_q^f, \tilde{N}_q^{-f}, \tau_q}_{\tilde{x}_q}, \xi_q^f)$$

for all $\tau_q \in \mathcal{T}$, all ξ_q^f and all q.

b) At all dates, the bundle $\sigma_q^f(x_q, \xi_q^f)$ is an anonymous bundle, $N_{n,q}^f$, as defined in 3.3, i.e., it is composed of the n products with largest product type. Mathematically,

$$|\sigma_q^f(x_q, \xi_q^f)| = n \Leftrightarrow \sigma_q^f(x_q, \xi_q^f) = N_{n,q}^f$$

for all x_q and ξ_q^f.

It is easy to see that if all firms play an AMS, then the size of the product bundles is enough to predict flow profits and forecast the evolution of the market structure. This raises the question of what conditions make it optimal for firms to play AMS's, and under what conditions does an equilibrium involving AMS's exist. Concerning the latter, I simply assume existence, noting that although there is no off-the-shelve equilibrium existence result for this type of model,[13] in practical applications equilibrium multiplicity is more often a concern than non-existence. Note, however, that in the empirical application, existence is guaranteed by the finite time assumption.

About the former question, I show in the appendix that assumption 3 below guarantees that an AMS is the best response to an AMS.

Assumption 3 (Constant Product Ranking)**.** *For every product j, product type τ_{jt}^f can be decomposed into:*

$$\tau_{jq}^f = \bar{\tau}_j^f + \bar{\tau}_q^f + \tilde{\tau}_{jq}^f \tag{27}$$

Where $\bar{\tau}_j$ is a product specific time invariant component, $\bar{\tau}_q^f$ is a firm-date(-market) specific product invariant component, and τ_{jq}^f is a product-firm-time varying component, which is independent over product-firm-time, and is revealed only after product decisions have been made.

[13] Doraszelski and Satterthwaite (2010) provide an existence result for a related model.

Under assumption 3, I reinterpret x_q as:

$$x_q \equiv (N^f, N^{-f}, (\tau_q^{N^f})_{f\in F}) \tag{28}$$

Importantly, assumption 3 does not require one product to be more profitable than the other in every single realization of the pricing game. It simply demands that whenever ranking reversals occur, they do so randomly and unpredictably from the perspective of all firms. A sufficient condition for assumption 3 is that α^f_{jq}'s and marginal cost c^f_{jq} are themselves decomposable into a product specific-time invariant component, a time specific-product invariant component, and a zero mean a shock that is independent and idiosyncratic component:

$$\begin{aligned}
\alpha^f_{jq} &\equiv \underbrace{\bar{\alpha}^f_j}_{\text{Time Invariant}} + \underbrace{\bar{\alpha}^f_q}_{\text{Product-Firm Invariant}} + \underbrace{\epsilon^f_{jq}}_{\text{Product-Firm-Time IID}} \\
c^f_{jq} &\equiv \overbrace{\bar{\eta}^f_j} + \overbrace{\bar{\eta}^f_q} + \overbrace{\tilde{\omega}^f_{jq}} \\
\tau^f_{jq} &\equiv \underbrace{\bar{\tau}^f_j}_{\bar{\alpha}^f_j - \lambda\bar{\eta}^f_j} + \underbrace{\bar{\tau}^f_q}_{\bar{\alpha}^f_q - \lambda\eta^f_q} + \underbrace{\tilde{\tau}^f_{jq}}_{\tilde{\epsilon}^f_{jq} - \lambda\tilde{\omega}^f_{jq}}
\end{aligned} \tag{29}$$

As shown in section 3, from a static perspective firm f will be better off by introducing (removing) the products with the highest (lowest) τ_j's first. In a dynamic context, f may be willing to violate this rule if it expects its lower profits today to be compensated by larger profits at a later date. In this model, these larger profits may come from two sources: the expectation of growth in a product's τ^f_j, or a strategic effect through f's impact over its competitors' behavior. Assumption 3 rules out the first type of incentive, as firms will not choose not-so-profitable products today in the expectation that they will become more profitable in the future. On the other hand, if f's competitors are playing AMS's, f gains nothing from choosing a less profitable product over a more profitable one. If f's competitors are not influenced by the particular products f introduces, only by the number of them, there is nothing to be gained by departing from anonymous bundles. In consequence, under assumption 3 and if all of f's competitors' play an AMS, firm f's optimal strategy is itself an AMS.

Proposition 3. *Assume 2 and 3 hold, and assume all $g \neq f$ play an AMS. Then, f's optimal strategy is also an AMS.*

Proof. See appendix. □

For the rest of the paper, I re-interpret all functions as depending directly on the size of f's bundles, n^f, understanding that for any bundle size, firms are always choosing the strictly dominant bundle N^f_n, as defined in (20). With this result at hand, I define the equilibrium concept that I will focus on:

Definition 4.3 (Anonymous Markov Equilibria (AME))**.** An Anonymous Markov Equilibria is a profile of AMS's, $(\sigma^f)_{f\in F}$, such that:

$$\sigma_q^f(x_q, \xi_q^f) \quad = \quad \underset{n_{q+1}^f}{\text{argmax}} \ \Pi^f(n_{q+1}^f, x_q) + \xi_q^f(n_{q+1}^f) + \beta EV_{q+1}^f(n_{q+1}^f, x_q, \xi_q^f)$$

for all f.

An AME effectively reduces the size of f's strategy space from $2^{|\mathcal{N}^f|}$ to $|\mathcal{N}^f|$. Furthermore, it reduces the state space (omitting the state of τ_q) from $2^{\sum_f |\mathcal{N}^f|}$ to $\prod_f |\mathcal{N}^f|$. This reductions imply large computational gains that make it feasible to study problems that would otherwise be impossible, like the empirical application I show in section 7.

Assuming that the data is generated by an AME comes with some costs. On top of the impositions implied by assumption 3, AMEs rules out equilibria where strategies depend on the identity of the products, be it directly or indirectly through product characteristics. An AME could not, for example, accommodate a situation like that studied by Sullivan (2018). There he provides evidence of collusion in the product space among ice cream manufacturers, each specializing in a different area of the product space. Since cooperation depends on firms not introducing products in their competitors' section of the product space, the characteristics of the products are necessary to describe the equilibrium, violating the definition of AME.

Before moving on to the empirical application, a word about the AME is in order. The concept of AME *is not* a refinement of MPE, although it is tempting to think so, and the reason is that the AME is not perfect (i.e., firms do not choose optimally out of path). To see this, consider a two firm version of the dynamic game described in this section, and assume 2 and 3 hold. Furthermore, assume that, as AMS's mandate, both firms keep track only of the bundle sizes. If firm 2 deviates from optimality and starts introducing products in reverse order, from lowest to highest τ_j^f, firm 1 would find it profitable to behave more aggressively and add more products to the market, as the market became less competitive. Since firm 1 only conditions its behavior on the size of N_q^2, however, its policy will be the same as if firm 2 were behaving optimally, so firm 1 will be responding sub-optimally to firm 2's deviation.

Note that this violation of perfection is not the result of some time inconsistency such as a non-believable threat or promise, so even if the AME is not perfect, players have no incentive to deviate from it. Rather, it is a direct consequence of the fact that firms do not condition their strategies on the product bundles, but only on their bundle sizes, and are constrained in the way they can respond to out of path behavior. In this sense, an AMS can be regarded as a boundedly rational strategy (and the AME a boundedly rational equilibrium), which, if assumptions 2 and 3 hold, is observationally equivalent to an equilibrium where firms track bundles rather than bundle sizes.

5 Data

For the remainder of the paper I focus on an application to the RTE-Cereal industry. Several reasons make the RTE-cereal industry a good laboratory in which to estimate this model and conduct counterfactuals. First, because horizontal differentiation is a prominent feature of the industry. Second, because its market structure is remarkably stable, being controlled by same four dominant players, Kellogs's, General Mills, Post, and Quaker Oats, going back as far as the 1950s (see Schmalensee (1978)). Jointly, they represent 80% of total sales. The third reason is that previous research has found that pricing equilibrium is consistent with static Nash-Bertrand (see Nevo (2001), Backus et al. (2018)), which matches my modeling assumptions. The fourth reason is that there is considerable variation in product bundles between markets and over time, which helps identify the parameters of the dynamic model.

The two largest companies in the market are Kellogg's and General Mills, each controlling roughly 30% of the national sales and supplying roughly 38 products. Both are large international food manufacturers, focused mainly on the production of cereal, but with considerable interests in other household consumption goods, such as snack bars and frozen foods. Some of General Mills' most popular products are Cherios and Trix. On the other side, some Kellogg's better known cereals are Corn Flakes and Frosted Flakes. Additionally, it is noteworthy that in 2000 Kellogg's bought organic cereal producer, Kashi, and has kept as a subsidiary company ever since.

The third largest firm is Post, which has a market share of roughly 15%. Control over Post has changed hands multiple times over the past decade and a half. Prior to 2007, Post was a part of Kraft, at which point it was sold to the Ralcorp Holdings. Later, in 2012, Post was spinned-off into its own company, after an IPO. Finally, in 2015, Post purchased cereal producer MOM-Cereals (previously Malt-O-Meal). Post is the producer of several cereal products, such as Cocoa Pebbles and Oh's.

Finally, Quaker is the fourth largest company, representing roughly 10% of the market. Owned by PepsiCo since 2001, Quaker represents a small portion of the total sales of the holding, ranging between 2 and 3%. Quaker is the producer of Cap'n Chrunch and Life Cereal, among others well known brands and hot cereals.

My primary data source is the Nielsen scanner sales data for the RTE-Cereal industry, provided by the Kilts Center for Marketing at the University of Chicago. The data records all weekly sales by *upc* at the store level for selected retailers covering all of the contiguous US from 2006 to 2016. Store location can be identified at the three-digit ZIP code level, and stores can be identified as belonging to the same parent-company/retailer, however the identity of this parent-company is unobserved.

My contractual agreement with Nielsen precludes me from revealing the data on the individual companies, so I randomly attach to each large player a label, $F1$, $F2$, $F3$ and $F4$, and group all products belonging to fringe competitors under the umbrella of firm zero, i.e., $F0$. Additionally, I follow Backus et al. (2018) in keeping only 'code F' retail establishments (supermarkets), and exclude other establishments that sometimes carry cereals like mass-market stores and pharmacies, which represent a small fraction of sales. To minimize sample selection concerns, I drop all Designated

Market Areas (DMAs)[14] where Nielsen covers less than 80% of total sales. I also drop all retailers that register average yearly cereal sales below $10 million dollars, and products that register annual sales below $10 thousand dollars, for representing a small fraction of total sales. Finally, I drop retailers that do not present sales every quarter in my sample period. This process leaves me with 1.37 million product-week-market observations, over distinct 530 products, 509 weeks, and 15 retailers.

I follow Nevo (2001) in aggregating different sizes of the same product into a single category and express all quantities in terms of servings, which I assume to be equivalent to $1.5oz$ of cereal. In particular, I define a product as a single *brand_ code_ uc*, which groups all sizes of the same product. I make this choice because product sizes have decreased during the sample period, and every time a product changes its size, a new *upc* is issued. If I were to define products as distinct *upc*'s, I'd be double-counting products in the transition period from a larger to a smaller size. By aggregating them under the same umbrella, I avoid this problem.

Table 1 presents summary statistics for each firm. Because the selected sample is composed by large supermarket chains, the share of the competitive fringe $F0$ is larger than the unconditional numbers would suggest. This is a result of large supermarkets carrying house-brands and more small brands than other retailers. As a result, the market share of $F0$, excluding the outside option, $s^{f|F}$, is 29% rather than 20%.

Table 1: Cross Market-Week Statistics by Firm

Firm	Obs.	n^f	p_j^f	s^f	$s^{f\|F}$	Servings
F0	575,799	82.8	0.470	0.079	0.289	0.38
		(20.7)	(0.180)	(0.026)	(0.087)	(0.35)
F1	282,296	39.6	0.419	0.067	0.243	0.33
		(4.3)	(0.095)	(0.025)	(0.077)	(0.31)
F2	267,715	37.2	0.407	0.075	0.268	0.36
		(4.7)	(0.117)	(0.032)	(0.084)	(0.33)
F3	161,837	22.7	0.330	0.036	0.130	0.18
		(5.6)	(0.085)	(0.023)	(0.060)	(0.20)
F4	71,107	9.9	0.366	0.019	0.069	0.10
		(1.7)	(0.124)	(0.016)	(0.048)	(0.12)

Note: Averages on top, standard deviation in parenthesis at the bottom.
p_j^f in \$/serving.
Servings in millions.

The rest of the figures match the description of the market outlined above. Note that $F0$ exhibits a considerably larger price per serving than the rest of the firms. This is probably a reflection of the fact that the final sample only contain supermarkets which offer more specialized products from small companies (organic, gluten-free, etc.) that can be sold at a higher price.

[14] A DMA is a contiguous group of counties, defined by Nielsen, that constitute a television market. DMAs cover all of the United States and are used mainly for advertisement targeting.

Table 2 shows summary statistics for the number of products for each firm. More importantly, the data shows that firms change their products frequently. The bundle size of one of these firms changes on average by roughly 1.5 per quarter, except for $F4$, for which this number is 0.7. Figure H.1 in the appendix shows a histogram of the absolute changes in n^f between quarters.

Table 2: Number of Products by Firm

	F0	F1	F2	F3	F4
Mean n^f	82.8	39.6	37.1	22.7	10.0
St.D. n^f	20.7	4.3	4.7	5.6	1.7
Min n^f	37.0	27.0	23.0	12.0	7.0
Max n^f	126.0	53.0	48.0	38.0	18.0
Mean Δn^f	3.7	1.8	1.7	1.4	0.6
St.D. Δn^f	4.2	1.4	1.6	1.5	0.7

I depart from most existing literature in that I equate a market to a retailer, rather than a geographic area. The rationale behind this decision is twofold. First, consumer choice sets are considerably more similar within retailers than within DMAs. Recent papers by DellaVigna and Gentzkow (2017) and Hitsch et al. (2017) have found that between retailer variation accounts for the vast majority of cross-sectional price variation for a product, while within retailer variation explains very little, indicating that consumers that go to different stores of the same retailer face roughly the same prices. I conducted an exercise of my own (see Appendix G), and found these patterns describe the cereal market well. In a nutshell, average within retailer variation is roughly 1%, while between retailer price variation is approximately 5%.

In the same vein, I investigate whether retailers offer the same bundle of products across geographic regions and found that, to a surprisingly large extent, they do. If a product is marketed by retailer r in a given DMA, the probability that it is marketed by retailer r in any other DMA is around 98%. In contrast, given that a product is found in a given store, the probability that it is found on another store in the same DMA is only slightly above 75% (again, see Appendix G).

A second reason for equating markets to retailers comes from anecdotal evidence obtained from the firms Financial Statements. It is clear from these documents that firms track their within retailer performance closely, while little mention there appears to be little interest in their geographic performance beyond broad international markets.

Put together, these pieces of evidence show that consumers buying in different stores within the same retailer face roughly the same choice set. In contrast, the choice sets of consumers buying from different retailers in the same DMA can differ substantially. Additionally, this evidence is at least consistent with the notion that producers tailor their strategies, both pricing and variety wise, to individual clients (retailers) rather than DMAs, or any other concept of geographic market for that matter. That being so, equating a market to a retailer seems a more natural choice. With this in mind, from now on, I use market and retailer interchangeably.

Table 3 presents market level summary statistics. Recall that retailers are not identified by name in the data, so I display each in the "Market" column with randomly chosen numbers. The table shows cross-market heterogeneity in market size, servings and sales. Average prices range from 0.365 dollars per serving, to 0.44. The share of the buy option, s^F, is stable across markets.[15].

Table 3: Market/Retailer Summary Statistics

Market	Market Size	Servings	Sales	p_j^f	s^F	$\sum_f n^f$	n^{F0}	n^{F1}	n^{F2}	n^{F3}	n^{F4}
0	244	68	59	0.432	0.276	221	108	41	38	24	10
1	101	31	24	0.440	0.280	182	74	42	35	22	10
2	1700	464	357	0.390	0.278	177	69	42	35	22	9
3	997	278	234	0.424	0.276	229	112	42	40	26	10
4	2490	681	502	0.370	0.273	205	79	43	43	29	11
5	1448	395	302	0.388	0.276	156	54	36	34	22	10
6	612	168	137	0.438	0.278	166	51	41	39	23	12
7	298	81	62	0.375	0.274	204	97	38	39	21	9
8	2271	622	426	0.393	0.274	208	100	40	37	21	9
9	2163	585	428	0.369	0.274	218	102	41	41	24	9
10	347	94	68	0.369	0.274	194	90	38	37	19	9
11	120	33	21	0.365	0.277	179	79	38	33	19	9
12	819	224	174	0.381	0.273	195	83	39	36	27	10
13	263	73	57	0.412	0.278	169	59	40	37	21	11
14	611	169	123	0.440	0.275	179	85	32	31	21	9

Note: Market Size and Servings in Millions of Quarterly Servings.
Sales in Millions of USD.
p_j^f in \$/serving.

Over the sample period and across the selected markets, the industry has been relatively stable, with the four largest firms accounting for roughly 80% of sales, and fringe firms a remaining relatively unimportant. Figure 4 shows that average market shares from table 1 are representative of the cross-market average market share for the whole period. Similar stories apply for number of products.

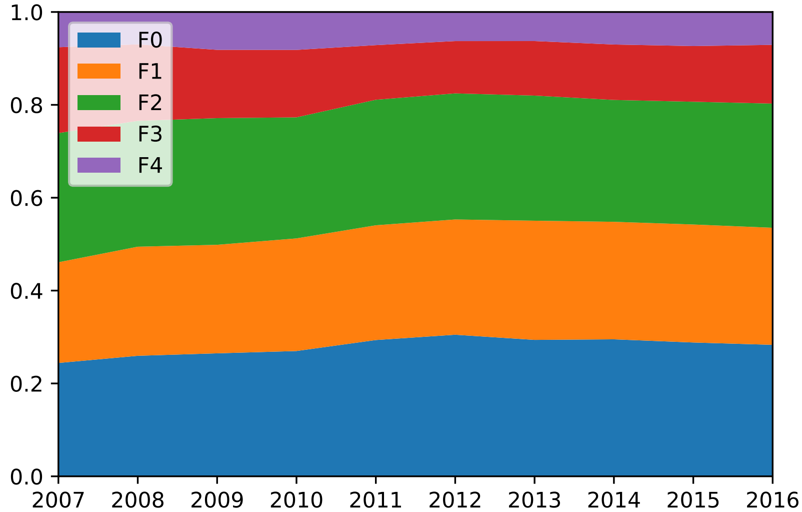
This plot, however, conceals significant between market variation, as well as within market time series variation. Figure 5, for example, shows the cross market mean (top) and standard deviation (bottom) of the number of products by firm-quarter. The top panel paints a picture of stability over time, similar to that of 4. The bottom panel, in contrasts, uncovers substantial cross-sectional variation in the number of products across retailer/markets for all firms except $F4$. In fact, for each of the large firms, the standard deviation in number of products is roughly equal to approximately 10% of its mean.

To highlight the within market/retailer time series variation of the number of products, I present the evolution of n^f for an arbitrary market/retailer in figure 6. This figure shows substantial variation in n^f, for all firms except $F4$. This is a pattern that is repeated in most markets.

Finally, for the estimation of the dynamic game, I aggregate data to the quarterly level to accelerate the estimation of the dynamic game and avoid registering as withdrawn products that

[15]See section 7 and appendix I for an explanation of how Market Size, s^0 and. $s^F = 1 - s^0$ are computed.

Figure 4: Market Shares Over Time



simply present zero sales. Tables H.3 and 1 in the appendix show some summary statistics for each year and for each firm, respectively. Additionally, table H.4 shows the mean number of distinct product each firm carries to a market for each year and the between market standard deviation. For every firm, the average deviation from the mean number of products is about 10% of the average product bundle, showing substantial cross-market variation in product availability.

6 Parametrization and Estimation - Demand and Marginal Cost

The estimation takes place in two stages. On a first stage, I use standard techniques in industrial organization, outlined by Berry (1994), to separately estimate the weekly demand and marginal cost. Then, I aggregate demand to the quarterly level and use these estimates to compute variable profits for each possible contingency of the dynamic game, i.e., for every (N^f_{kq}, N^{-f}_{kq}), for every k and q. In the second stage, I take the estimated variable profits and use them as an input in the estimation procedure of the dynamic parameters embedded in FC and AC via Nested Fixed Point (NFXP), Rust (1987).

6.1 Demand

As is standard, the utility that consumer i gets from product $j \in N^f$ at market k during week w is parametrized by a linear index as follows:

$$u^f_{ijkw} = \overbrace{\alpha_j + \alpha^f_{kw} + \epsilon^f_{jkw}}^{\alpha^f_{jkw}} - \lambda p^f_{jkw} + \nu^f_{ijkw} \tag{30}$$

Figure 5: Number of Products Over Time

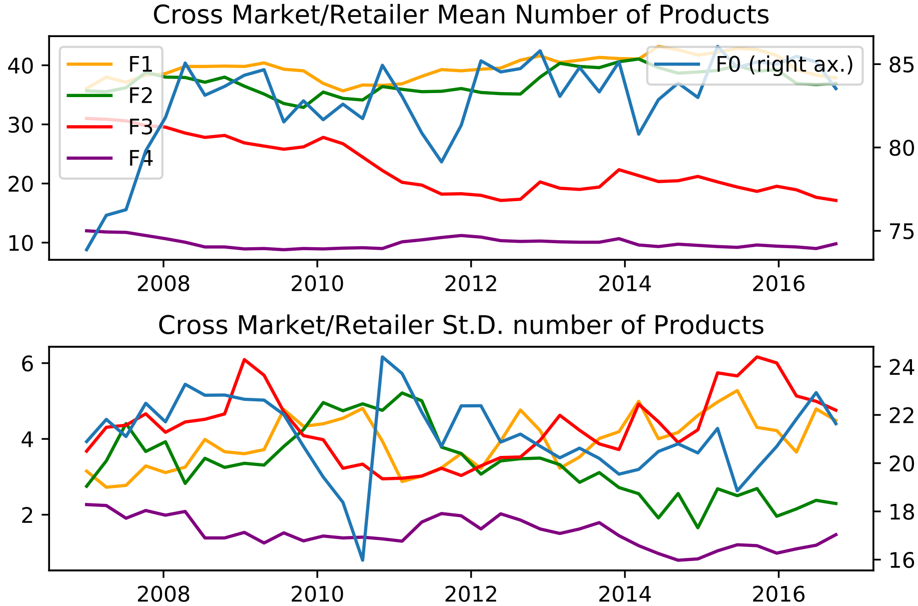
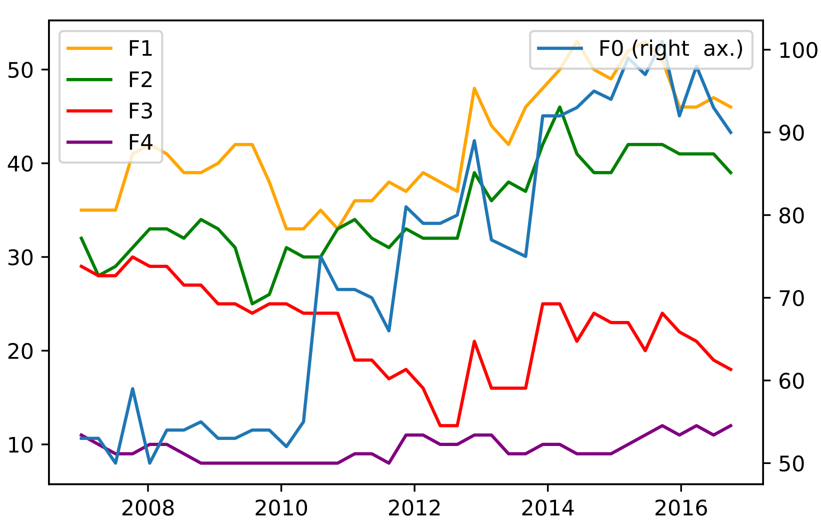


Figure 6: Number of Products Over Time: Single Market



In line with assumption 3 and expression 29, I decompose α^f_{jkw} into three elements: a time invariant product specific fixed effect, α_j, a market-time-firm varying component, α^f_{kw}, and a product specific utility shock common to all consumers within a product-firm-market-week, ϵ^f_{jkw}. As before, ν^f_{ijkw}, is a consumer idiosyncratic taste shock that follows a GEV distribution.

The utility model parametrized in (30) matches the nested logit model from Berry (1994), and yields the well known estimable equation:

$$\log\left(\frac{s_{jkw}}{s^0_{kw}}\right) = \alpha^f_j + \alpha^f_{kw} - \lambda p^f_{jkw} + (1-\rho)\log(s_{j|F,kw}) + \epsilon^f_{jkw} \tag{31}$$

where $s_{j|F}$ is the market share of product j in the inside nest, and s^0 is the market share of the outside option.[16] Note that the parametrization in (31) assumes that there are no firm level nests, i.e., $\gamma^f = 1$ for all f. I estimate a single nest specification, rather than the double nest presented in section 2, because estimates obtained from latter are inconsistent with the random utility model, i.e. $\gamma^f \notin [0, \rho]$. This result is not surprising. The implicit assumption in the double nest specification is that within-firm substitution is larger than between-firm substitution, which is unlikely to hold in the cereal market.

To calculate the market size and the market share of the outside option I follow Backus et al. (2018). In particular, I take the market size, $\mathcal{D}_{kw}$, to be proportional to the number of consumers who transit through market k during week w, as measured by the sales of milk and eggs, two commonly purchased goods. To do this, I regress the (log) cereal sales in servings of 1.5 ounces, against the (log) sales of milk and eggs, and I set the market size to be proportional to the predicted cereal sales from this regression, as shown below:

$$\mathcal{D}_{kw} = \kappa_{\mathcal{D}}\, e^{\hat{\kappa}_k + \hat{\kappa}_{\text{month}} + \hat{\kappa}_{\text{milk}} \log(\text{milk}_{kw}) + \hat{\kappa}_{\text{eggs}} \log(\text{eggs}_{kw})}$$

where the $\hat{\kappa}$'s are the estimates of the aforementioned OLS regression, with $\hat{\kappa}_k$ and $\hat{\kappa}_{\text{month}}$ representing a market and month fixed effect, and $\kappa_{\mathcal{D}}$ is a constant chosen to match an average share of the outside option to 72.4% (= 1−27.6%), which corresponds to the average probability of not purchasing cereals in the Nielsen Consumer Panel Dataset. Finally, I set s^0 equal to the total number of servings sold divided by the market size $\mathcal{D}_{kw}$. More details are provided in appendix I.

The term ϵ^f_{jkw} represents random utility shock to product j common to all consumers on market k and week w, and it explains the deviations of the observed market shares from their means. I assume ϵ^f_{jkw} is observable to the firm but not to the econometrician, so that that price is endogenous.

To solve the endogeneity problem, I use a 2SLS estimation. A standard price instrument is a marginal cost shifter, such as input prices. In this line, I use commodity prices of corn, oats, sugar, and wheat (see figure J.1 in the appendix for a time series plot of these prices). The advantage of these instruments is that they have a firm theoretical justification, and have a very plausible basis for exogeneity, provided that the firms behave as price takers with respect to input prices, which

[16]For an a derivation of the estimable expression in the double nest case see Appendix F.

seems reasonable in this industry. The disadvantage is that if these inputs represent a low fraction of the marginal cost, or if firms hedge their exposure to commodity prices through the financial markets, then they are likely to be weak. Furthermore, because commodity prices are common to all markets and products, they identify the parameters of interest only through variation over time, adding to the weak instrument concern.

In addition to commodity prices, I follow Nevo (2001) and create instruments in the spirit of Hausman (1996), i.e., I instrument p_{jkw} using the average price of good j at week w in markets different from k, $\frac{1}{K-1}\sum_{k'\neq k} p_{jk'q}$. Relevance of the so-called Hausman instrument comes from the fact that production is centralized, so marginal cost shocks hit all markets, and hence, prices in different markets are correlated through these exogenous marginal cost shocks. This is better seen in expression (33) in the next subsection. If marginal cost shocks, ω, are correlated across markets, then a large ω^f_{jkw} will tend to be accompanied by large $\omega^f_{jk'w}$ for $k' \neq k$, which will induce an exogenous price increase for product j throughout all markets. That being so, prices in one market are correlated to prices on another market through this exogenous source of variation, making the first a relevant instrument for the second.

Exogeneity of the Hausman instrument requires that conditional on observables, marginal cost shocks are the only source of variation that generates correlation between prices in different markets. In other words, exogeneity requires taste shocks to be uncorrelated across markets, conditional on product, market, and time fixed effects, as a failure of this would induce endogenous correlation between prices in different markets. There are many ways in which exogeneity may fail. For example, if firms engage in advertising campaigns that boost demand across markets. To mitigate this concern, I include time-firm specific dummies in some specifications to capture the temporary effect of a firms' advertising over its demand; results are remarkably stable to this adjustment (see table 6.1). It should be noted, however, that recent findings by Shapiro et al. (2018) show that the effectiveness of television advertising in increasing sales is much lower than previous research has found, suggesting that the worries of exogeneity violations in Hausman instruments due to cross-market advertising may have been overblown.

Alternatively, one may worry that taste shifts over time due to increasing taste for healthy breakfast choices, for example, may induce across market demand correlation, causing exogeneity to fail. To address this concern, I estimate demand using a four-year window, from 2007 to 2010, to minimize the incidence of taste changes on estimation. Although price sensitivity is larger in these estimations, and the nesting parameter is smaller, it is reassuring that the ratio λ/ρ is largely unchanged. (See table J in the appendix for these results).

Variable $\log(s_{j|F,kw})$ is also endogenous in equation 31. I instrument it using quarterly lagged aggregate number of products, i.e., $\sum_{f\in F} n^f_k$.[17] Intuitively, the more products inside a nest, the less likely it is that a particular product j is chosen, decreasing its share in the nest, $s_{j|F,kw}$, however, this does not affect the intrinsic attractiveness of product j and therefore is uncorrelated to ϵ_{jkw}. I use last quarter's value of n_q, rather than the current value, to ameliorate concerns that n_q

[17] I choose a quarterly lag for consistency with the dynamic game, which is played with quarterly frequency.

might be influenced by unobserved (by the econometrician) demand shocks, i.e. firms introduce more products when they expect high demand. Also, I introduce firm-year dummies to capture firm-specific variation in demand over time, which may induce firms to adjust their product set.

Table 6.1 below shows the result for the second stage regression estimating equation (30). The specifications presented in each column differ from each other only in the fixed effects that they include. Results look remarkably stable across the first four specifications, yielding similar values for λ, and ρ. Notably, ρ is relatively small, indicating little substitution between buying and the outside option. Appendix K shows histograms of the estimated elasticities for each product-week-market, and the market elasticity for each market-week, for specification 4. The average price elasticity is 5.04, while the average market elasticity is 0.22.

For first-stage estimation results see appendix J. In a nutshell, first-stage results show that instruments have high predictive power. A common concern when in the presence of two or more endogenous variables is that only one instrument may be predicting the two endogenous variables, in which case identification the whole set of instruments is weak. To rule out this concern, I ran the test of weak instruments in the presence of multiple endogenous variables proposed by Sanderson and Windmeijer (2016). Under the null of weak instruments, the proposed statistic follows an F distribution. Results soundly reject the null hypothesis of weak instruments, with p-values well below 1%.

Table 4: Second Stage Nested Logit

	(1) $\log(s_j/s^0)$	(2) $\log(s_j/s^0)$	(3) $\log(s_j/s^0)$	(4) $\log(s_j/s^0)$
λ	-0.873***	-0.916***	-0.904***	-0.933***
	(0.019)	(0.021)	(0.021)	(0.021)
ρ	0.06***	0.13***	0.13***	0.13***
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	1358754	1358754	1358754	1358754
λ/ρ	-14.03	-6.99	-6.85	-7.00
ProductFE	Yes	Yes	Yes	Yes
MarketFE	Yes	Yes	Yes	No
QuarterFE	No	Yes	No	Yes
FirmQuarterFE	No	No	Yes	No
FirmMarketFE	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

6.2 Marginal Cost

As is standard, I back out marginal costs using demand estimates and firms' first-order conditions. To do this, I invert the firms' first-order condition to find the marginal cost that is consistent with the demand estimates, the observed price, and the behavioral assumption underlying the pricing equilibrium concept, in this case, Nash-Bertrand. For the particular case of the nested-logit demand this is:

$$c_{jkw}^{f} = p_{jkw}^{f} - \frac{1}{\lambda} \frac{1}{1 - s^{f|F}(1 - \rho s^{0})} \tag{32}$$

for c_{jkw}^{f}, for all k and all w.

Following Berry et al. (1995), I parametrize marginal cost as:

$$\log(c_{jkw}^{f}) = \eta_{j}^{f} + \eta_{kw}^{f} + \omega_{jkw}^{f} \tag{33}$$

where η_{j}^{f} is a product specific fixed effect, η_{kw}^{f} is a market-time fixed effect, and ω_{jkw}^{f} is an IID random shock. Since there are no endogeneity concerns in (33), I simply estimate by OLS.

Figure J.2, in the appendix, presents a histogram of the implied marginal costs of every jkw in the sample, obtained using demand specification (4). Less than 0.003% of these marginal costs are negative. The average marginal cost is roughly 0.34 dollars per serving.

Marginal cost in this estimation includes costs of raw materials and other inputs, labor, packaging, transportation, and retailer margin. Regarding retailer margins and, in particular, the double marginalization problem, the implicit assumption is that retailers charge a markup that is independent from the wholesale price. Alternatively, one could assume that retailers charge a common margin to all of their cereals, so that the retail price is $(1 + r)p_j$, where p_j is the wholesale price chosen by the producer and r is the retailer's margin. It is easy to see that this is equivalent, from the firms perspective, to a model where the price sensitivity parameter is $\lambda' = (1 + r)\lambda$, so all results showed in the previous sections will still hold. Regarding estimation, the 2SLS estimate $\hat{\lambda}$ would be consistent for $(1 + r)\lambda$ rather than λ, however, predicted retail prices, market shares, and profits will be correct, and the predicted consumer surplus would contain consumer surplus and retailers' profits.

It is worth noting that there is an extensive literature trying to measure deviations from competitive pricing by searching for the conduct parameters that better predict pricing behavior given demand estimates, i.e., by attempting to infer the ownership matrix implied by observed pricing behavior. In this line, Backus et al. (2018) use the same data as I do to study the effect of common ownership on price competition by trying-out different ownership structures. They find that "classic competition is most consistent with demand and supply data", supporting my behavioral assumption, but that "the implied effects of common ownership" are large relative to mergers. Also, Miller and Weinberg (2017) use the conduct parameter approach to study the effect on pricing of the joint venture between Coors and Miller in the beer industry. They find that the observed price increase after the merger is larger than what can be explained by a pure Nash-Bertrand competition, implying that part of the price increase is the result of a deviation from static pricing.

To the extent that coordinated effects, arising from common ownership or mergers, play a significant role, the pricing model proposed here will be misspecified, and merger simulations will be biased. Allowing for coordinated pricing, however, is inconsistent with the aggregative result that is at the heart of this paper. Rather than attempting to move towards some form of coordinated pricing, I simply acknowledge this shortcoming and move on under the assumption that firms price competitively pre and post-merger.

7 Parametrization and Estimation - Dynamic Game

As mentioned above, I use demand and marginal cost estimates to compute variable profits and use NFXP to estimate the parameters embedded in FC and AC taking as given the variable profits. In what follows, I parametrize functions FC and AC, and explain how I compute variable profits. Additionally, I make explicit the econometric assumptions underlying the estimation procedure.

7.1 Fixed Cost and Adjustment Cost Parametrization

I begin by recalling that, in light of proposition 3, I am re-interpreting all function as depending on bundle sizes, n^f, rather than bundles themselves, understanding that firms will always choose anonymous bundles, as defined in expression (20)

With that said, I move forward to define the firm's feasible set. I assume that firms can adjust the number of products they introduce to the market by no more than Δ each period, so the feasible set given n_t^f is the set:

$$C(n_t^f) \equiv \left\{ n : \text{abs}(n - n_t^f) \leq \Delta, \;\; 0 \leq n \leq |\mathcal{N}^f| \right\}$$

where abs$(\cdot)$ represents the absolute value operator. In estimation, I group products in bins of size 5, and set $\Delta = 2$, meaning that firms can introduce/remove as many as 10 products from one period to the next (for more details see subsection 7.2).

The adjustment cost is proportional to the change in n^f and potentially asymmetric around zero. This asymmetry accounts for the fact that it might be easier to remove products from the market than to introduce them.

$$AC^f(n_{q+1}^f, n_q^f) = \begin{cases} \theta_{AC+}^f (n_{q+1}^f - n_q^f) & \text{if} \quad n_{q+1}^f - n_q^f > 0 \\ 0 & \text{if} \quad n_{q+1}^f - n_q^f = 0 \\ \theta_{AC-}^f (n_{q+1}^f - n_q^f) & \text{if} \quad n_{q+1}^f - n_q^f < 0 \end{cases} \tag{34}$$

This adjustment cost is meant to capture the difficulties firms face when changing their product offering in one market. It includes things like the cost of adjusting logistics and negotiations with retailers.

Finally, regarding fixed costs I follow the literature on product-line length (Draganska and Jain (2005), Draganska and Jain (2006), Kadiyali et al. (1999)) and approximate FC by a second order polynomial:

$$FC^f(N^f) = \theta^f_{FC0} + \theta^f_{FC1} N^f + \theta^f_{FC2} {N^f}^2 \tag{35}$$

Clearly θ^f_{FC0} cannot be identified, so it its normalized to zero. A consistent finding of this literature is that $\theta_{FC2} > 0$, meaning that there are decreasing returns to scale in the number of products.

7.2 Variable Profits, Bundle Size and Product Menu

As mentioned before, to reduce the computational burden of the estimation of the dynamic parameters I pre-compute variable profits using demand and marginal cost estimates of each firm in every contingency, i.e. $(\pi^f_{kq}(n_{kq}, \tau_{kq}))_{\forall N_{kq},k,q,f}$, and use them as an input to the dynamic model. To do this, I specify here how observations in the sample, and demand-marginal cost estimates, are mapped into the dynamic model.

First, I use static estimates to construct a time-invariant product type for each product in my sample as follows:[18]

$$\bar{\tau}^f_j = \bar{\alpha}_j - \lambda \bar{\eta}_j$$

Then, for each firm-year, I rank products observed across all markets in the data according to their time invariant type and assign each an index equivalent to its ranking in decreasing order, i.e. to the product with the largest τ_j I assign $j = 1$, to the second largest I assign $j = 2$, etcetera. To maintain the state space manageable, I group products into bins of five contiguous products (in terms of ranking), so that bin b contains products $5b + 1, \dots, 5b + 5$. With some abuse of notation define the type of bin b for firm f as:

$$\bar{\tau}^f_b \equiv \sum_{j=5b+1}^{5(b+1)} \bar{\tau}^f_j$$

Note that, by construction, the ranking of bin types is: $\tau^f_1 > \tau^f_2 > \dots > \tau^f_{B^f}$, where B^f is the number of bins into which f's products where grouped.

The collection of binds constitutes f's product menu, and f bundle type for a bundle of size n^f, corresponds to the partial sum from $b = 1$ to $b = n^f$ of τ^f_{bkt}, multiplied by the firm-market-date shifters. This is:

$$\hat{\tau}^{n^f}_{kq} = \left(\sum_{b=1}^{n^f} e^{\bar{\tau}^f_b / \gamma^f} \right) \times e^{\bar{\tau}^f_{kq}/\gamma^f}$$

where:

$$\bar{\tau}^f_{kq} = \bar{\alpha}^f_{kq} - \lambda \bar{\eta}^f_{kq}$$

[18] Recall that the demand specification used in estimation has a single nest, so γ^f is set to 1 for all f.

Here I implicitly assume that the IID component of product type, $\tilde{\tau}^f_{kq}$, as defined in expression (29), is simply 0. This is equivalent to assuming that a firm chooses its bundle size assuming that idiosyncratic demand and marginal cost shocks are zero.

Using $(\hat{\tau}^{n^f}_{kt})_{\forall f,n,k,t}$ I solve the static game for each possible state $(n^f_{kt})_{f\in F}$ for each k and t, and calculate the corresponding profits (which I plug into the dynamic estimation), markups, market shares, and consumer welfare.

For the sample value of n^f_{kq}, I take the average across all weeks of quarter q of the number of distinct products provided by f in market k, and I set n^f_{kq} equal to the bin number where this value lies. Finally, I set a lower bound on n^f equal to the minimum number of products that firm f is observed to have put in a market-quarter over the whole sample, and an upper bound equal to the maximum number of products that f is observed to have in a market-quarter.

7.3 Model Solution and Estimation

I estimate the model by NFXP, using as input the variable profits implied by demand and marginal cost estimates obtained in section 6. To do this I pre-compute the variable profits at each possible contingency $(n_{kq})_{\forall k,q}$ (see subsection 7.2 for details on this procedure), and make assumptions. on the nature of the transition probabilities and the game.

Before specifying my econometric assumption, it is useful to define a firm's Conditional Choice Probability (CCP). Firm f's CCP is the probability that it chooses bundle n^f_{q+1} given the public state x_q. This is:

$$CCP^f_q(n^f_{q+1}|x_q) \quad = \quad \int_{\xi^f_q} \mathbf{1}\left\{\sigma(x_q,\xi^f_q) = n^f_{q+1}\right\} d\mathcal{F}(\xi^f_q|x_q) \tag{36}$$

With this in hand, I make the following assumptions:

Assumption 4. *Assume the following holds:*

a) The data of each market comes from an AMS.

b) Conditional Independence (CI). Each firm's Markov transition kernel satisfies:

$$\begin{aligned} P^f_q(n^{-f}_{q+1},\tau_{q+1},\xi^f_{q+1}|n^f_{q+1},n_q,\tau_q,\xi^f_q) \quad &= \quad \mathcal{F}(\xi^f_{q+1}|n^f_{q+1},n^{-f}_{q+1}\tau_{q+1}) \\ &\times \quad \prod_{g\neq f} CCP^g_q(n^g_{q+1}|x_t) \\ &\times \quad P_q(\tau_{q+1}|\tau_q) \end{aligned}$$

c) ξ^f_t is Extreme Value (EV):

$$\mathcal{F}(\xi^f_q|n_q) = \prod_{n^f_{q+1}\in C(n^f_q)} e^{-\exp(-\xi^f_q(n^f_{q+1}))}$$

d) Firms alternate moves in a predetermined order. The firm moving at time q is denoted by

$$f_q = q \mod F$$

e) The game has finite time, with terminal date Q and terminal values:

$$V_Q^f(x_Q) = \frac{\pi^f(x_Q) - FC\left(n_Q^f\right)}{1-\beta} \tag{37}$$

f) Firms can perfectly forecast the evolution of demand and marginal cost parameters (except for the idiosyncratic shocks ϵ and ω). This is:

$$P_q(\tau_{q+1}|\tau_q) = 1 \tag{38}$$

Along the predetermined path of τ_q.

g) Every market $k = \{1, \ldots, K\}$ is an independent realization of the dynamic game.

The importance of assumption a) is straightforward; the whole of the estimation and solution of the model rests on it. Assumptions b) and c) are standard in the literature. Intuitively, CI says that conditioning on the present state, all past information is irrelevant for forecasting current and future choice specific shocks. This assumption would fail if, for example, introducing a product during t makes it more expensive, on average, to introduce a product on $q+1$.

Assumptions e) and d) are not fundamental for estimation, but they considerably reduce computational burden. In particular, finite time allows me to estimate the model using an NFXP algorithm, as it allows me to solve it by backward induction. Furthermore, it guarantees equilibrium existence making the simulation of counterfactual experiments simpler.

Assumption d), while not fundamental for estimation, considerably simplifies the model solution. Furthermore, the alternate move assumption is quite innocuous in this context. Researchers avoid asynchronous move assumptions for fear of imposing an asymmetry among players that might contaminate the estimates. Such asymmetry stems, for example, from the deterring effect that an early entrant would have over its potential competitors in an entry game. If a firm is assumed to move early when, in reality, firms move simultaneously, the estimation algorithm will tend to rationalize the relatively low propensity to enter the market for an early mover by imputing a large fixed cost. Similarly, the fixed cost of the second mover would be underestimated to rationalize what would look like an excessive propensity to enter for the game's follower.

Two considerations, however, make me argue that this concern is not important in the context of this paper. First, I take the initial state of the market as given, so I impose no asymmetry in the process of arriving at the current market structure. Second, the deterrent effect of introducing a product in a game where firms hold many products at any given time is unimportant compared to the deterrent effect of being the incumbent on a market. Furthermore, since the game involves a relatively long time horizon, the benefit of having an early move should dissipate.

Accommodating merger counterfactuals into the model is not straightforward under alternate moves since one has to decide what happens to the "dead date" that is left when two firms merge. To deal with this, I assume the game is played quarterly and that each quarter is divided into F stages. At each stage, one, and only one, firm gets to adjust its product offering, and firms obtain market profits equivalent to $1/F$'th of the quarterly profits they would obtain under the stage's initial product offering. The move cycles once every quarter among the F firms in the same predetermined order. Firm f's quarterly profits are:

$$\frac{1}{F}\sum_{f_q=1}^{F}\Big(\pi^f(n_{q+1}^{\le f_q}, n_q^{>f_q}, \tau_q) - \mathbf{1}\{f_q < f\}FC^f(n_q^f) - \mathbf{1}\{f_q \ge f\}FC^f(n_{q+1}^f)\Big) - AC^f(n_{q+1}^f, n_q^f) \quad (39)$$

where $n^{>f_q} = (n^f)_{f>f_q}$ and $n^{\le f_q} = (n^f)_{f\le f_q}$. This approach has the advantage that the model adjusts naturally to mergers by decreasing by the number intra quarter stages by one, and increasing stage profits proportionally.

Assumption f) is not fundamental but allows me to reduce the computational burden of estimation. I could relax this assumption by introducing an underlying Markov process that affects the evolution of τ_{kq}^f in a way that is common to every firm. This would be equivalent to adding a market level first-order autoregressive component to the demand shock ϵ_{jkq}.

Finally, assumption g) states that the actions firms choose in one market have no predictive power over their actions on other markets. This assumption would fail, for example, if my empirical market definition does not match the relevant definition used by the firms, or if there are cross-market economies of scale internalized by the firm.

Given the alternate moves assumption, the transition of n_t to n_{t+1} is described by:

$$n_{q+1}^f = \begin{cases} n_q^f & \text{if } \; f \ne f_q \\ \sigma_q^f(x_q, \xi_q^f; \theta) & \text{if } \; f = f_q \end{cases} \quad (40)$$

hence, I define the probability that f chooses n_{q+1}^f, conditional on state n_q, dubbed Conditional Choice Probability (CCP) in the literature, as:

$$CCP_q^f(n_{q+1}^f, x_q; \theta) \equiv \begin{cases} 1\{n_{q+1}^f = n_q^f\} & \text{if } \; f \ne f_q \\ \displaystyle\int_{\xi_q^f} \mathbb{1}\left\{\sigma_q^f(n_q^f, \xi_q^f; \theta) = n_{q+1}^f\right\} d\mathcal{F}(\xi_t^f | x_q) & \text{if } \; f = f_q \end{cases} \quad (41)$$

The transition probability at time t, from state n_t to state n_{t+1} is given by the products of all the individual CCPs. Using CI and EV, this corresponds to:

$$CCP_q(n_{q+1}|x_q;\theta) \quad = \quad \prod_{f\neq f_q} 1\{n_{q+1}^f = n_q^f\} \frac{\exp\left\{\frac{\Pi^{f_q}(n_{q+1}^{f_q},x_q)+\beta EV_q^{f_q}(n_{q+1}^{f_q},x_{q+1})}{\theta_{scale}^f}\right\}}{\sum_{\tilde{n}} \exp\left\{\frac{\Pi^{f_q}(\tilde{n},x_q)+\beta EV_q^{f_q}(\tilde{n},x_{q+1})}{\theta_{scale}^f}\right\}} \tag{42}$$

With the transition probabilities in place, we are in a position to define the elements of the likelihood function. A sample for market k is a sequence of bundle sizes for each firm, $(n_{k,q}^f)_{f\in F, q=0,\dots,Q}$. Since the sample is generated by a Markov process, its joint probability can be written as the product of the probabilities of each transition observed in the data. The log-likelihood of market k can be written as:

$$l_k\left(\theta\right) \equiv \sum_{q=0}^{Q-1} \log\left(CCP_q(n_{k,q+1}|x_{k,q};\theta)\right) \tag{43}$$

Finally, by the independence across markets assumption, the full log-likelihood function can be written as the sum of the market individual log-likelihoods.

$$L(\theta) \equiv \sum_{k=1}^{K} l_k(\theta) \tag{44}$$

The Maximum Likelihood Estimator, $\hat{\theta}$ corresponds to the maximizer $L(\theta)$, and has all the well-known properties of these class of estimators. It's variance matrix is:

$$\Sigma \equiv E\left[\nabla_\theta L(\theta)\ \nabla_\theta L(\theta)'\right] \tag{45}$$

where $\nabla_\theta L(\theta)$ represents the gradient of the likelihood with respect to θ. Since I have no closed form solution for the gradient I compute these derivatives numerically by introducing a small perturbation to $\hat{\theta}$.

Since NFXP requires me to solve the model for each iteration of the numerical optimizer, speed in solving the model can make the difference between what is feasible and what is not. I take several steps to speed up computation. First, I wrote down the code that solves the model and evaluates the log-likelihood in a C language routine, faster than MATLAB of Python by a factor as large as 1,000, embedded in a standard MATLAB numerical optimizer. Second, I parallelize the solution algorithm by noting that the solution to the game played in any one market, does not need inputs from the solution in another market so that they can be solved independently. Depending on the hardware and the implementation, this process yields speed gains between 8 and 32 times. Third, grouping products in bins of size 5 reduces the state space by a factor of 5^F.

7.4 Dynamic Model Results

To estimate the model I set the value of the quarterly discount factor to $\beta = 0.985$, and the number of firms to $F = 5$, with four firms representing the four most prominent players in the market, and a fifth firm aggregating all remaining product (accounting for roughly 20% of the market) into a single producer. To ameliorate the concern that estimation is contaminated by the dynamics of the last few quarters leading to the end date, I solve the model for the full sample period but drop the last eight quarters of CCPs when calculating the likelihood function (I estimated the model using the full sample and results remained largely unchanged).

Table 5 shows the estimates of the dynamic model obtained using demand specification 4. Profits are measured in millions To explore the fit of the model I simulate every market 100 times

Table 5: Dynamic Model Estimation

Firm	θ_{FC1}	θ_{FC2}	θ_{AC-}	θ_{AC+}	θ_{scale}
F0	0.008***	0.002***	0.111	0.051***	1.2432***
	(0.00001)	(0.00000)	(0.0029)	(0.0036)	(0.0171)
F1	0.034***	0.010***	0.045	0.022	1.0784***
	(0.00008)	(0.00000)	(0.0220)	(0.0227)	(0.0119)
F2	0.040***	0.012***	0.059	0.029	1.4370***
	(0.00002)	(0.00000)	(0.0358)	(0.0375)	(0.0263)
F3	0.017***	0.007***	0.033**	0.016	1.4667***
	(0.00001)	(0.00000)	(0.0163)	(0.0151)	(0.0217)
F4	0.031***	0.050***	0.009***	0.004***	1.3606***
	(0.00061)	(0.00026)	(0.0028)	(0.0019)	(0.0427)

and subtract the average number of products in each market's simulations, from the actual number of products observed in the data, i.e.

$$d_k^f = T^{-1} \sum_t \left(n_{k,t}^f - S^{-1} \sum_{s=1}^{S} \tilde{n}_{k,t,s}^f \right)$$

where $n_{f,t}^f$ is the observed number of products that f had in. market k at quarter t, and $\tilde{n}_{k,t,s}^f$ is the corresponding value from simulation s.

The results of this exercise are presented in table 6. Recall that the number of products were grouped in bins of 5 for each firm, hence, a 0.1 difference in. table 6 represents a difference of 1 product. The model systematically under predicts the number of products $F0$ keeps in each market, as reflected in a positive d_k^0. For the rest of the firms, however, the model seems to provide a reasonable fit along this dimension.

Figure 7 presents the the evolution of the number of products for retailer/market 0, averaged across all simulations, for the first 24 quarters in the data (year 2007 through 2012). In terms of levels, the model does a reasonable job of explaining the number of varieties for GML, KEL, and PST. For QKR, the model seems to over predict the number of varieties that

Table 6: Model Fit by Market

Market	F0	F1	F2	F3	F4
0	0.255	-0.130	0.098	0.325	-0.018
1	-0.076	0.088	-0.064	-0.337	0.208
2	0.823	-0.398	-1.070	-0.695	-0.224
3	0.555	-0.369	0.149	0.051	-0.339
4	1.010	-0.211	-0.574	-1.109	-0.517
5	0.622	0.034	-1.392	-1.193	-0.417
6	1.690	0.023	-0.719	0.756	0.183
7	0.250	0.266	-0.673	-0.860	0.718

Figure 7: Dynamic Fit: Market 0

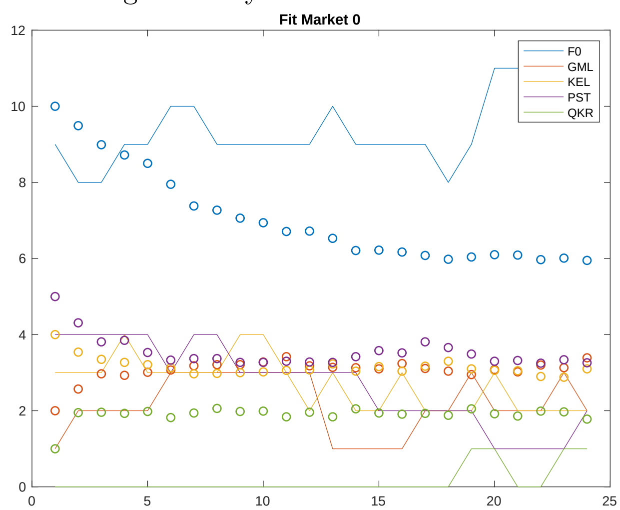
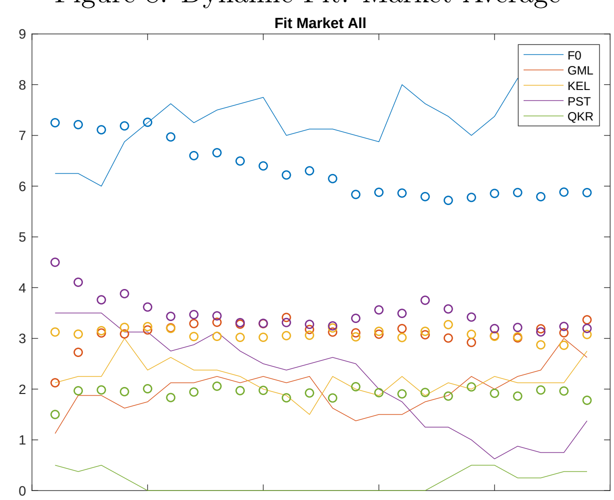


Figure 8: Dynamic Fit: Market Average



8 Counterfactual Analysis

I perform three main counterfactual experiments. First, to highlight the importance of endogeneity in merger assessments, I compare two hypothetical merger scenarios. One in which the merged entity and its competitors have to stick to their pre-merger product choice policy and can only change prices, and another in which both the product policy and prices can change after the merger. The impact of product endogeneity, then, is measured by the differences in market outcomes between these two policies.

Second, to asses the impact of a merger over welfare, I simulate a merger between two firms and compare the market outcomes to a situation with no merger. And third, to asses that dynamic strategic incentives play on this market, I remove dynamic strategic incentives by fixing the strategy of one firm.

8.1 Exogenous Post-Merger CCPs

To assess the importance of taking into account product endogeneity when assessing mergers, I simulate a merger between $F1$ and $F4$ under two assumptions. First, I simulate a merger with *pre-merger CCPs*, in which firms adjust their pricing policies to the new ownership structure, but introduce-remove products from the market according to the as dictated by the equilibrium without the merger. Second, I simulate a merger with *post-merger CCPs*, where firms adjust both their pricing policies and their CCPs to a new post-merger equilibrium.

To simulate a merger, I create a new entity m whose product menu consists of the union of the product menus of the individual merging firms. Say f and g are merging to form m, then the product menu of the merged entity is $\mathcal{N}^m = \mathcal{N}^f \cup \mathcal{N}^g$, leaving unchanged the demand and marginal cost parameters of the products of the merging firms. Using the newly merged entity and the non-merged firms, I recompute markups, profits, and consumer surplus from the static equilibria, and conduct simulation with fixed pre-merger CCPs. Then, I solve the dynamic model under the new ownership structure and simulate the model using post-merger CCPs. I account for efficiency gains on the fixed cost by allowing the merged entity to combine its independent fixed cost functions optimally. So, the fixed cost of the merged entity is:

$$FC^m(n^m) = \min \left\{ FC^f(n^f) + FC^g(n^g) \text{ subject to } n^f + n^g = n^m \right\}$$

This approach to modeling efficiency gains is akin to the one proposed by Perry and Porter (1985), and later McAfee and Williams (1992), for mergers under Cournot assumption.

Figure 9 shows the averaged result of 100 simulations of each market, comparing several market outcomes in the case where post-merger conditional choice probabilities are endogenous, i.e., firms adjust their strategy to the new economic environment, to a situation where firm only adjust their pricing strategies but keep their product policy invariant. For any particular variable, say n^f, the figure plots the values of:

$$\frac{\frac{1}{SK}\sum_{s,k} n^f_{k,t,s}(\text{Post-Merger CCPs})}{\frac{1}{SK}\sum_{s,k} n^f_{k,t,s}(\text{Pre-Merger CCPs})}$$

for every firm across time. In essence, what I show here is the evolution of an index for each variable; if the index is greater than 1, then the Post-Merger CCP simulations predict a larger value of the corresponding variable than the simulation with Pre-Merger CCPs.

Figure 9: Market Outcomes. Pre-Merger vs. Post-Merger CCPs

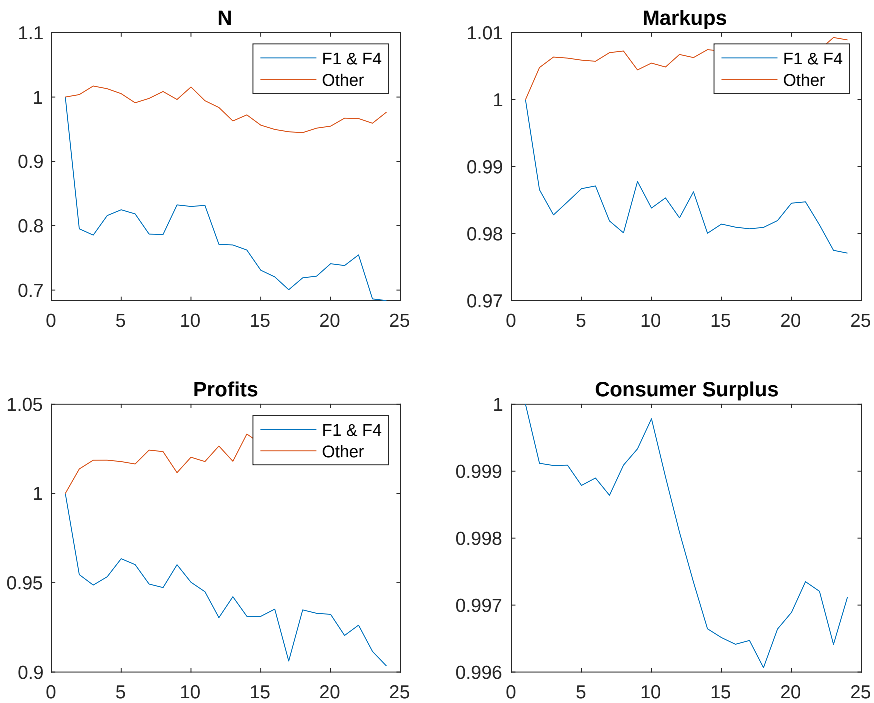


Figure 9 shows that the simulations with post-merger CCPs predict a substantially lower level of product variety. After two and a half years, these suggest that the merged entity supplies roughly 30% fewer varieties than in the pre-merger CCPs simulations. The removed varieties, however, are the lower end of the product type distribution, so not enough demand is freed in this process to justify non-merged competitors to introduce products of their own. This is reflected by the fact that they do not respond to m's jettisoning of products by adding products of their own.

As a result of the product removal in the post-merger CCP simulation, the merged entity charges markups 3.5% lower than in the pre-merger CCPs simulation. In other words, the price increase triggered by the merger is somewhat mitigated by the fact that the merged firm removes products from the market. In contrast, the non-merged firms increase their markups between 1 and 2% more, as compared to the pre-merger CCP simulations. Similarly, the merged firm's (variable) profits are a 12% lower when post-merger CCPs are used, and the non-merged firms have 5% higher profits. Table 7 presents the average ratio for each outcome, starting two years after the merger.

Table 7: Average Post-merger CCPs/Pre-merger CCPs

Firm	N	MU	PI
'F1-F4'	0.75	0.98	0.93
'F0'	0.96	1.01	1.02
'F2'	0.10	1.01	1.04
'F3'	1.00	1.01	1.04

8.2 Merger vs. No merger

Here I present simulations analogous to the ones shown in the previous subsection, except now I compare the case when firms 1 and 4 merge, to the case when no merger takes place. The aim is to give a sense of what an actual merger evaluation would look like. To simulate the merger, I follow the same procedure described in the previous subsection.

The results of these simulations can be found in figure 10. The first (top-left) panel shows the evolution of the number of products for each firm. This panel matches the results shown in figure 9, because the same CCPs are being compared in both experiments. The top-right panel shows a jump in markups at the time of the merger $t = 0$, followed by a mild decline as the merged entity removes products from the market. The profits of the merged firm exhibit a pattern similar to the previous simulation. It should be noted that the merger between firms 1 and 4 is particularly unattractive because firm 4 is very inefficient regarding FC, so there is no efficiency gain for firm 1 from the merger. As s result, when the merger occurs, the bundle of the merged firm increases drastically, and so does her fixed costs.

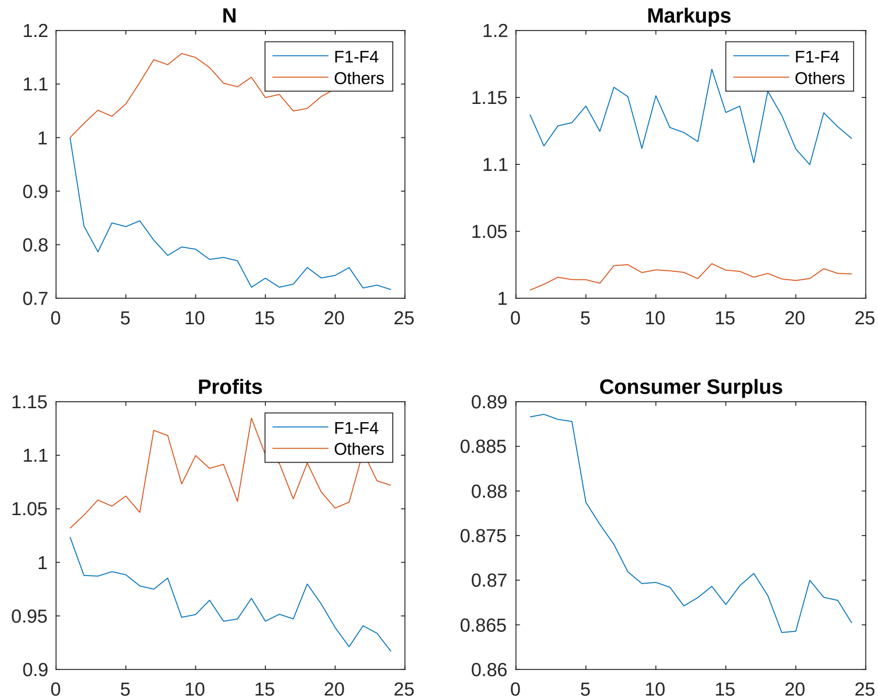
The most interesting result is the comparison in consumer surplus. Following the merger, consumer surplus drops discretely by 11%, as a result of the price increase. A further 3% drop over the next ten quarters can be attributed to the adjustments in product variety and its effects on prices that take place over this period. In other words, product endogeneity explains roughly 19% ($\approx 2.5/(10 + 3)$) of the total decrease in consumer surplus.

8.3 Product Proliferation

Schmalensee (1978) observed the RTE-cereal industry of the 60s and 70s was dominated by four large firms that controlled 80% of the market, and although they enjoyed high profits, no significant entry of new competitors occurred during this period. He explained this by arguing that firms were building artificial barriers through "the practices of proliferating brands, differentiating similar products, and promoting trademarks through intensive advertising."

I test for the presence of these strategies, which I gather under the umbrella of *product proliferation*, by simulating the model while fixing the $F0$'s (fringe firms) actions to what the path that is observed in the data. By doing this, I eliminate the dynamic incentive to flood the market with products to hinder the introduction of new products, so if large firms regard product proliferation an important concern when deciding on their strategies in the strategies on large firms, under this counterfactual,

Figure 10: Merger vs. No Merger



we should observe that they introduce fewer/take out products of the market.

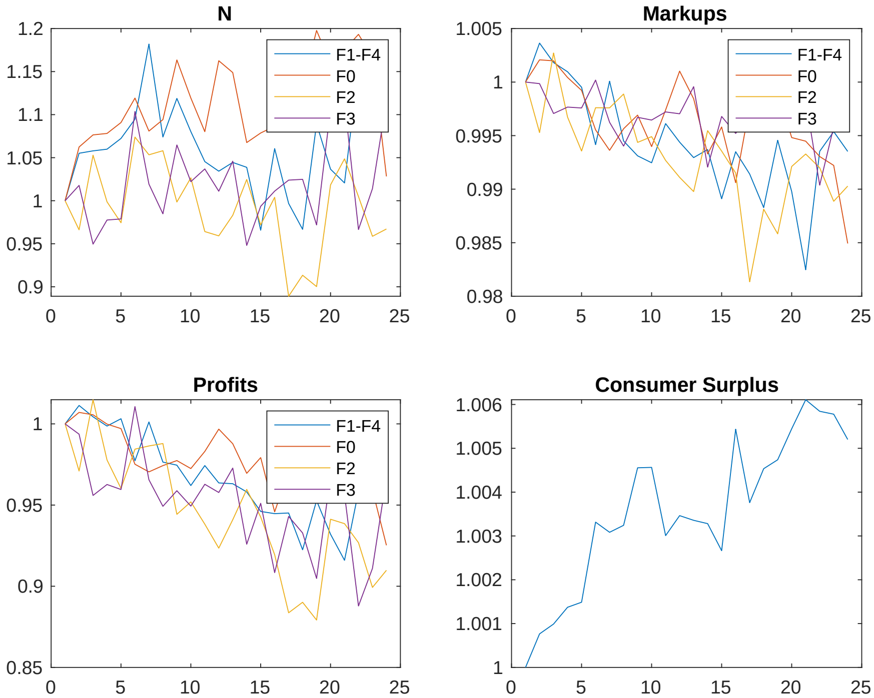
Figure 11 presents the results on this counterfactual, relative to the standard case. The simulations do not support the product proliferation hypothesis. Firms react in different directions, but the total magnitude appears small, product variety increases for firm 2 roughly 7%, while it decreases for firm 3 by a similar amount. If anything, the market appears to become more competitive under this equilibrium, as reflected by the fact that both markups and profits tend to decrease.

9 Closing Remarks

This paper proposes a flexible model through which merger evaluations can be conducted, allowing firms to rebalance their post-merger product offering. The model is flexible in that it is not overly idiosyncratic, and it can be transferred to other industries with relative ease. Furthermore, I show that in equilibrium, only a small subset of all possible bundles is played, which reduces the computational complexity of the model, and allows for applications with many more products per firm than previous research has.

I estimate the model using data from the RTE-cereal industry, and conduct several counterfactual experiments. Results show that after the merger, the merged entity cuts product variety by 30% over two and a half years. As a result, non-merged firms raise their prices more than what can be explained by the change in ownership structure alone, and amelioration in the price increase of the merged entity. These effects result in an additional loss of consumer surplus of roughly 19%.

Figure 11: Open Loop Equilibrium



The model can be extended in several directions, which would allow its application to study different questions. A particularly important one is to endogenize product development and use the model to study the effects of market structure on the rate of new product release. The interplay between market structure and innovation has received increasing attention in recent years, as evidence of increased market power in the U.S. economy has fueled voices calling for stricter enforcement of merger policy. The extent to which this could negatively impact the generation of new products, however, is not well understood and deserves a much closer examination before any substantive policy changes are made.

References

Aguirregabiria, V. and C.-Y. Ho (2010). A dynamic game of airline network competition: Hub-and-spoke networks and entry deterrence. *International Journal of Industrial Organization 28*(4), 377–382.

Aguirregabiria, V. and P. Mira (2007, January). Sequential estimation of dynamic discrete games. *Econometrica 75*(1), 1–53.

Anderson, S. P. and A. de Palma (1992). The logit as a model of product differentiation. *Oxford Economic Papers 44*(1), 51–67.

Backus, M. R., C. Conlon, and M. Sinkinson (2018). Common ownership and competition in the ready-to-eat cereal industry.

Bajari, P., C. L. Benkard, and J. Levin (2007, August). Estimating dynamic models of imperfect competition. *Econometrica 75*(5), 1331–1370.

Barwick, P. J., M. Kalouptsidi, and N. B. Zahur (2019, July). China's industrial policy: an empirical evaluation. Working Paper 26075, National Bureau of Economic Research.

Benkard, C. L., P. Jeziorski, and G. Y. Weintraub (2015). Oblivious equilibrium for concentrated industries. *The RAND Journal of Economics 46*(4), 671–708.

Benkard, L., A. Bodoh-Creed, and J. Lazarev (2014, April). The Long Run Effects of U.S. Airline Mergers. Research Papers 3005, Stanford University, Graduate School of Business.

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

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

Ciliberto, F., C. Murry, and E. Tamer (2018, Nov). Market structure and competition in airline markets. *Available at SSRN: https://ssrn.com/abstract=2777820 or http://dx.doi.org/10.2139/ssrn.2777820*.

Collard-Wexler, A. (2013, May). Demand fluctuations in the ready-mix concrete industry. *Econometrica 81*(3), 1003–1037.

DellaVigna, S. and M. Gentzkow (2017, November). Uniform pricing in us retail chains. Working Paper 23996, National Bureau of Economic Research.

Doraszelski, U. and M. Satterthwaite (2010). Computable markov-perfect industry dynamics. *The RAND Journal of Economics 41*(2), 215–243.

Draganska, M. and D. C. Jain (2005). Product-line length as a competitive tool. *Journal of Economics & Management Strategy 14*(1), 1–28.

Draganska, M. and D. C. Jain (2006). Consumer preferences and product-line pricing strategies: An empirical analysis. *Marketing Science 25*(3), 1–28.

Draganska, M., M. Mazzeo, and K. Seim (2009). Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *QME 7*(2), 105–146.

Federal Trade Commision and Department of Justice (2011). Hart-Scott-Rodino Annual Report.

Federal Trade Commision and Department of Justice (2015). Hart-Scott-Rodino Annual Report.

Federal Trade Commision and Department of Justice (2017). Hart-Scott-Rodino Annual Report.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. Cambridge, MA: MIT Press. Translated into Chinesse by Renin University Press, Bejing: China.

Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In *The Economics of New Goods*, pp. 207–248. National Bureau of Economic Research, Inc.

Hendel, I. and A. Nevo (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica 74*(6), 1637–1673.

Hitsch, G. J. (2006). An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Science 25*(1), 25–50.

Hitsch, G. J., A. Hortaçsu, and X. Lin (2017). Prices and promotions in u.s. retail markets: Evidence from big data.

Ifrach, B. and G. Y. Weintraub (2016). A Framework for Dynamic Oligopoly in Concentrated Industries. *The Review of Economic Studies 84*(3), 1106–1150.

Igami, M. (2015). Offshoring under oligopoly. *Working Paper*.

Igami, M. and K. Uetake (2019, 09). Mergers, Innovation, and Entry-Exit Dynamics: Consolidation of the Hard Disk Drive Industry, 1996–2016. *The Review of Economic Studies*. rdz044.

Jeziorski, P. (2014, 1). Estimation of cost efficiencies from mergers: application to us radio. *The RAND Journal of Economics 45*(4), 816–846.

Kadiyali, V., N. Vilcassim, and P. Chintagunta (1999). Product line extensions and competitive market interactions: An empirical analysis. *Journal of Econometrics 89*(1), 339 – 363.

Li, S. Y., J. Mazur, Y. Park, J. W. Roberts, A. Sweeting, and J. Zhang (2018, January). Endogenous and selective service choices after airline mergers. Working Paper 24214, National Bureau of Economic Research.

Maskin, E. and J. Tirole (2001). Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory 100*(2), 191 – 219.

Mazzeo, M. J., K. Seim, and M. Varela (2018). The welfare consequences of mergers with endogenous product choice. *The Journal of Industrial Economics 66*(4), 980–1016.

McAfee, R. P. and M. A. Williams (1992). Horizontal mergers and antitrust policy. *The Journal of Industrial Economics 40*(2), 181–187.

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

Miller, N. H. and M. C. Weinberg (2017). Understanding the price effects of the millercoors joint venture. *Econometrica 85*(6), 1763–1791.

Nevo, A. (2001, March). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Nevo, A. and F. Rossi (2008). An approach for extending dynamic models to settings with multi-product firms. *Economics Letters 100*(1), 49–52.

Nocke, V. and N. Schutz (2018). Multiproduct-firm oligopoly: An aggregative games approach. *Econometrica 86*(2), 523–557.

Nosko, C. (2010). Competition and quality hoice in the cpu market. *Unpublished Manuscript*.

Pakes, A. and P. McGuire (1994). Computing markov-perfect nash equilibria: Numerical implications of a dynamic differentiated product model. *The RAND Journal of Economics 25*(4), 555–589.

Perry, M. K. and R. H. Porter (1985). Oligopoly and the incentive for horizontal merger. *The American Economic Review 75*(1), 219–227.

Rust, J. (1987, September). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica 55*(5), 999–1033.

Rust, J., S. Cho, G. Lee, and M. Yu (2018). Optimal dynamic hotel pricing.

Sanderson, E. and F. Windmeijer (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics 190*(2), 212–221.

Schmalensee, R. (1978, Autumn). Entry deterrence in the ready-to-eat breakfast cereal industry. *Bell Journal of Economics 9*(2), 305–327.

Shapiro, B., G. Hitsch, and A. Tuchman (2018, 01). Generalizable and robust tv ad effects. *SSRN Electronic Journal*.

Sullivan, C. (2018). The ice-cream split: Empirically distinguishing price and product space collusion. *Working Paper*.

Sweeting, A. (2013, September). Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica 81*(5), 1763–1803.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.

Williams, K. R. (2017, August). Dynamic Airline Pricing and Seat Availability. Cowles Foundation Discussion Papers 2103, Cowles Foundation for Research in Economics, Yale University.

Wollmann, T. G. (2018, June). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review 108*(6), 1364–1406.

A Optimal Markup Derivation

By the nested logit assumption, the market share of product j is:

$$s_j(m_j, h^f, H) = \frac{e^{(\tau_j - \lambda m_j)/\gamma^f}}{e^{h^f}} \frac{e^{\frac{\gamma^f}{\rho} h^f}}{e^H} \frac{e^{\rho H}}{1 + e^{\rho H}} \tag{A.1}$$

where

$$h^f = \log\left(\sum_{j \in N^f} e^{(\tau^j - \lambda_j m_j)/\gamma^f}\right) \tag{A.2}$$

$$H = \log\left(\sum_{f \in F} e^{\frac{\gamma^f}{\rho} h^f}\right) \tag{A.3}$$

Profits are:

$$\pi^f((m_j)_{j \in N^f}, h^f, H) = \sum_{j \in N^f} m_j s_j(m_j, h^f, H) \tag{A.4}$$

Taking f.o.c. with respect to m_l:

$$0 = s_l + m_l \frac{\partial s_l}{\partial m_l} + \sum_{j \in N^f} m_j \left(\frac{\partial s_j}{\partial h^f} \frac{\partial h^f}{\partial m_l} + \frac{\partial s_j}{\partial H} \frac{\partial H}{\partial h^f} \frac{\partial h^f}{\partial m_l}\right) \tag{A.5}$$

The derivatives in the expression above are:

$$\frac{\partial s_l}{\partial m_l} = -\frac{\lambda}{\gamma^f} s_l \tag{A.6}$$

$$\frac{\partial s_j}{\partial h^f} = -\left(1 - \frac{\gamma^f}{\rho}\right) s_j \tag{A.7}$$

$$\frac{\partial s_j}{\partial H} = -s_j(1 - \rho s^0) \tag{A.8}$$

$$\frac{\partial h^f}{\partial m_l} = -\frac{\lambda}{\gamma^f} s_{l|f} \tag{A.9}$$

$$\frac{\partial H}{\partial h^f} = \frac{\gamma^f}{\rho} s^{f|F} \tag{A.10}$$

Where $s^0 = (1 + e^{\rho H})^{-1}$ is the market share of the outside option, $s_{l|f} = \frac{e^{\tau_l - \lambda m_l}}{e^{h^f}}$ is the market share of product l within firm f, i.e. the probability of choosing l given that firm f was chosen, and

$s_{j|F}$ is the market share of product j. Putting these terms together:

$$\begin{aligned}\frac{\partial s_j}{\partial h^f}\frac{\partial h^f}{\partial m_l} &= \left(1-\frac{\gamma^f}{\rho}\right)\frac{\lambda}{\gamma^f}s_j s_{l|f} \\ &= \left(1-\frac{\gamma^f}{\rho}\right)\frac{\lambda}{\gamma^f}\frac{s_j s_l}{s^f} \qquad \text{(A.11)}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial s_j}{\partial H}\frac{\partial H}{\partial h^f}\frac{\partial h^f}{\partial m_l} &= \left(-s_j(1-\rho s^0)\right)\times\left(\frac{\gamma^f}{\rho}s^{f|F}\right)\times\left(-\frac{\lambda}{\gamma^f}s_{l|f}\right) \\ &= (1-\rho s^0)\frac{\lambda}{\rho}\frac{s_j s_l}{s^F} \qquad \text{(A.12)}\end{aligned}$$

Summing the two:

$$\begin{aligned}\frac{\partial s_j}{\partial h^f}\frac{\partial h^f}{\partial m_l}+\frac{\partial s_j}{\partial H}\frac{\partial H}{\partial h^f}\frac{\partial h^f}{\partial m_l} &= \left(1-\frac{\gamma^f}{\rho}\right)\frac{\lambda}{\gamma^f}\frac{s_j s_l}{s^f}+(1-\rho s^0)\frac{\lambda}{\rho}\frac{s_j s_l}{s^F} \\ &= s_j s_l\frac{\lambda}{\rho}\left(\left(1-\frac{\gamma^f}{\rho}\right)\frac{1}{\gamma^f}\frac{\rho}{s^f}+\frac{1-\rho s^0}{s^F}\right)\end{aligned}$$

Replacing in (A.5), and dividing by s_l:

$$\frac{\lambda}{\gamma^f}m_l = 1+\frac{\lambda}{\rho}\left(\left(1-\frac{\gamma^f}{\rho}\right)\frac{1}{\gamma^f}\frac{\rho}{s^f}+\frac{1-\rho s^0}{s^F}\right)\sum_{j\in N^f}m_j s_j \qquad \text{(A.13)}$$

Since the RHS does not depend on the identity of l, it follows that for any l there is a constant m^f such that:

$$m^f = m_l, \forall l\in N^f \qquad \text{(A.14)}$$

Using $s_f=\sum_{s_j\in N^f}s_j$, and the common markup to replace in the f.o.c.:

$$\begin{aligned}m^f\frac{\lambda}{\gamma^f} &= 1+\frac{\lambda}{\rho}\left(\left(1-\frac{\gamma^f}{\rho}\right)\frac{1}{\gamma^f}\frac{\rho}{s^f}+\frac{1-\rho s^0}{s^F}\right)m^f s^f \qquad \text{(A.15)}\\ m^f\frac{\lambda}{\gamma^f} &= 1+m^f\frac{\lambda}{\rho}\left(\left(1-\frac{\gamma^f}{\rho}\right)\frac{\rho}{\gamma^f}+s^{f|F}(1-\rho s^0)\right) \qquad \text{(A.16)}\\ m^f\frac{\lambda}{\gamma^f} &= 1+m^f\frac{\lambda}{\rho}\frac{\rho}{\gamma^f}-m^f\frac{\lambda}{\rho}\frac{\gamma^f}{\rho}\frac{\rho}{\gamma^f}+m^f\frac{\lambda}{\rho}s^{f|F}(1-\rho s^0) \qquad \text{(A.17)}\\ m^f\frac{\lambda}{\rho} &= 1+m^f\frac{\lambda}{\rho}s^{f|F}(1-\rho s^0) \qquad \text{(A.18)}\end{aligned}$$

Using common markup on the firm inclusive value defined in (A.2):

$$\begin{aligned} h^f &= \log\left(\sum_{j\in N^f} e^{\tau_j - \lambda m^f}\right) & \text{(A.19)} \\ &= \log\left(\sum_{j\in N^f} e^{\tau_j}\right) - \lambda m^f & \text{(A.20)} \\ &= \tau^{N^f} - \lambda m^f & \text{(A.21)} \end{aligned}$$

Finally, substituting s^0 and $s^{f|F}$:

$$\frac{\lambda}{\rho} m^f = 1 + \frac{\lambda}{\rho} m^f \frac{e^{(\gamma^f \tau^{N^f} - \lambda m^f)/\rho}}{e^H} \left(1 - \frac{\rho}{1 + e^{\rho H}}\right) \tag{A.22}$$

B Equilibrium Uniqueness

Recall that the equilibrium H is determined by the solution to:

$$H = \log\left(\sum_{f\in F} e^{(\gamma^f \tau^{N^f} - \lambda m^f(H))/\rho}\right)$$

Taking the derivative of the right-hand side with respect to H:

$$\frac{\partial RHS}{\partial H} = \frac{\sum_{j\in N^f} e^{(\gamma^f \tau^{N^f} - \lambda m^f(H))/\rho}}{e^H} \frac{-\lambda}{\rho} \frac{\partial m^f}{\partial H}$$

From (C.5) below, $\frac{\partial m^f}{\partial H}$ is:

$$\frac{\partial m^f}{\partial H} = -\frac{\phi^f}{H}\left(1 - \rho^2 \frac{s^0(1-s^0)}{1-\rho s^0}\right)$$

where $\phi^f = \frac{{m^f}^2 - m^f}{{m^f}^2 - m^f + 1} \in [0,1]$. Substituting:

$$\frac{\partial RHS}{\partial H} = \left(1 - \rho^2 \frac{s^0(1-s^0)}{1-\rho s^0}\right) \sum_{f\in F} s^{f|F} \phi^f < 1$$

where the inequality comes from the fact that both $\sum_f s^f \phi^f$, lie in $[0,1]$.

C Proof of Proposition 1

Throughout this proof I denote with the subindex 0 (resp. 1), the equilibrium value of a variable after a firm f has introduced a new product.

Proof of (1,a)). Note that the introduction of a product by f, affects g only through H, hence the markup function $m^g(H)$, defined as the solution to (A.22), does not change. Let $\Gamma(H) = \log\left(\sum_{f\in F} e^{(\gamma^f \tau^{N^f} - \lambda m^g(H))/\rho}\right)$. An equilibrium aggregator H^* solves:

$$H^* = \Gamma(H^*) \tag{C.1}$$

Nocke and Schutz (2018) show that for a Nested Logit specification, the pricing game has a unique equilibrium, i.e. H^* is unique. Furthermore, they show in lemma J (p. 553) that for $H < H^*$

$$H < \Gamma(H) \tag{C.2}$$

For a given H, if f introduces a product it increases $e^{\tau^{N^f} - \lambda m^f(H)}$, and has no effect over $e^{\tau^{N^g} - \lambda m^g(H)}$ for $g \neq f$, It follows that $\Gamma_1(H) > \Gamma_0(H)$ for all H. At the original equilibrium market inclusive value, call it H_0^*, we have $H_0^* = \Gamma_0(H_0^*) < \Gamma_1(H_0^*)$, which implies, by lemma J, that the new equilibrium market inclusive value must be larger, $H_1^* > H_0^*$, as desired. □

Proof of (1,b)). Throughout this proof I will assume that $s_1^{g|F} < s_0^{g|F}$ for all $g \neq f$. I show that this is in fact the case when proving c) below.

Since $s^{g|F}$ decreases for all $g \neq f$, and $s^{f|F} = 1 - \sum_{g\neq f} s^{g|F}$, we have $s_1^{f|F} > s_0^{f|F}$. Additionally, since H increases, $s^F = \frac{e^{\rho H}}{1+e^{\rho H}}$ also increases, $s_1^F > s_0^F$. By definition, $s^f = s^{f|F} s^F$. Since both terms on the right hand side increased, $s_1^f > s_0^f$.

From (A.22):

$$\frac{\frac{\lambda}{\gamma^f} m^f - 1}{\frac{\lambda}{\gamma^f} m^f} = s^{f|F}(1 - \rho s^0)$$

Since $s^{f|F}$ increases, and s^0 decreases, the term in the right hand side must increase, which implies that the left hand side increases too:

$$\frac{\frac{\lambda}{\gamma^f} m_1^f - 1}{\frac{\lambda}{\gamma^f} m_1^f} > \frac{\frac{\lambda}{\gamma^f} m_0^f - 1}{\frac{\lambda}{\gamma^f} m_0^f}$$

which in turn implies $m_1^f > m_0^f$.

Finally $\pi_1^f > \pi_0^f$ follows immediately from $m_1^f > m_0^f$ and $s_1^f > s_0^f$. □

Proof of (1,c)). For the purpose of this proof, I re-interpret the markup m^f to $\frac{\lambda}{\rho} m^g$. It suffices to show that m^g is decreasing in H. Rearranging equation (A.22) for firm g.

$$\frac{m^g - 1}{m^g} = s^{g|F}(1 - \rho s^0) \tag{C.3}$$

Differentiating with respect to H and letting $m^{g\prime} = \frac{\partial m^g}{\partial H}$:

$$\frac{m^{g\prime}}{(m^g)^2} = -m^{g\prime} s^{g|F}(1-\rho s^0) - s^{g|F}(1-\rho s^0) + \rho^2 s^{g|F} s^F s^0$$

Using equation (C.3) and solving for $m^{g\prime}$:

$$\begin{aligned}
\frac{m^{g\prime}}{(m^g)^2} &= -m^{g\prime}\frac{m^g-1}{m^g} - \frac{m^g-1}{m^g} + \rho^2 s^{g|F} s^0(1-s^0) \\
m^{g\prime}\left(\frac{1}{(m^g)^2} + \frac{m^g-1}{m^g}\right) &= -\frac{m^g-1}{m^g} + \rho^2\frac{m^f-1}{m^f}\frac{s^0(1-s^0)}{1-\rho s^0} \\
m^{g\prime}\left(\frac{1+(m^g)^2-m^g}{(m^g)^2}\right) &= -\frac{m^g-1}{m^g} + \rho^2\frac{m^g-1}{m^g}\frac{s^0(1-s^0)}{1-\rho s^0} \\
m^{g\prime} &= -\phi^g\left(1-\rho^2\frac{s^0(1-s^0)}{1-\rho s^0}\right) \qquad \text{(C.4)}
\end{aligned}$$

where :

$$\phi^g = \frac{(m^g)^2-m^g}{(m^g)^2-m^g+1} \in [0,1]$$

A decreasing m requires the term in parenthesis to lie in the $[0,1]$ interval. To show that this is the case, note that this is equivalent to:

$$1-\rho s^0 - \rho^2 s^0(1-s^0) > 0$$

which is strictly decreasing in ρ for a fixed s^0 and $\rho > 0$, so the minimum value it can take is: $(1-s^0)^2 > 0$. Going back to the original interpretation of m^f:

$$m^{g\prime} = -\phi^g\frac{\rho}{\lambda}\left(1-\rho^2\frac{s^0(1-s^0)}{1-\rho s^0}\right) \qquad \text{(C.5)}$$

where:

$$\phi^g = \frac{(\frac{\lambda}{\rho}m^g)^2 - \frac{\lambda}{\rho}m^g}{(\frac{\lambda}{\rho}m^g)^2 - \frac{\lambda}{\rho}m^g + 1} \in [0,1]$$

To evaluate the effect over g's market share consider first, $s^{g|F} = \frac{e^{(\gamma^g \tau N^g - \lambda m^g)/\rho}}{e^H}$. Differentiating with respect to H:

$$\begin{aligned}
\frac{\partial s^{g|F}}{\partial H} &= -s^{g|F} - s^{g|F}\frac{\lambda}{\rho}\frac{\partial m^g}{\partial H} \\
&= -\frac{s^{g|F}}{H}\left(1-\phi^g\left(1-\rho^2\frac{s^0(1-s^0)}{1-\rho s^0}\right)\right)
\end{aligned}$$

By the result above, the term in the inner parenthesis lies in the interval $[0,1]$. This, together with $\phi^g \in [0,1]$ implies $\frac{\partial s^{g|F}}{\partial H} < 0$.

For the market share note that $s^g = s^{g|F} s^F$ and differentiate with respect to H:

$$\frac{\partial s^g}{\partial H} = \frac{\partial s^{g|F}}{\partial H} s^F + s^{g|F} \frac{\partial s^F}{\partial H}$$

note that $\frac{\partial s^F}{\partial H} = \frac{\partial}{\partial H} \frac{H^\rho}{1+H^\rho} = \frac{\rho}{H} s^F (1 - s^F)$, hence:

$$\begin{aligned} \frac{\partial s^g}{\partial H} &= -\frac{s^{g|F}}{H} \left[1 - \phi^g \left(1 - \rho^2 \frac{s^0(1-s^0)}{1-\rho s^0} \right) \right] s^F + s^{g|F} \frac{\rho}{H} s^F (1 - s^F) \\ &= -\frac{s^g}{H} \left[1 - \rho s^0 - \phi^g \left(1 - \rho^2 \frac{s^0(1-s^0)}{1-\rho s^0} \right) \right] \end{aligned}$$

Finally, the fact that g's profits decrease follows immediately from the. fact that both m^g and s^g decrease. □

D Proof of Proposition 2

Proof of Proposition Strict Dominance of. N_n^f. Without loss of generality, let index j rank products in $\mathcal{N}^f$ from largest to smallest type, i.e. $\tau_1 > \tau_2 > \ldots > \tau_{|\mathcal{N}^f|}$.

Fix an arbitrary bundle size n, and define reduced form profits, $\tilde{\pi}^f$ as:

$$\tilde{\pi}^f(\tau^{Nf}, \tau^{N^{-f}}) \equiv \pi^f(N^f, N^{-f})$$

From proposition 1 we know that increasing $\tilde{\pi}^f$ is increasing in τ^{Nf} for all N^{-f}, hence, finding the bundle of size n that maximizes π^f is equivalent to finding the bundle of size n that maximizes τ^{Nf}. Recalling that:

$$\tau^{Nf} = \log\left(\sum_{j \in N^f} e^{\tau_j/\gamma^f}\right)$$

it is clear that the only such bundle is:

$$N_n^f \equiv \{1, 2, \ldots, n\}$$

Since FC^f depends on N^f only through its size, $FC^f(N^f) = FC^f(N_n^f)$ for all N^f such that $|N^f| = n$. Put. together, these two observations imply:

$$\tilde{\pi}^f(\tau^{N_n^f}, \tau^{N^{-f}}) - FC^f(N_n^f) > \tilde{\pi}^f(\tau^{Nf}, \tau^{N^{-f}}) - FC^f(N^f)$$

for all $\tau^{N^{-f}}$, and all $N^f \neq N_n^f$, of size n. □

E Proof that AMS is a Best Response to AMS

Proof. Let us begin by refreshing the definition of N_n^f as used in the one-shot game. Bundle N_n^f is the bundle size n containing f's products of n largest product size.

There are two elements to an AMS:

1. It only depends on the vector bundle sizes.

2. It always chooses bundles that are stricly dominant in the two stage game as described in 3.

First, note that by product type decomposition assumption, the bundle that maximize static variable profits conditional on a bundle size is time invariant, and it is always equal to the strictly dominant bundle of the one shot game. This is:

$$N_n^f = N_{n,t}^f, \forall t$$

So, under an AMS, the n is a sufficient statistic for $N_{n,t}^f$ for all t.

Given this argument, if all of f's competitors follow an AMS it follows immediately that f's strategy will satisfy need to track only its competitors' bundle seizes n_t^{-f}. Intuitively, if f chooses an AMS then two things are true. First, f only needs to track vector n_t to forecast the behavior of $g \neq f$, and $|N_t^g|$ is a sufficient statistic for composition of g's bundle at every t, so it contains all the information it needs to make pricing decisions. Since this holds for all $g \neq f$, then keeping track of the tuple n_t^{-f} is enough to forecast the evolution of the industry. More formally, note that by definition of AMS, the conditional choice probability of $g \neq f$ can be written as:

$$CCP^g\left(N_{t+1}^g \middle| N_t, |N_{t+1}^g| = n^g\right) = \begin{cases} 0 & \text{if} \quad N_{t+1}^g \neq N_{n^g}^g \\ Pr\left(N_{t+1}^g = N_n^g | (|N_t^g|)_{g\in F}\right) & \text{otherwise} \end{cases} \quad \text{(E.1)}$$

where $(|N_t^g|)_{g\in F}$ is the vector of bundle sizes at t, and N_n^g is the optimal bundle of size n as defined in the two stage game. So, to predict N_{t+1}^g, for an $g \neq f$, f only needs to track $(|N_t^g|)_{g\in F}$. Furthermore, since by definition of AMS, all $g \neq f$ choose bundles of the N_n^f type, f also knows the composition of its competitors' bundles.

Now, it still could be the case that f's strategy depends on its own bundle rather than just its size. To show that this does not occur, I argue that any a strategy σ^f is weakly worse than a strategy σ^{f^*} satisfying $|\sigma^{f^*}(N_t^f, n_t^{-f})| = |\sigma^f(N_t^f, n_t^{-f})|$, and which always introduces (removes) the products with the highest (lowest) types first. To see this let $q^g(n_{t+1}^g|n_t) = Pr(N_{t+1}^g = N_{n^g}^g|N_t)$, and fix an strategy for firm f, σ^f which depends on (N^f, n^{-f}). For such strategy, f's discounted profit is:

$$\begin{aligned} EV_t^f(N_t;\sigma^f) &= \sum_{i=t}^{\infty}\beta^{i-t}\sum_{h\in\mathcal{H}_i(N_t)}\sum_{N_i\in C(h)}\prod_{g\neq f} q^g(N_i^g|N_{i-1})q^f(N_i^f|N_{i-1},\sigma^f) \\ &\times \; E_{\sigma^f}\left[\pi(N_i^f, N_i^{-f}) - FC^f(N_i^f) - AC^f(N_{i+1}^f, N_i^f) + \xi_t^f(N_{i+1}^f)\middle| N_i\right] \end{aligned} \quad \text{(E.2)}$$

where $\mathcal{H}_i(N_t)$ is the set of all possible histories from t to $i-1$, starting at state N_t, h is a particular history, $C(h)$ is the set of possible next period states given history $h \in \mathcal{H}_i(N_t), q^f(\cdot|\cdot, \sigma^f)$ is the CCP of firm f for strategy σ^f, and $E_{\sigma^f}[\cdot|N_i]$ is the expected period payoff given state N_i, and given strategy σ^f. Using the fact that $g \neq f$ only track bundles sizes, and that they only choose bundles of the type N_n^g, we can write:

$$\begin{aligned} EV_t^f(N_t;\sigma^f) &= \sum_{i=t}^{\infty} \beta^{i-t} \sum_{h\in\mathcal{H}_i(N_t)} \sum_{n_i^{-f}\in C(h)} \prod_{g\neq f} q^g(n_i^g|n_{i-1}) \sum_{N_i^f\in C(h)} q^f(N_i^f|N_{i-1}^f, n_{i-1}^{-f}, \sigma^f) \\ &\times \; E_{\sigma^f}\left[\pi(N_i^f, n_i^{-f}) - FC^f(N_i^f) - AC^f(N_{i+1}^f, N_i^f) + \xi_t^f(N_{i+1}^f)\middle| N_i\right] \end{aligned} \tag{E.3}$$

Where I write everything in terms of bundle sizes except N^f. The inner most sum can be decomposed into:

$$\sum_{N_i^f\in C(h)} = \sum_{n^f\in C(h)} \sum_{N_i^f:|N_i^f|=n^f}$$

where $n \in C(h)$ is short for $N_i^f \in C(h)$ satisfying $|N_i^f| = n$. So (E.3) can be written as:

$$\begin{aligned} EV_t^f(N_t;\sigma^f) &= \sum_{i=t}^{\infty} \beta^{i-t} \sum_{h\in\mathcal{H}_i(N_t)} \sum_{n_i^{-f}\in C(h)} \prod_{g\neq f} q^g(n_i^g|n_{i-1}) \sum_{n^f\in C(h)} \sum_{N_i^f:|N_i^f|=n} q^f(N_i^f|N_{i-1}^f, n_{i-1}^{-f}, \sigma^f) \\ &\times \; E_{\sigma^f}\left[\pi(N_i^f, n_i^{-f}) - FC^f(N_i^f) - AC^f(N_{i+1}^f, N_i^f) + \xi_t^f(N_{i+1}^f)\middle| N_i\right] \end{aligned} \tag{E.4}$$

Note now that an expression equivalent to (E.3) can be written for σ^{f*}:

$$\begin{aligned} EV_t^f(N_t;\sigma^{f*}) &= \sum_{i=t}^{\infty} \beta^{i-t} \sum_{h\in\mathcal{H}_i(N_t)} \sum_{n_i^{-f}\in C(h)} \prod_{g\neq f} q^g(n_i^g|n_{i-1}) \sum_{n_i^f\in C(h)} q^f(n_i^f|n_{i-1}^f, n_{i-1}^{-f}, \sigma^{f*}) \\ &\times \; E_{\sigma^{f*}}\left[\pi(n_i^f, n_i^{-f}) - FC^f(n_i^f) - AC^f(n_{i+1}^f, n_i^f) + \xi_t^f(n_{i+1}^f)\middle| n_i\right] \end{aligned} \tag{E.5}$$

And, by the definition of σ^{f*} and σ^f we can write:

$$\sum_{N_i^f:|N_i^f|=|\tilde{N}_i^f|} q^f(N_i^f|N_{i-1}^f, n_{i-1}^f, \sigma^f) = q^f(\tilde{N}_i^f|N_{i-1}^f, n_{i-1}^f, \sigma^{f*}) \tag{E.6}$$

Using this, we can subtract the two present value expressions to get:

$$
\begin{aligned}
&EV_t^f(N_t;\sigma^{f*}) - EV_t^f(N_t;\sigma^f) = \sum_{i=t}^{\infty}\beta^{i-t}\sum_{h\in\mathcal{H}_i(N_t)} \\
&\sum_{n_i^{-f}\in C(h)}\prod_{g\neq f} q^g(n_i^g|n_{i-1})\sum_{n^f\in C(h)} q^f(n_i^f|n_{i-1}^f, n_{i-1}^{-f}, \sigma^{f*})\times\Bigg(\\
&E_{\sigma^{f*}}\left[\pi(n_i^f, n_i^{-f}) - FC^f(n_i^f) - AC^f(n_{i+1}^f, n_i^f) + \xi_t^f(n_{i+1}^f)\middle| n_i\right] \\
&- \sum_{N_i^f:|N_i^f|=n^f}\Big\{ q^f(N_i^f|N_{i-1}^f, n_{i-1}^{-f}, |N_i^f| = n_i^f\sigma^f) \\
&E_{\sigma^f}\left[\pi(N_i^f, n_i^{-f}) - FC^f(N_i^f) - AC^f(N_{i+1}^f, N_i^f) + \xi_t^f(N_{i+1}^f)\middle| N_i\right]\Big\}\Bigg)
\end{aligned}
\tag{E.7}
$$

Within the large parenthesis the size of N^f is constant so FC, AC and ξ^f cancels out:

$$
\begin{aligned}
&EV_t^f(N_t;\sigma^{f*}) - EV_t^f(N_t;\sigma^f) = \sum_{i=t}^{\infty}\beta^{i-t}\sum_{h\in\mathcal{H}_i(N_t)} \\
&\sum_{n_i^{-f}\in C(h)}\prod_{g\neq f} q^g(n_i^g|n_{i-1})\sum_{n^f\in C(h)} q^f(n_i^f|n_{i-1}^f, n_{i-1}^{-f}, \sigma^{f*})\times\Bigg(\pi(n_i^f, n_i^{-f}) \\
&- \sum_{N_i^f:|N_i^f|=n^f} q^f(N_i^f|N_{i-1}^f, n_{i-1}^{-f}, |N_i^f| = n_i^f\sigma^f)\pi(N_i^f, n_i^{-f})\Bigg)
\end{aligned}
\tag{E.8}
$$

By the definition of σ^{f*}, and σ^f the term in parenthesis is greater than 0, because $\pi(n_i^f, n_i^{-f}) \geq \pi(N_i^f, n_i^{-f})$. Hence

$$EV_t^f(N_t;\sigma^{f*}) - EV_t^f(N_t;\sigma^f) \geq 0$$

and the AMS σ^{f*} is better than σ, as desired. □

F Derivation of Estimable Demand Equation

We have:

$$s_j = \frac{e^{(\alpha_j - \lambda_j p_j)/\gamma^f}}{e^{h^f}} \frac{e^{\frac{\gamma^f}{\rho} h^f}}{e^H} \frac{e^{\rho H}}{1 + e^{\rho H}} \tag{F.1}$$

Let $s^0 = \frac{1}{1+e^{\rho H}}$, so:

$$\frac{s_j}{s^0} = e^{(\alpha_j - \lambda_j p_j)/\gamma^f} e^{\left(\frac{\gamma^f}{\rho} - 1\right) h^f} e^{(\rho - 1)H}$$

Note that:

$$\begin{aligned} s_{j|f} &= \frac{e^{(\alpha_j - \lambda_j p_j)/\gamma^f}}{e^{h^f}} \\ \Rightarrow e^{h^f} &= \frac{e^{(\alpha_j - \lambda_j p_j)/\gamma^f}}{s_{j|f}} \end{aligned}$$

and

$$\begin{aligned} s_{j|F} &= e^{(\alpha_j - \lambda_j p_j)/\gamma^f} e^{\left(\frac{\gamma^f}{\rho} - 1\right) h^f} e^{-H} \\ \Rightarrow e^H &= e^{(\alpha_j - \lambda_j p_j)/\gamma^f} \left(\frac{e^{(\alpha_j - \lambda_j p_j)/\gamma^f}}{s_{j|f}} \right)^{\frac{\gamma^f}{\rho} - 1} (s_{j|F})^{-1} \\ \Rightarrow e^H &= e^{(\alpha_j - \lambda_j p_j)/\rho} (s_{j|f})^{1 - \frac{\gamma^f}{\rho}} (s_{j|F})^{-1} \end{aligned}$$

Replacing:

$$\begin{aligned} \frac{s_j}{s^0} &= e^{(\alpha_j - \lambda_j p_j)/\gamma^f} \times \left(\frac{e^{(\alpha_j - \lambda_j p_j)/\gamma^f}}{s_{j|f}} \right)^{\frac{\gamma^f}{\rho} - 1} \\ &\times \left(e^{(\alpha_j - \lambda_j p_j)/\rho} (s_{j|f})^{1 - \frac{\gamma^f}{\rho}} (s_{j|F})^{-1} \right)^{\rho - 1} \end{aligned}$$

Rearranging:

$$\frac{s_j}{s^0} = e^{\alpha_j - \lambda_j p_j} (s_{j|f})^{\left(1 - \frac{\gamma^f}{\rho}\right)\rho} (s_{j|F})^{1-\rho} \tag{F.2}$$

Finally, taking logs:

$$\log\left(\frac{s_j}{s^0}\right) = \alpha_j - \lambda_j p_j + (\rho - \gamma^f) \log(s_{j|f}) + (1 - \rho) \log(s_{j|F})$$

G Market Definition

I depart from previous studies and take each retailer to be an individual market. I base this decision on recent findings in the retail pricing literature (DellaVigna and Gentzkow (2017), and Hitsch et al. (2017) showing that there is little price variation for a given product across locations, within retailers, and that most of the price variation of a product is explained by between retailer variation. This findings suggest that, at least pricing wise, the retailer is a more important strategic unit than geographic markets.

To further support this decision I perform the following exercise. Let $\tilde{p}_{jrdt}$ be the log-price of good j sold by retailer r, at geographic area d,[19] during period t. For each product I calculate the within retailer mean absolute price deviation:

$$\Delta_j^W = (RDT)^{-1} \sum_{r,d,t} \left| \tilde{p}_{jrdt} - R^{-1} \sum_r \tilde{p}_{jrdt} \right|$$

and the between retailer mean absolute deviation:

$$\Delta_j^B = (DT)^{-1} \sum_{d,t} \left| R^{-1} \sum_r \tilde{p}_{jrdt} - R^{-1} D^{-1} \sum_r \sum_d \tilde{p}_{jrdt} \right|$$

The histogram in figure G.1 show the result of this exercise. It shows that within retailer price variation consistently lower than 1%, whereas the between retailer variation. is considerably larger, with an average between 5 and 6 percent.

Figure G.1: Price Variation Histogram

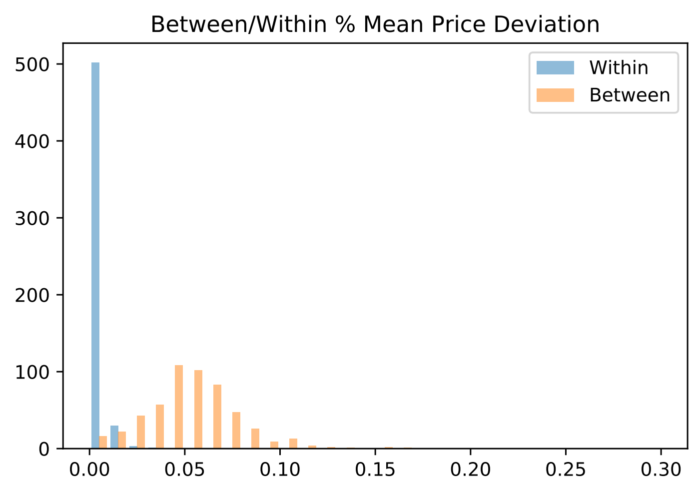


Figure G.2 shows a similar measure only aggregated over product rather than over time. This

[19] For geographic area I use the Designated Market Area (DMA) from the Nielsen dataset. A DMA is roughly defined as the contiguous geographic area that receives the same television programming.

is:

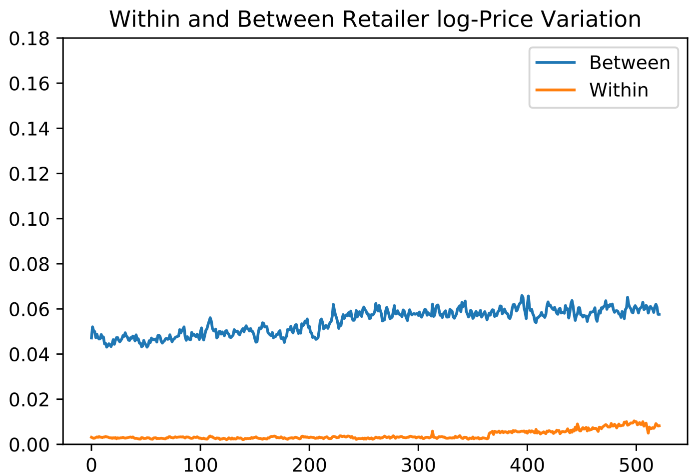
$$\Delta_t^{W'} = (N_t RD)^{-1} \sum_{j,r,d} \left| \tilde{p}_{jrdt} - R^{-1} \sum_r \tilde{p}_{jrdt} \right|$$

and a measure of average deviation of the mean price within the retailer to the overall mean.

$$\Delta_t^{B'} = (N_t D)^{-1} \sum_{j,d} \left| R^{-1} \sum_r \tilde{p}_{jrdt} - R^{-1} D^{-1} \sum_r \sum_d \tilde{p}_{jrdt} \right|$$

It shows that the average between and within retailer variations are reasonably stable over time, and that the between retailer variation between two and three times the within retailer variation.

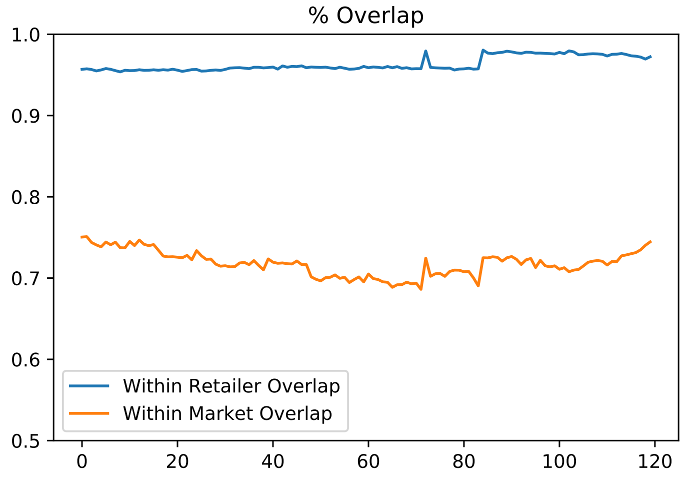
Figure G.2: Price Variation Series



Another cause of concern when defining the relevant markets is the overlap in product variety between the different aggregating units. To explore the overlap overlap in product availability within and between retailers I conducted the following calculation. First, I asked of every product j, if j is sold by retailer r in any DMA d, on what proportion of DMAs in which r is present j be found. Second, I asked about every product j, if j is sold in DMA d, what is the proportion of DMAs on which it is sold. Figure G.3 presents the average over products-retailer and product-dma for each of these exercises.

From the demand perspective, the evidence in figures G.1, G.2 and G.3 suggests that consumers buying from the same retailer in different geographic areas face a more similar choice set than those buying in the same geographic area from different retailers. Similarly, it suggests that pricing and product assortment strategies take place at the retailer level more than at the geographic level.

Figure G.3: Price Variation Series



H Summary Statistics

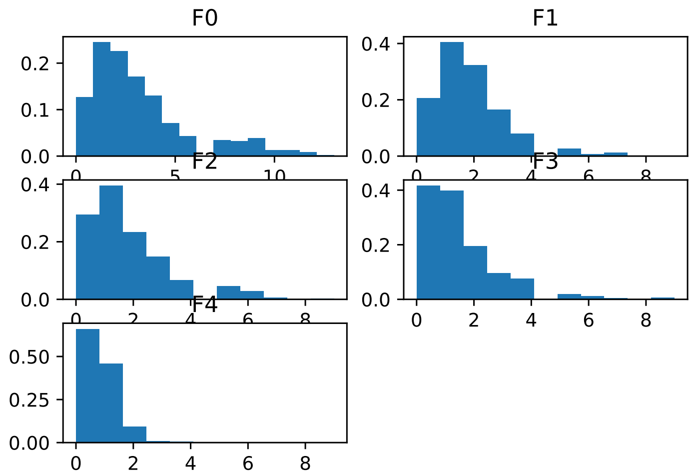


Table H.1: Aggregate Servings (millions) by Firm

year	F0	F1	F2	F3	F4	Total
2007	207.52	184.36	236.50	157.19	64.52	850.08
2008	295.68	267.61	308.68	187.69	79.38	1139.05
2009	304.54	268.69	313.67	169.05	93.72	1149.67
2010	301.90	271.32	291.64	162.51	91.33	1118.70
2011	319.94	269.38	294.63	128.27	77.80	1090.02
2012	312.91	254.73	278.57	115.62	64.29	1026.13
2013	294.97	257.69	270.47	117.92	62.84	1003.89
2014	293.70	251.93	261.24	118.87	69.71	995.45
2015	287.04	253.09	263.27	119.48	73.06	995.94
2016	276.37	246.06	261.00	123.44	69.09	975.97

Note: The large increase between 2007 and 2008 is an artifact of the fact that the quarter of 2007 is dropped when taking the quarterly lagged number of products.

Table H.2: Market Shares Over Time

year	F0	F1	F2	F3	F4
2007	0.24	0.22	0.28	0.18	0.08
2008	0.26	0.23	0.27	0.16	0.07
2009	0.26	0.23	0.27	0.15	0.08
2010	0.27	0.24	0.26	0.15	0.08
2011	0.29	0.25	0.27	0.12	0.07
2012	0.30	0.25	0.27	0.11	0.06
2013	0.29	0.26	0.27	0.12	0.06
2014	0.30	0.25	0.26	0.12	0.07
2015	0.29	0.25	0.26	0.12	0.07
2016	0.28	0.25	0.27	0.13	0.07

Table H.3: Yearly Averages and Standard Deviation

year	Obs.	N	price_per_serving	sF	serv.	market size
2007	104,625	190.9	0.389	0.284	1.45	5.13
		(23.9)	(0.024)	(0.052)	(1.13)	(4.14)
2008	143,743	198.8	0.411	0.291	1.46	5.14
		(27.2)	(0.025)	(0.052)	(1.16)	(4.27)
2009	140,766	195.0	0.421	0.292	1.47	5.16
		(28.6)	(0.024)	(0.046)	(1.20)	(4.43)
2010	134,869	188.8	0.415	0.293	1.43	4.95
		(25.0)	(0.030)	(0.047)	(1.21)	(4.31)
2011	135,405	185.7	0.428	0.289	1.37	4.75
		(24.5)	(0.031)	(0.044)	(1.18)	(4.26)
2012	134,316	187.4	0.435	0.275	1.32	4.72
		(23.8)	(0.036)	(0.042)	(1.18)	(4.31)
2013	140,259	194.3	0.434	0.268	1.29	4.75
		(22.8)	(0.039)	(0.036)	(1.17)	(4.37)
2014	141,698	196.1	0.433	0.280	1.28	4.61
		(22.0)	(0.038)	(0.045)	(1.13)	(4.01)
2015	143,118	195.7	0.435	0.274	1.28	4.76
		(23.0)	(0.039)	(0.051)	(1.16)	(4.28)
2016	139,955	189.2	0.437	0.217	1.23	5.60
		(26.2)	(0.041)	(0.043)	(1.14)	(4.96)

Note: Averages on top, standard deviation in parenthesis at the bottom.

Table H.4: Mean Number of Products

year	F0	GML	KEL	PST	QKR
2007	75.36	37.07	35.80	30.82	11.84
	(20.60)	(2.90)	(4.00)	(3.48)	(2.08)
2008	82.45	39.13	37.97	28.92	10.30
	(21.91)	(3.38)	(4.35)	(3.48)	(1.99)
2009	83.55	39.85	35.78	26.78	9.00
	(21.81)	(3.86)	(5.40)	(3.43)	(1.39)
2010	82.20	37.08	34.22	26.30	9.03
	(18.08)	(4.55)	(3.67)	(4.73)	(1.34)
2011	81.95	37.75	35.86	20.05	10.13
	(22.09)	(3.36)	(2.96)	(4.77)	(1.75)
2012	83.95	39.70	35.45	17.68	10.67
	(21.05)	(3.86)	(3.25)	(3.32)	(1.86)
2013	84.25	41.07	39.43	19.47	10.13
	(19.79)	(3.67)	(4.14)	(3.13)	(1.59)
2014	83.17	42.00	39.98	21.12	9.83
	(19.54)	(4.31)	(4.24)	(2.51)	(1.20)
2015	84.75	42.37	39.23	19.88	9.42
	(19.51)	(4.67)	(5.33)	(2.53)	(1.05)
2016	84.80	39.25	37.46	18.30	9.37
	(21.05)	(4.42)	(5.17)	(2.34)	(1.20)

Note: Averages on top, standard deviation in parenthesis at the bottom.

I Market Size

I follow Backus et al. (2018) in defining market size as the number of consumers that enter a retailer in a give week, which I approximate using the two most commonly bought items, milk and eggs. To do this, I ran the following OLS regression on weekly data, at the retailer level:

$$\log(\text{cereal servings}_{kw}) = \kappa_{\text{month}} + \kappa_{\text{market}} + \kappa_1 \log(\text{milk}_{kw}) + \kappa_2 \log(\text{eggs}_{kw}) \tag{I.1}$$

The market size, is proportional to the predicted sales of cereals obtained from this OLS estimation. This is:

$$\mathcal{D}_{kw} = \kappa_{\mathcal{D}} \overbrace{e^{\kappa_{\text{month}}+\kappa_k+\kappa_1 \log(\text{milk}_{kw})+\kappa_2 \log(\text{eggs}_{kw})}}^{\text{predicted servings}} \tag{I.2}$$

Finally, $\kappa_{\mathcal{D}}$ is chosen so that the average market share of the value of the outside option is 27.6%, this is:

$$\begin{aligned} \frac{\sum_{kw} \text{servings}_{kw}}{\sum_{kw} \mathcal{D}_{kw}} &= 0.276 \\ \Rightarrow \frac{\sum_{kw} \text{servings}_{kw}}{\kappa_{\mathcal{D}} \sum_{kw} \text{predicted servings}_{kw}} &= 0.276 \\ \Rightarrow \kappa_{\mathcal{D}_{kw}} &= \frac{\sum_{kw} \mathcal{D}_{kw}}{0.276 \sum_{kw} \text{servings}_{kw}} \end{aligned} \tag{I.3}$$

Table I presents the resuls of the OLS regression. Columns differ in the fixed effects that are included in the estimation.

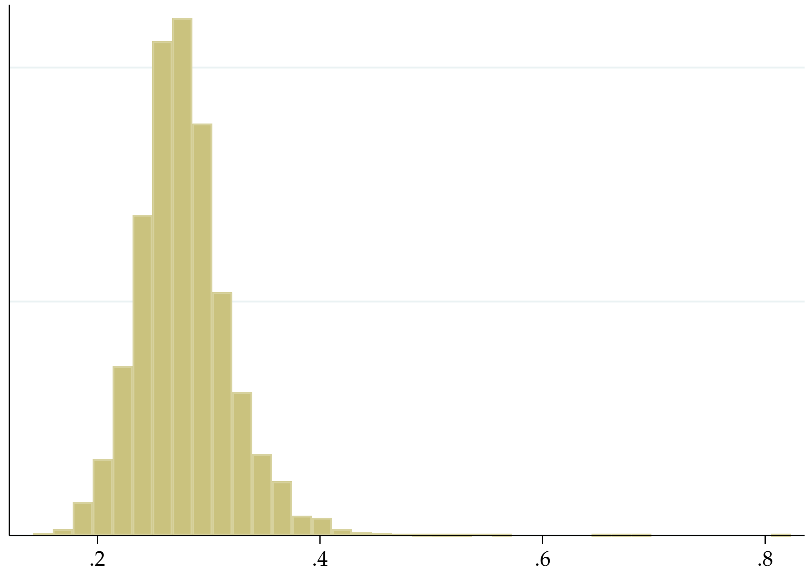
	(1) log(cereal)	(2) log(cereal)	(3) log(cereal)	(4) log(cereal)
log(eggs)	0.352***	0.186***	0.431***	0.278***
	(0.011)	(0.009)	(0.011)	(0.009)
log(milk)	0.574***	0.606***	0.507***	0.562***
	(0.010)	(0.009)	(0.010)	(0.009)
Observations	7633	7633	7633	7633
r2	0.946	0.977	0.957	0.985
MarketFE	No	Yes	No	Yes
MonthFE	No	No	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Table I.1: Estimation of Market Size

Figure I.1: Market Share Outside Option



J First Step Demand Estimation and Marginal Cost

Figure J.1: Monthly Commodity Price Index. 2007/01 - 2016/12, Log-Scale

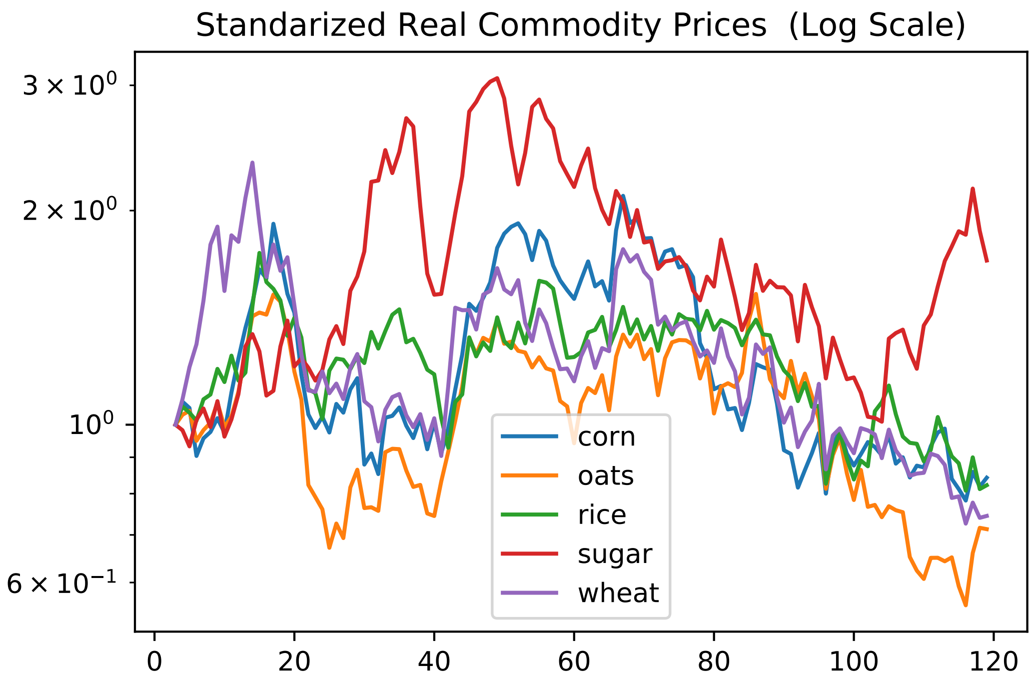


Table J.1: First Stage: log_sjF

	(1) log_sjF	(2) log_sjF	(3) log_sjF	(4) log_sjF
Corn	-0.003	-0.009	-0.011	-0.008
	(0.003)	(0.010)	(0.010)	(0.010)
Oats	-0.017***	-0.001	-0.000	-0.001
	(0.004)	(0.009)	(0.009)	(0.009)
Rice	0.037***	0.010**	0.011**	0.010**
	(0.003)	(0.005)	(0.005)	(0.005)
Sugar	-0.068***	-0.004	-0.003	-0.003
	(0.001)	(0.005)	(0.005)	(0.005)
Wheat	0.035***	-0.007	-0.007	-0.008
	(0.003)	(0.007)	(0.007)	(0.007)
Hausman	-2.396***	-2.355***	-2.139***	-2.393***
	(0.019)	(0.019)	(0.020)	(0.019)
Lag # products	-0.007***	-0.007***	-0.007***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1358754	1358754	1358754	1358754
SW	0.00	0.00	0.00	0.00
ProductFE	Yes	Yes	Yes	Yes
MarketFE	Yes	Yes	Yes	No
QuarterFE	No	Yes	No	Yes
FirmQuarterFE	No	No	Yes	No
FirmMarketFE	No	No	No	Yes

Standard errors in parentheses

SW: p-value of Sanderson-Windmeijer test of weak instruments

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Table J.2: First Stage: Price

	(1) price	(2) price	(3) price	(4) price
Corn	0.008***	0.000	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.001)
Oats	0.004***	-0.000	-0.000	-0.000
	(0.000)	(0.001)	(0.001)	(0.001)
Rice	-0.000**	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Sugar	0.000***	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Wheat	-0.014***	0.000	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.001)
Hausman	0.406***	0.388***	0.366***	0.391***
	(0.001)	(0.001)	(0.001)	(0.001)
Lag # products	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1358754	1358754	1358754	1358754
SW	0.00	0.00	0.00	0.00
ProductFE	Yes	Yes	Yes	Yes
MarketFE	Yes	Yes	Yes	No
QuarterFE	No	Yes	No	Yes
FirmQuarterFE	No	No	Yes	No
FirmMarketFE	No	No	No	Yes

Standard errors in parentheses

SW: p-value of Sanderson-Windmeijer test for weak instruments

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Table J.3: First Stage: log_sjF - 4 Year Sample

	(1) log_sjF	(2) log_sjF	(3) log_sjF	(4) log_sjF
Corn	-0.040***	-0.005	-0.009	-0.004
	(0.006)	(0.014)	(0.014)	(0.013)
Oats	0.068***	-0.033**	-0.029*	-0.035**
	(0.008)	(0.016)	(0.016)	(0.015)
Rice	0.002	0.007	0.008	0.006
	(0.004)	(0.006)	(0.006)	(0.006)
Sugar	-0.064***	-0.003	-0.003	-0.003
	(0.002)	(0.006)	(0.006)	(0.006)
Wheat	0.038***	0.006	0.006	0.007
	(0.004)	(0.009)	(0.009)	(0.006)
Hausman	-2.352***	-2.207***	-2.017***	-2.269***
	(0.033)	(0.033)	(0.033)	(0.032)
Lagg # products	-0.002***	-0.001***	-0.001***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	590816	590816	590816	590816
SW	0.00	0.00	0.00	0.00
ProductFE	Yes	Yes	Yes	Yes
MarketFE	Yes	Yes	Yes	No
QuarterFE	No	Yes	No	Yes
FirmQuarterFE	No	No	Yes	No
FirmMarketFE	No	No	No	Yes

Standard errors in parentheses

SW: p-value of Sanderson-Windmeijer test of weak instruments

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Table J.4: First Stage: Price - 4 Year Sample

	(1) price	(2) price	(3) price	(4) price
Corn	0.011***	0.001	0.001	0.001
	(0.000)	(0.001)	(0.001)	(0.001)
Oats	-0.016***	-0.003**	-0.004***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Rice	0.003***	0.001	0.001	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
Sugar	0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Wheat	-0.002***	0.002**	0.002**	0.002***
	(0.000)	(0.001)	(0.001)	(0.000)
Hausman	0.321***	0.306***	0.292***	0.311***
	(0.002)	(0.002)	(0.002)	(0.002)
Lagg # products	-0.000***	-0.000***	-0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	590816	590816	590816	590816
SW	0.00	0.00	0.00	0.00
ProductFE	Yes	Yes	Yes	Yes
MarketFE	Yes	Yes	Yes	No
QuarterFE	No	Yes	No	Yes
FirmQuarterFE	No	No	Yes	No
FirmMarketFE	No	No	No	Yes

Standard errors in parentheses

SW: p-value of Sanderson-Windmeijer test for weak instruments

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

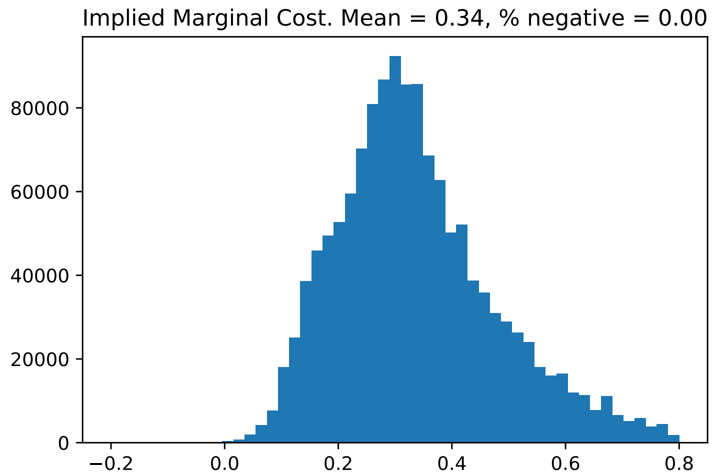
Table J.5: Second Stage Nested Logit - 4 Year Sample

	(1) log_sjs0	(2) log_sjs0	(3) log_sjs0	(4) log_sjs0
price	0.341***	-2.561***	-2.579***	-2.955***
	(0.047)	(0.084)	(0.085)	(0.102)
log_sjF	1.048***	0.697***	0.687***	0.645***
	(0.005)	(0.010)	(0.011)	(0.013)
Observations	590816	590816	590816	590816
rho	-0.05	0.30	0.31	0.35
lambda_rho	-7.17	-8.45	-8.25	-8.33
ProductFE	Yes	Yes	Yes	Yes
MarketFE	Yes	Yes	Yes	No
QuarterFE	No	Yes	No	Yes
FirmQuarterFE	No	No	Yes	No
FirmMarketFE	No	No	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Figure J.2: Histogram of Implied Marginal Costs



K Elasticities

Here I derive the formulas used to compute elasticities and show the elasticity histograms.

K.1 Own Price Elasticity

This corresponds to the standard demand elasticity:

$$\eta_{jj} = \frac{ds_j}{dp_j}\frac{p_j}{s_j}$$

Recall that:

$$s_j = \frac{e^{(\alpha_j - \lambda p_j)/\gamma^f}}{e^{h^f}} \frac{e^{\frac{\gamma^f}{\rho}h^f}}{e^H} \frac{e^{\rho H}}{1+e^{\rho H}}$$

So:

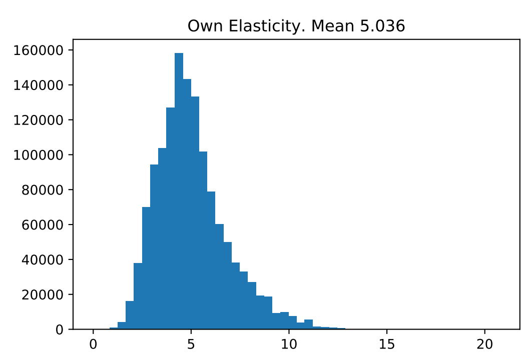
$$\frac{ds_j}{dp_j} = \frac{\partial s_j}{\partial p_j} + \frac{\partial s_j}{\partial h^f}\frac{\partial h^f}{\partial p_j} + \frac{\partial s_j}{\partial H}\frac{\partial H}{\partial h^f}\frac{\partial H}{\partial p_j}$$

I compute it as follows:

$$\begin{aligned} \frac{ds_j}{dp_j} &= -\frac{\lambda}{\gamma^f}s_j + s_j s_j \frac{\lambda}{\rho}\left(\left(1-\frac{\gamma^f}{\rho}\right)\frac{1}{\gamma^f}\frac{\rho}{s^f} + \frac{1-\rho s^0}{s^F}\right) \\ &= -\frac{\lambda}{\gamma^f}s_j\left[1-\left(1-\frac{\gamma^f}{\rho}\right)s_{j|f} - \frac{\gamma^f}{\rho}s_{j|F}(1-\rho s^0)\right] \end{aligned} \tag{K.1}$$

So, the elasticity is:

$$\eta_{jj} = -\frac{\lambda}{\gamma^f}p_j\left[1-\left(1-\frac{\gamma^f}{\rho}\right)s_{j|f} - \frac{\gamma^f}{\rho}s_{j|F}(1-\rho s^0)\right] \tag{K.2}$$



K.2 Market Elasticity

As before, let $\bar{p}$ be some reference price level, and define w_j such that:

$$p_j = w_j \bar{p}$$

The market elasticity measures the percentage change in aggregate demand as a result of a percentage change in the price level.

$$\eta_{FF} = \frac{ds^F}{d\bar{p}} \frac{\bar{p}}{s^F}$$

Re-writing $s^F = e^{\rho H}/(1 + e^{\rho H})$, and differentiating with respect to $\bar{p}$:

$$\begin{aligned} \frac{ds_F}{d\bar{p}} &= \frac{\partial s_F}{\partial H} \sum_{f \in F} \frac{\partial H}{\partial h^f} \sum_{j \in N^f} \frac{\partial h^f}{\partial p_j} \frac{\partial p_j}{\partial \bar{p}} \\ &= \rho s^F (1 - s^F) \left(\sum_{f \in F} \frac{\gamma^f}{\rho} s^{f|F} \sum_{j \in N^f} -\frac{\lambda}{\gamma^f} s_{l|f} w_j \right) \\ &= -\lambda s^F (1 - s^F) \left(\sum_{f \in F} \sum_{j \in N^f} s_{l|f} s^{f|F} w_j \right) \end{aligned}$$

Hence:

$$\begin{aligned} \eta_{FF} &= -\lambda \bar{p} (1 - s^F) \left(\sum_{f \in F} \sum_{j \in N^f} s_{l|F} w_j \right) \\ &= -\lambda (1 - s^F) \left(\sum_{f \in F} \sum_{j \in N^f} s_{l|F} p_j \right) \end{aligned} \tag{K.3}$$