Identification and Estimation of Intra-Firm and Industry Competition via Ownership Change*

Christian Michel
Universitat Pompeu Fabra, Barcelona GSE

This version: May 09, 2016

Abstract

In this paper, I estimate the degree of organizational integration of merging firms and the intensity of industry competition in the ready-to-eat cereal industry. Using pre- and post-merger industry data surrounding the 1993 Post-Nabisco merger, I separate merging firms' intra-organizational pricing considerations from industry pricing considerations. I find an increasing degree of joint-profit maximization of the merged entities over the first two years post-merger, eventually leading to close to full joint-profit maximization. Moreover, between 18.3 and 21.1 percent of the median markups in the industry can be attributed to cooperative industry behavior.

JEL Classification: L11, L22, C52

*Email: christian.michel@upf.edu. The paper is based on chapter 2 of my thesis. I am indebted to Volker Nocke, Philipp Schmidt-Dengler, and Yuya Takahashi for their guidance and advice. I benefited from conversations with Steven Berry, Pierre Dubois, Georg Duernecker, Rosa Ferrer, Amit Gandhi, David Genesove, Andre Romahn, Alex Shcherbakov, André Stenzel, and Stefan Weiergraeber, and received helpful comments from seminar participants at Bocconi, Toulouse, Pompeu Fabra, Mannheim, Tilburg, Compass Lexecon Europe, Copenhagen Business School, ESSEC Business School, Vienna, VfS 2013 in Duesseldorf, EEA 2015 in Mannheim, the ESWC 2015 in Montreal, and the IIOC 2016 in Philadelphia. I gratefully acknowledge financial support from the Marie Curie Intra-European Fellowship FP7-PEOPLE-2013-IEF-623912, and from the Spanish Ministry of Economy and Competitiveness Grant ECO2013-43011.

1 Introduction

There is a long-standing debate about the extent to which behavior within firms deviates from profit maximization (e.g., Machlup, 1967). Several theories predict less than full maximization of joint profits between different divisions of a firm (see e.g., Schwartz and Thompson, 1986; Fauli-Oller and Giralt, 1995; and Wickelgren, 2005). Chandler (1962) already documents competition between divisions within General Motors. However, potentially competitive within-firm interactions have so far not received much attention from the empirical literature. A key difficulty of any empirical approach is that the optimal within-firm behavior also depends on how intensely the firm as a whole competes with its industry rivals.

Within-firm behavior is of particular interest following a merger. The organizational integration of merging divisions can be complicated by different corporate cultures and procedures, and by division managers that compete for the same position in the firm (e.g., Shrivastava, 1986). This can lead to less than full joint-profit maximization between different divisions, and can in turn influence the behavior of other firms in an affected industry.

In this paper, I estimate structures of both within-firm and industry behavior using data from the ready-to-eat (RTE) cereal industry. First, I examine to what extent merging firms jointly maximize their profits after a horizontal merger. This sheds light on the speed and intensity of organizational integration. Second, I directly estimate industry conduct in differentiated products industries. I show that when pre- and post-merger industry data is available, one can make use of the ownership change to estimate rich organizational and industry structures instead of assuming them. This allows for estimating how closely merging divisions cooperate and how intensely firms in an industry compete.

The paper particularly focuses on estimating the competitive intensity both within and between firms while relying on existing approaches to recover industry demand and marginal costs, see for example Berry *et al.* (1995) and Nevo (2001). I recover marginal costs using firms' first-order conditions. Different forms of industry conduct or degrees of profit internalization of merging firms, respectively, lead to different markups charged by firms. The difference between predicted and observed prices amounts to an unobservable structural cost term that is thus contingent on the underlying supply-side specification. I generate moments based on this term to identify the supply-side parameters of interest. This allows me to estimate a time-variant degree of merging firms' joint-profit maximization or to estimate a rich structure of how intensely different firms compete with each other in an industry.

I use *differentiation instruments* as recently introduced by Gandhi and Houde (2015), who build on the work of Berry and Haile (2014), to identify both consumer preferences and supply-side behavior. In particular, I combine measures of a product's relative proximity in the characteristics space with store-specific data on other products' promotion periods to form demand and supply-side instruments. The differentiation instruments should be correlated to a product's demand, and therefore also to its price, but not to product-specific transitory demand and cost shocks. I find that these instruments are particularly useful for identifying supply-side parameters.[1] On the demand side, I exploit variation in input and wholesale prices, and use the ownership change itself to form additional instruments.

For my estimations I use data from the RTE cereal industry covering the 1993 Post-Nabisco merger. In January 1993, Kraft foods with its Post cereal line purchased the Nabisco RTE cereal branch from RJR Nabisco. I use data from the Dominick's Finer Foods (henceforth DFF) database. One very attractive property of the data is that I observe wholesale prices in addition to retail prices. This allows me to focus in detail on both competition and merger integration of different manufacturing firms while exploiting the retail data to accurately estimate consumer demand.

With respect to the post-merger integration of the merging firms, my results indicate an increase in the degree of joint-profit maximization over time, leading to close to full joint maximization two years after the merger. This finding is robust both to different market structures, and to accounting for potential cost synergies resulting from the merger. When focusing on industry competitiveness, I find that between 18.3 and 21.1 percent of median manufacturer markups in the industry can be attributed to cooperative industry behavior, while the remaining markups are due to product differentiation of multi-product firms. Furthermore, the data indicates that DFF as the retailer also benefited from the merger via a retail price that increased more steeply than the wholesale price.

To my knowledge, this is the first paper to focus on estimating the degree of joint-profit maximization of horizontally merging firms. This links the empirical industrial organization literature to the theoretical organizational economics literature on intra-firm coordination and horizontal integration by allowing for frictions between different divisions of a firm.[2] Conceptually, this approach differs from most existing empirical organizational economics models in that its focus is on behavior within a single organization rather than on correlations across firms and industries.[3] This allows for estimating how merging firms cooperate rather

[1]In ongoing work, Gandhi and Houde (2016) focus on the creation of relevant differentiation instruments to identify industry conduct, and propose a direct approximation of firms' supply relations.

[2]Crawford *et al.* (2016) analyze the welfare effects of vertical integration in the U.S. cable and satellite industry. They account for internalization effects using a structural bargaining model and find a less than optimal increase in internalization after a merger.

[3]There is a large empirical literature in organizational economics focusing on the determinants for specific organizational structures across firms and industries, see for example Lafontaine and Slade (2012), and Bloom *et al.* (2012) for an overview over the literature. Thomas (2011) argues that a reduction in the brand portfolios of firms in the laundry detergent industry across different countries would lead to an increase in their profits.

than having to assume it. I further show that differentiation instruments can identify rich supply-side structures. Even if there is no variation in the product space, a convenient way to compute such instruments is by using data on other products' promotional activity at the market level. Moreover, using supply-side variation, I estimate industry conduct directly in a differentiated products industry, which can typically not be achieved using demand-side variation due to a lack of sufficiently many demand rotators; see Nevo (1998) for a discussion.

On an organizational level, existing empirical merger models assume that merging firms fully internalize their pricing externalities immediately after a merger. This follows the view of a merged firm as a single agent that can jointly maximize the profits of all of its divisions. Several theories predict that full internalization of joint profits cannot be achieved after a merger. Incentive structures that give managers bonuses based on the performance of their own division rather than the performance of the firm as a whole can cause different horizontal divisions to compete with each other. Fauli-Oller and Giralt (1995), for example, analyze a headquarter's choice of the optimal incentive scheme for division managers. Whenever products of different divisions are substitutes to each other, managers' bonuses will be partly based on their own division's performance. There is further a growing literature in organizational economics that focuses on the trade-offs between coordination of decision-making through a headquarter and strategic communication of division managers.[4]

When estimating industry conduct, I maintain the assumption as in the literature that merging firms fully internalize the profits after the merger. Given marginal cost estimates and price-elasticities obtained from the demand side-estimation, I can predict the effects of an ownership change on prices ex-post. By varying the form of supply-side competition (i.e. industry conduct), and accounting for input price changes on the cost side, I look for the form of competition that most accurately predicts the "unilateral" effects of the merger-induced ownership change on prices. I use the differences of observed prices and predicted model prices to form moments in order to obtain the model's underlying conduct parameters using a Generalized Methods of Moments estimator.[5]

Previous attempts to estimate both marginal costs and industry conduct rely mostly on demand side variation. Bresnahan (1982) and Lau (1982) provide identification results for estimating conduct in the homogeneous good case. In differentiated product industries,

[4]Alonso *et al.* (2008) study the optimal degree of centralization when managers communicate strategically. They show that while centralization can improve horizontal outcomes, it will lead to adverse vertical effects. Dessein *et al.* (2011) show the existence of endogenous incentive conflicts between headquarter managers and division managers within multi-divisional firms. Fershtman and Judd (1987) show that in a Cournot oligopoly, firm owners in most cases have no incentive to set incentive structures to their managers that maximize within-firm profits as to favorably influence rival firms' actions.

[5]This is a refined version of my previous work in Michel (2013), which to my knowledge is the first to estimate flexible parameters inside a supply-side ownership matrix to make inference on the underlying form of competition.

these approaches usually face two kinds of problems. The first problem is the difficulty to find a sufficient number of demand rotators to identify industry conduct. The second problem relates to the estimation techniques, which only estimate the economic parameters of interest accurately in special cases. Nevo (1998) discusses advantages and disadvantages of a direct conduct estimation compared to a non-nested menu approach. He argues that in practice estimating conduct directly will be impossible due to a lack of sufficiently many distinct demand shifters. As an alternative, he proposes the use of selection tests for a "menu" of pre-specified models, see for example Gasmi *et al.* (1992) and Rivers and Vuong (1988). Berry and Haile (2014) show that to nonparametrically identify consumer preferences in discrete choice models when observing only market level data, it is necessary but not sufficient to exploit information on the characteristics space. Using additional instruments such as cost shifters leads to a sufficient condition for identification. On the supply side, they show the potential to identify underlying industry conduct in differentiated product industries for two different identifying assumptions. In the first case, they assume identification of cost parameters independent of the oligopoly model. In the second case, they use exclusion restrictions on the unobservable cost term similar to my model, and show that identification of the supply side can be theoretically achieved. In my model, I make further use of variation in the ownership of merging firms. The observable change in market structure helps identifying the underlying supply side by accounting for the change of the firm's pricing decisions.

There is a small literature related to the estimation of industry conduct using supply-side variation. Bresnahan (1987) estimates a structural model for both demand and supply to test whether multi-product Bertrand Nash pricing or full collusion can better explain the US car industry around a price war in 1955. For 1954 and 1956 his results indicate a collusive industry outcome, and indicate multi-product Nash pricing for 1955. Miller and Weinberg (2015) account for a single supply-side parameter to assess the effects of a joint-venture on industry behavior in the beer industry. This can be seen as complementary to my approach: While I either estimate the degree of profit internalization of merging firms given a specific form of pre-merger conduct, or the underlying form of industry conduct also pre-merger, they estimate the change in conduct from a known form of pre-merger competition. Furthermore, their identification relies on regional differences in transportation costs and variation in regional income over time, instead of exploiting the characteristics space. Ciliberto and Williams (2014) develop an approach that relies on multi-market contact for estimating conduct in the airline industry. Their model includes conduct parameters that can have three different values, accounting for different degrees of cooperation among profit-maximizing firms.

Corts (1999) critically discusses the identification of conduct parameters. He argues that the estimated parameters usually differ from the "as-if conduct parameters" and therefore do not reflect the economic parameters of interest. The static conduct estimation models are furthermore not able to detect all dynamic forms of collusion. While I cannot account for the latter point due to the static nature of the model, my approach can overcome the former. A related caveat is that it abstracts from dynamic storage considerations. The underlying assumption with respect to consumer choice is that each consumer decides each period whether or not to purchase a specific product. If some consumers anticipate sale periods and use them to store products, this would change both the substitution behavior of consumers and optimal price setting of firms.[6]

The remainder of the paper is organized as follows. Section 2 gives an overview over the industry and the merger. Section 3 introduces the model setting. Section 4 discusses identification both for estimating the degree of joint-profit maximization of merging firms and for estimating industry conduct, respectively. Section 5 introduces the data and my estimation technique. Section 6 presents my empirical results. Section 7 concludes with a discussion of the results in a more general context. In the related online appendix I give proofs of minimum rank requirements and examples of the different supply-side specifications, and discuss alternative ways to identify industry conduct using other sources of variation. I further provide computational details on the estimation routines, and present additional figures and tables.

2 Industry Overview and the 1993 Post-Nabisco Merger

In this section, I first give a general overview of the RTE cereal industry. I then discuss the effects of potential synergies due to the merger and why a structural model is needed to answer the questions of interest.

2.1 General industry overview

There are several factors that make the RTE cereal industry an excellent starting point for a detailed oligopoly analysis. Economies of scale in packaging different cereals, as well as in the distribution of products, cause barriers to entry for new firms. This industry has already been studied extensively, see for example Schmalensee (1978), Corts (1995), and Nevo (2000). The cereals differ with respect to their product characteristics, such as sugar content or package design. In the beginning of my sample period, the industry consists of 6 main nationwide manufacturers: Kellogg's, General Mills, Post, Nabisco, Quaker Oats, and Ralston Purina. It

[6] In particular, such storage behavior could overestimate own-price elasticities and downward bias the cross-price elasticities in absolute magnitude, which would affect the estimated markups and marginal costs, see for example Hendel and Nevo (2006). To focus on more long-term price elasticities, I aggregate data on the monthly level.

is common to classify the cereals into different groups, such as adult, family, and kids cereals. Kellogg's as the firm with the biggest market share has a strong presence in all segments. General Mills is mainly present in the family and kids segments, whereas Post and Nabisco have their main strengths in the adult segment. The products furthermore differ in the type of main cereal grain, and type of processing. The main types of cereal grains are corn, wheat, rice, and oat. The main production processes are flaking, puffing, shredding, and baking. In the late 1980's, shortly before my sample period, the industry was characterized by a high rate of new product introductions, however, these introductions were only rarely successful. On a retail level, RTE cereal products are primarily distributed via supermarkets. According to Nevo (2000), while more than 200 brands were available to consumers, more than 55 percent of sales could be attributed to the 25 most popular brands.

2.2 The 1993 Post-Nabisco merger and industry price development

Throughout my analysis, I use deflated prices using a regional consumer price index. In both 1991 and 1992, i.e. in the two years before the merger, prices remained relatively stable in the industry. On November 12, 1992, Kraft Foods made an offer to purchase RJR Nabisco's RTE cereal line. The acquisition was cleared by the FTC on January 4, 1993. On February 10th, 1993, the New York State attorney sued for a divestiture of the Nabisco assets, which was finally turned down 3 weeks later. Figure 2 in the online appendix shows the average (deflated) retail price development across all stores for the merging firms' brands. Retail prices slightly increased directly after the merger, and increased more steeply roughly one year after the merger. Online appendix A presents supporting reduced-form estimation results with respect to both retail and wholesale price development. The average increase in merging firms retail prices over time is in line with a unilateral effects merger model. The retail price development of other firms is more heterogeneous. Prices for Kellogg's and Ralston on average increased. Prices for most of General Mills products slightly decreased. According to Corts (1995), in 1993 demand for General Mills' products relative to its competitor's was slumping, such that the decrease in prices can be interpreted as a response to negative demand shocks. This is supported by the product-specific average retail and wholesale price movements over time in Table 6 in the online appendix. While Kellogg's market share remained almost unchanged, General Mills lost four percent market share over time, as can be seen from Figure 3 in the online appendix.

In March 1995, 26 months after the merger, two US congressmen started a public campaign to reduce cereal prices, which received relatively high media attention. There is evidence that industry prices decreased after this campaign, see for example Cotterill and

Franklin (1999) for a detailed analysis. In April 1996, Post decreased the prices for its products nationwide by up to 20%, thereby also increasing its markets share. This was followed by significant price cuts two months later by the market leader Kellogg's.

These facts suggest a systematic change in industry pricing, potentially caused by the public campaign for lower prices, which I discuss in detail in section 6.4. Such a change in industry pricing would violate my identification assumption of a stable form of industry conduct over time apart from the merging firms internalization of pricing externalities. For this reason, I only consider the period until February 1995 for my estimations, i.e. until 25 months after the merger.

Exogeneity of merger and possible synergies From an estimation standpoint, it is important to discuss concerns and potential effects of merger endogeneity. If the merger leads to productivity gains not accounted for in my model, this could bias the results of my estimations. I do not observe any factory closures within the first five years of the merger. Nabisco's main production faciltiy in Naperville, Illinois continues to produce the same products after the merger as before. The merging firms' products also have different production technologies. Post's products primarily require flaking and baking processes, while Nabisco's products mainly rely on shredding. Therefore, I argue that marginal cost synergies due to joint factory production or transportation are unlikely to be achieved. In my baseline setting I thus do not account for such synergies.[7] As a robustness check, I also estimate the degree of profit-internalization and industry conduct for a known level of marginal cost synergies. When accounting for a 5 percent decrease in marginal costs for the merging firms after the merger, I find the results to be close to my main specification.

I explicitly rule out synergies due to increased bargaining power of the merged firm with suppliers of inputs. Because the production facilities of the different firms are geographically separated, the need of using different suppliers with respect to wheat, sugar, and energy seems reasonable. If such synergies existed, an increase in prices would overestimate marginal costs of the merging firms post-merger and underestimate the degree of profit internalization of merging firms. There are further synergies the model allows for. First, if the merger leads to savings in fixed costs, this has no effect on the pricing. This is because savings in fixed costs do not affect the first order pricing conditions of the different firms. An example for such savings are savings in administrative staff or in renting office space. Similarly, savings in financing costs due to a larger firm size should not affect marginal costs of production in

[7] See Rubinfeld (2000) for a detailed description of the arguments brought forward in the merger case. There, synergies are not mentioned as an argument in favor of the merger. Rather, the case heavily relied on the substitution patterns between the different cereals, which I estimate in detail.

the short run. One can furthermore test for synergies in brand value of the merging firms products. In my case, this does not lead to significant changes in the results.

A clear non-synergy rationale for the merger can be a reduction in debt for the Nabisco's former parent company, RJR Nabisco. After the 1988 leveraged buyout of RJR Nabisco which at this point was the largest leveraged buyout of all time, the ownership group accumulated a substantial pile of debt. Divesting different branches of the company such as the RTE cereal branch was thus a strategy to reduce the overall debt level.

Reasons for setting up a structural model There are several reasons for why prices can increase after a merger. These can be merger unrelated changes in consumer demand for a product that increase the consumers' willingness to pay, increases in input prices, or merging divisions jointly internalizing their pricing externalities. If the competitive constraints posed by rival firms' products are high for a specific product of one of the merging firms, this limits the potential to increase the price of this product post-merger. Similarly, reduced form regressions do not give information about the price-cost margins in the industry and their determinants. If these margins are high, this can be due to a high degree of product differentiation between different products that gives firms some local market power, or because of coordinated behavior between different firms. To gain insights about the different channels, I set up a structural model that allows me to disentangle the different effects.

3 Empirical model

In this section I introduce the main empirical model. I first focus on the demand side. Second, I introduce the cost properties of the model. Finally, I focus in detail on the supply-side properties, i.e. on both the degree of joint-profit maximization of merging firms and on industry conduct.

3.1 Demand side

On the demand side, I estimate a random coefficients logit model, with a specification closely related to Nevo (2001). The biggest strength of this model is that it allows for very flexible substitution patterns. An accurate estimation of cross-price elasticities is crucial in my model, because I am interested in the internalization of different firms' pricing externalities.

There is a total number of J brands in the market. Denote the number of individual consumers in every market by I, and denote the number of markets by T, where a market is defined as a time-store combination. Using a random coefficients logit model, individual i's indirect utility of consuming product j at market t is given by

$$u_{ijt} = x_j \beta_i^* + \alpha_i^* p_{jt} + \xi_{jt} + \epsilon_{ijt},\ i = 1, .., I; j = 1, .., J; t = 1, .., T. \tag{1}$$

x_j denotes a K-dimensional vector of firm j's observable brand characteristics, p_{jt} denotes the price of product j at market t, and ξ_{jt} the brand-specific mean valuation at market t that is unobservable to the researcher but observable to the firms. I decompose the brand-specific unobservable component into a market-invariant fixed component and a market-specific component: $\xi_{jt} = \bar{\xi}_j + \Delta\xi_{jt}$. In this case, $\bar{\xi}_j$ denotes a mean unobservable brand-valuation across all stores, while $\Delta\xi_{jt}$ is a store and market-specific component that is treated as an error term. Finally, ϵ_{ijt} is an idiosyncratic error term that is Type I extreme value distributed. The coefficients β^* and α^* are individual specific coefficients. These coefficients depend on the mean valuations, on a vector of demographic variables, D_i and their associated parameter coefficients Φ that measure how the tastes vary with demographics, as well as on an unobserved vector of shocks v_i that is interacted with a scaling matrix Σ:

$$\begin{pmatrix} \alpha_i^* \\ \beta_i^* \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Phi D_i + \Sigma v_i, \quad v_i \sim\ N(0, I_{K+1}). \tag{2}$$

I assume that the taste parameters follow a multivariate normal distribution conditional on the demographic variables, which follows the literature, see for example Nevo (2001). Because not all of the potential consumers purchase a good in each period, I also require an outside good. The indirect utility of not purchasing any product and thus consuming the outside good can be written as $u_{i0t} = \xi_0 + \phi_0 D_i + \sigma_0 v_{i0} + \epsilon_{i0t}$. As is common in the literature, I normalize ξ_0 to zero. I denote the vector of all demand side parameters by θ. This vector can be decomposed into a vector of parameters obtained from the linear part of the estimation, $\theta_1 = (\alpha, \beta)$, and a vector of parameters obtained from the nonlinear part of the estimation, $\theta_2 = (vec(\Phi), vec(\Sigma))$, respectively. The indirect utility of consuming a product can be decomposed into a mean utility part δ_{jt} and a mean-zero random component $\mu_{ijt} + \epsilon_{ijt}$ that takes into account heterogeneity from demographics and captures other shocks. The decomposed indirect utility can be expressed as $u_{ijt} = \delta_{jt}(x_j, p_{jt}, \xi_{jt}, \theta_1) + \mu_{ijt}(x_j, p_{jt}, v_i, D_i; \theta_2)$; with

$$\delta_{jt} = x_j \beta - \alpha p_{jt} + \xi_{jt},\ \mu_{ijt} = [p_{jt}, x_j]' * (\Pi D_i + \Sigma v_i), \tag{3}$$

where $[p_{jt}, x_j]$ is a $(K + 1) \times 1$ vector. Consumers either buy one unit of a single product or take the outside good. They choose the option which yields the highest indirect utility. This characterizes the set A_{jt} of unobservables that yield the highest utility for a specific choice j: $A_{jt}(x_{.t}, p_{.t}, \delta_{.t}, \theta_2) = \{(D_i, v_i, \epsilon_{it}) | u_{ijt} \geq u_{ilt} \forall l \in \{0, .., J\}\}$, where dotted indices indicate vectors over all J brands. The market shares predicted by the model can be obtained via integrating over the different shocks, using population moment functions $P^*(\cdot)$:

$$s_{jt}(x_{.t}, p_{.t}, \delta_{.t}, \theta_2) = \int_{A_{jt}} dP^*_\epsilon(\epsilon) dP^*_v(v) dP^*_D(D). \tag{4}$$

3.2 Industry technology

The J brands in the industry are produced by $N \leq J$ firms. Each brand can only be produced by one firm. The manufacturer marginal costs of all brands in the industry are unobserved to the researcher but common knowledge among firms. Therefore, they have to be backed out via first order conditions. As is common in the literature, I assume that marginal costs can be decomposed into cost factors that are observed to the researcher , such as input prices interacted with product characteristics, and a component that is unobserved to the researcher while known to the firms. I use a linear relationship between marginal costs and the observable cost component. This reflects a relatively weak substitutability of input production factors over the medium- and short-run in the RTE cereal industry. Denote the vector of brand j's observed cost drivers in market t by w_{jt}, and j's unobserved cost component at market t by ω_{jt}. The marginal costs of brand j in market t can be written as

$$mc_{jt} = w_{jt}\gamma + \omega_{jt}; \tag{5}$$

where γ denotes a vector of marginal cost parameters. I decompose the unobserved cost component ω for product j in market t in a market-invariant product fixed effect and a transitory market-specific component: $\omega_{jt} = \bar{\omega}_j + \Delta\omega_{jt}$. Further denote the cost component that relates to the observable cost characteristics and input prices by $\tilde{mc}_{jt} = w_{jt}\gamma$.

3.3 Industry behavior

On the supply side, I use a flexible price competition model which nests the assumption of multi-product Nash pricing. In each market, the manufacturing firms set wholesale prices for their products and the retailer sets a product-specific retail markup over the wholesale price. Denote by rx_{jt} the retail markup component of product j in market t. This component includes the marginal costs of the retailer, which I do not separately identify. I assume linear wholesale prices that are not contingent on the overall quantity sold in a period. This is similar to previous work regarding similar data, see for example Meza and Sudhir (2010) and Goldberg and Hellerstein (2013).[8] Henceforth I always denote a manufacturing firm simply as a firm for brevity. Each firm f owns a portfolio of brands $\mathbb{F}_f$. Denote by λ_{ij} the degree to which brand i takes into account brand j's profits when setting its optimal price. This implicitly defines a pre-merger ownership matrix Λ^t with the entries $\Lambda^t_{jr} = \lambda^t_{jr}$. Each of the parameters within Λ^t are normalized to lie in between 0 and 1, where 0 implies no internalization of profits, and 1 implies full internalization of profits. Not allowing for negative internalization parameters implies that a firm does not derive a positive utility from "ruining" another firm. Given Λ^t, the objective function for product j in market t can be written as:

$$\Pi_{jt} = (p_{jt} - mc_{jt} - rx_{jt})s_{jt}\overline{M} + \sum_{r \neq j} \lambda^t_{jr}(p_{rt} - rx_{rt} - mc_{rt})s_{rt}\overline{M}, \tag{6}$$

where s_{rt} denotes the market share of brand r, $\overline{M}$ the market size, and rx_{rt} the retailer's markup per unit sold by brand r in market t. The difference between the retail price and the retail markup, $p_{jt} - rx_{jt}$, amounts to the wholesale price.

The first order condition for product j's objective function with respect to its own price can be written as:

$$s_{jt} + \sum_{r=1}^{J} \lambda^t_{jr}(p_{rt} - rx_{jt} - mc_{rt})\frac{\partial s_{rt}}{\partial p_{jt}} = 0. \tag{7}$$

Define $\Omega^t_{jr} \equiv -\lambda^t_{jr} * \frac{\partial s_{rt}}{\partial p_{jt}}$. Given the demand parameters θ, the vector of marginal costs of production for all products in market t, $mc_{.t}$, conditional on the ownership matrix Λ^t, is

$$mc_{.t}(\theta, \Lambda^t, p_{.t}, x_{.t}, rx_{.t}) = p_{.t} - rx_{.t} - \Omega^{-1}(\theta, \Lambda^t)s_{.t}(\theta). \tag{8}$$

For notational simplicity, I henceforth always refer to the pre-merger conduct matrix as Λ. I assume that this matrix is invariant for all pre-merger markets. When estimating the degree of joint-profit maximization, I assume that this matrix is fully known, which is similar to most of the work in the literature. When estimating industry conduct, I relax this assumption for the matrix entries that reflect the interaction between brands of different

[8] See Villas-Boas (2007) for a framework to test for different forms of vertical relations using retail and input prices. The variation in both wholesale and retail prices make cases of zero wholesale or zero retailmargins very unlikely. I cannot distinguish between collusion and competition between retailers, because I only focus on data from a single retailer. However, since I take the retail markup as an exogenous outcome, this should not affect the results of the supply-estimations of the manufacturing firms. In an extension, I further analyze the case in which a manufacturing firm maximizes the profits of its products jointly with the retailer.

firms. Given the above assumptions, the pre-merger matrix Λ has the form

$$\Lambda = \begin{pmatrix} 1 & \lambda_{12} & .. & \lambda_{1J} \\ \lambda_{21} & 1 & .. & \lambda_{2J} \\ .. & .. & .. & .. \\ \lambda_{J1} & \lambda_{J2} & .. & 1 \end{pmatrix}.$$

Naturally, not all parameters within the matrix Λ can be estimated. In most empirical industrial organization models, Λ is fully assumed. Henceforth, I focus on two ways how to relax this assumption. First, I focus on estimating a (potentially) time-variant degree of joint-profit maximization of merging firms post-merger. Second, I focus on estimating the overall industry conduct, i.e. the parameters within Λ. Throughout the paper I make the assumption that the underlying conduct and the degree of joint-profit internalization of merging firms is identical across all stores for a given time period. This rules out cases in which manufacturing firms collude in some stores, and compete in others. I next explain the two different approaches in detail.

3.3.1 Internalization of merging divisions' profits

One key objective of the paper is to estimate the degree of profit internalization post-merger. Assume that before a merger, all firms internalize each others' pricing externalities with a symmetric factor $\lambda \in [0, 1]$, while fully maximizing joint profits of their own product portfolio.[9] The most natural baseline assumption is to set $\lambda = 0$, which reflects multi-product Nash pricing. I assume that non-merging firms will not change their competitive behavior after a merger, but adapt to the change in the merging firms' pricing. Denote by $\tilde{\lambda}^t$ the degree of joint-profit maximization between any product j of one merging firm and any product r of the other merging firm in market t, which I assume to be unobserved to the researcher, but common knowledge among firms in the industry. I allow for this degree of joint-profit maximization to be variable over time. This leads to the following definition for the internalization parameters in the ownership matrix:

$$\lambda_{jr}^t = \begin{cases} \lambda & \text{if } j \text{ and } r \text{ belong to different firms} \\ \tilde{\lambda}^t & \text{if } j \text{ belongs to one merging firm and } r \text{ belongs to the other} \\ & \quad \text{merging firm; and merger occured} \\ 1 & \text{if } j \text{ and } r \text{ belong to same firm already pre-merger.} \end{cases}$$

[9] I later relax the symmetry assumption to account for more heterogeneous forms of industry behavior.

Denote by $\tilde{\lambda}$ the vector of the internalization parameters for the merging firms across all post-merger markets. The model's implied vector of prices in market t given the parameters θ and γ can thus be written as

$$p_{.t}(\theta, \gamma, \Lambda, \tilde{\lambda}^t) = \tilde{mc}_{.t}(\theta, \gamma, \Lambda, \tilde{\lambda}^t) + rx_{.t} + \Omega^{-1}(\theta, \Lambda, \tilde{\lambda}^t) s_{.t}(\theta) + \omega_{.t}. \tag{9}$$

Here, $\omega_{.t}$ reflects the vector of the brand-specific unobserved cost components for each brand's cost function in market t.[10]

Example: Post-merger internalization of profits Assume there are three firms and four brands, and firm 1 owns the first two brands. There are two periods, one pre-merger and one after a merger between firms 3 and 4. Under initial multi-product Nash pricing, the pre- and post-merger ownership matrices can be written as

$$\Lambda = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \qquad \Lambda^{post} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tilde{\lambda} \\ 0 & 0 & \tilde{\lambda} & 1 \end{pmatrix}.$$

Here, $\tilde{\lambda}$ represents the degree to which a merging firm internalizes the effects it has on the profits of its merging partner into its pricing decisions after the merger.

3.3.2 Industry conduct

In many cases it is important to understand how intensely different firms in an industry compete. When estimating industry conduct, I assume that neither the form of pre-merger industry conduct, Λ, nor the post-merger industry conduct are known to the researcher. A key identifying assumption is that even though the researcher does not know the underlying form of industry conduct, he knows exactly how this conduct will be affected by a merger. When estimating industry conduct, I assume that merging firms fully internalize their pricing externalities after a merger. Non-merging firms compete with each other in the same fashion as before.[11] The change between merging and non-merging firms after the merger is

[10] Even if the degree of joint-profit maximization, $\tilde{\lambda}$, is part of a post-merger industry conduct matrix, I will explicitly state it in the model. This is to clearly distinguish the case of estimating the joint-profit-maximization parameters from the case when estimating the industry conduct matrix Λ.

[11] Naturally one can also use different identifying assumptions, which I discuss in more detail in online appendix B. Assuming no change in competitive conduct apart from the merging firms' joint-profit maximization seems like a natural first benchmark.

also known to the researcher. Online appendix B provides an extensive discussion of these assumptions. Overall, this leads to the following definition for the internalization parameters in the ownership matrix:

$$\lambda_{jr}^t = \begin{cases} \lambda_{jr} & \text{if } j \text{ and } r \text{ belong to different firms both pre- and post-merger} \\ 1 & \text{if } j \text{ and } r \text{ belong to different merging firms and merger occured} \\ 1 & \text{if } j \text{ and } r \text{ belong to same firm already pre-merger.} \end{cases}$$

Let the underlying form of pre-merger conduct, Λ, be an element of the $J \times J$-dimensional space B. The post-merger conduct can then be expressed via a transition function $b : B \to B$ which maps the pre-merger conduct matrix Λ into the post-merger conduct matrix $\Lambda^{post} = b(\Lambda)$. Consequently, for any pre-merger market t, $\Lambda^t = \Lambda$, and for any post-merger market t', $\Lambda^{t'} = b(\Lambda)$. Under these assumptions, for a specific form of pre-merger industry conduct Λ and transition function b, the model's implied vector of prices in market t given the parameters θ and γ can be written as:

$$p_{.t}(\theta, \gamma, \Lambda^t) = \tilde{mc}_{.t}(\theta, \gamma, \Lambda^t) + rx_{.t} + \Omega^{-1}(\theta, \Lambda^t) s_{.t}(\theta) + \omega_{.t}. \tag{10}$$

It is worth mentioning the key difference to conjectural variation models. I treat the conduct parameters in Λ as part of the firms' underlying objective functions rather than as behavioral responses with respect to the competitors' price setting behaviors. I then form moment conditions to recover the underlying "level" conduct parameters using a Generalized Method of Moments estimator.

Example: symmetric industry behavior Assume again that there are three firms and four brands, and firm 1 owns the first two brands. There are two periods, one pre-merger and one after a merger of firms 3 and 4. Pre-merger, all firms in the market internalize each others' pricing externalities with a factor λ. After the merger of firm 2 and firm 3, these firms fully internalize their pricing externalities. The pre- and post-merger ownership matrices can thus be written as

$$\Lambda = \begin{pmatrix} 1 & 1 & \lambda & \lambda \\ 1 & 1 & \lambda & \lambda \\ \lambda & \lambda & 1 & \lambda \\ \lambda & \lambda & \lambda & 1 \end{pmatrix}; \qquad \Lambda^{post} = \begin{pmatrix} 1 & 1 & \lambda & \lambda \\ 1 & 1 & \lambda & \lambda \\ \lambda & \lambda & 1 & 1 \\ \lambda & \lambda & 1 & 1 \end{pmatrix}.$$

While the cross-firm entries of Λ are not known, different values for λ lead to different predictions with respect to the change in post-merger pricing. This is because the underlying conduct also has effects on how much the internalization of merging firms' products changes pricing of all firms in the industry post-merger.

Assumptions on industry behavior To reduce the number of parameters to estimate, I impose some conditions on the interactions between different firms. In particular, I only consider cases in which a firm treats all brands of a specific competitor's firm in the same way. This excludes the possibility that single brands of different firms collude while others play against each other competitively. From a pure rank condition perspective the number of parameters I would have to estimate when accounting for brand-specific collusion between firms would easily exceed the number brands in the market. This makes it impossible to identify the parameters. Overall, there is a trade-off between the flexibility of industry conduct and the number of parameters that have to be estimated.

The most restrictive specification assumes that the cross-conduct parameters are identical for all brands in the market. The biggest advantage is that this returns a single cross-conduct parameter instead of a complicated matrix, and thus always meets the rank conditions. One disadvantage is that very often this parameter severely restricts the set of estimable economic models. For example, one will not be able to test for collusion between only a subset of firms in the market, or for differences in competitive behavior between firms.

Naturally there are many different ways one could restrict the interactions with respect to the internalization parameters inside the ownership matrix to be able to estimate these parameters. My main model specification assumes that all brands of two firms play against each other in the same way, which I refer to as "bilateral firm symmetry". As a consequence, all brands have the same cross-conduct parameters for all of their brand pairs. This still allows for partial collusion between two firms, but does not allow for more elaborate strategies, such as for example collusion only between a subset of brands of two firms. In terms of the parameter space, this reduces the number of cross-conduct parameters to $\frac{N(N-1)}{2}$. If the product space is sufficiently large compared to the number of firms in the industry, as is the case for the industry I consider, this allows me to estimate the different interactions inside the ownership matrix.[12]

[12] Another possibility is a case in which each firm behaves in the same way to all of its competitors. The advantage of this specification is that it reduces the number of parameters to only N different cross-conduct parameters. However, there are also several problems associated with this assumption. First, it is again no longer possible to detect partial cooperation between a subset of firms in the industry. Second, there is a consistency problem with respect to a mutual responsiveness: Under this assumption, it is possible that firm 1 is acting collusively with firm 2, and firm 2 on the other hand acts competitively towards firm 1, something which is harder to justify from an economic perspective.

4 Identification

Berry *et al.* (1995) and Berry (1994) show that the vector of unobserved brand components can be obtained via first inverting the market share equation: $\delta = s_{.t}^{-1}(\delta_{.t}, \theta_2)$. From equation 3 and via decomposing the vector of unobserved brand valuations into the market-invariant brand fixed effect vector $\bar{\xi}$ and a market-variant component, one can define the unobserved error term stacked across all markets as $\delta - x\beta - \bar{\xi} - \alpha p \equiv \varpi_{\xi}(\theta)$.[13] Depending on whether focusing on estimating the degree of merging firms profit maximization or the underlying industry conduct, the analogous vector of unobserved cost errors can be recovered from equation 9 or equation 10, respectively.

Because of the endogeneity of price with respect to the unobserved demand and cost components ξ and ω, appropriate instruments are needed. Note that I consider a constant product space over time. In general, having product entry requires making assumptions about how the new product is treated by all the other firms. Product exit would not pose such a problem, and under some conditions provides another source of variation to identify industry conduct. I discuss alternative sources of identification in online appendix B.

My data includes several markets from a single metropolitan area.[14] A constant product space over time across different stores implies that using first- and second-order polynomials of the rival firms' product characteristics as instruments, i.e. the classic "BLP-instruments" as in Berry *et al.* (1995), is not sufficient to identify the demand side parameters. Similarly, one cannot rule out that price movements across different stores and pricing zones of DFF are caused by the same demand shocks, which renders the use of zone-price instruments as for example in Nevo (2001) infeasible as well. For the above reasons, I exploit input price variation, the ownership change itself, and available data on wholesale prices to create demand side instruments. Furthermore, I create "differentiation IV's" which have been recently introduced by Gandhi and Houde (2015). These instruments exploit the relative proximity of the different products' observable non-linear product characteristics in the product space, i.e. for the characteristics for which I estimate random coefficients. I next discuss both demand and supply-side identification and particularly the computation of the differentiation instruments in detail.

[13]Note that this term also depends on the observed industry prices and product characteristics, respectively. These are left out for notational simplicity.

[14]The big advantage of this data, however, is that I observe also a measure for the wholesale prices.

4.1 Identification of demand parameters

On the demand side, I assume that at the true demand parameter values θ_0, the demand residual vector $\varpi_\xi(\theta_0)$ is uncorrelated with respect to a M_ξ-dimensional set of exogenous demand side instruments, Z_ξ. This leads to the identifying moment condition:

$$E[Z_\xi' \varpi_\xi(\theta_0)] = 0. \tag{11}$$

I use four different sets of instruments on the demand side. First, I exploit input price variation over time, which I interact with specific product characteristics. The economic assumption is that input price variation should be correlated with variation in retail prices, but not with consumers' preferences for unobservable product characteristics. Specifically, I interact the gasoline price at a given time with the distance of the production facilities and the brand-specific fixed effects. This measure should give a proxy of the transportation cost component that is variable over time, also see for example Goldberg and Hellerstein (2013).

Second, I exploit the ownership change itself as an instrument. The ownership change should cause a change in the firms' pricing decisions. For the instrument to be valid, the ownership change itself must not change the willingness to pay for all the brands in the industry. For all non-merging firms this assumption seems to be uncritical. For Post's products, also there is no significant change in branding or packaging after the merger. After being taken over by Post, Nabisco Shredded Wheat eventually is labeled as Post Shredded Wheat, while the rest of the packaging including size does not change.[15]

Third, I exploit information about wholesale prices from the data and interact them with product-fixed effects to generate another set of instruments. The underlying intuition of these instruments is that the changes in the transitory store- and time-specific brand valuations of consumers, $\Delta\xi$, should be primarily picked up by the retailer through a higher retail markup. This is because of more accurate information about the local change in consumers' preferences, see Chintagunta *et al.* (2003) for a use of wholesale prices in a similar context. To account for the danger of potential correlation between the wholesale price and temporary consumer valuation changes, I further conduct a hedonic wholesale price estimation in a first stage in which I regress the wholesale prices on demand and supply characteristics and use the predicted wholesale prices from this regression to construct the instruments.

Fourth, I create "differentiation instruments" as introduced by Gandhi and Houde (2015). Using the information of the observable product characteristics, these instruments make use

[15]If a change in the firm label for the caused a change in consumers' brand valuation for Nabisco products, this would naturally imply that the instrument is at least weekly correlated with the endogenous price. However, in my case this effect seems to be unlikely. In particular, on the cereal boxes the brand name ("Shredded Wheat") is much bigger than the respective firm name both before and after the merger.

of the relative isolation of the different products in the product space. Gandhi and Houde show that the optimal instruments are symmetric functions of the empirical distribution of product characteristic differences.

Following Gandhi and Houde's notation, define by $\mathbf{x_{jt}}$ the vector of product characteristics of product j in market t. Define by $d^x_{ij,t} \equiv x_i - x_j, i \neq j$ the difference in the non-linear product characteristic x between products i and j in market t for which consumers have heterogeneous tastes, i.e. a characteristic for which I account for a random coefficient in the estimation. Let $C^x = \{c^x_1, .., c^x_\upsilon\}$ denote υ equally spaced percentiles of the entire distribution of characteristics differences $d^x_{ij,t}$ with respect to product characteristic x, $i \neq j$ pooled across all markets. The differentiation instruments z^x_{jt} for a continuous non-linear characteristic x in market t can then be computed as

$$z^x_{jt} = \left\{ \sum_{i \neq j}^{J_t} \mathbb{1}(d^x_{ij,t} < c^x_k) \cdot \mathbf{x_{it}} \right\}_{k=1,..,\upsilon},$$

where $\cdot$ is the element-by-element product indicator. Note that in my case, the set of products J is invariant across all markets. The observable product characteristics set by the different manufacturer's is market-invariant as well, which implies $d^x_{ij,t}$ is constant for each non-linear product characteristic x. However, I observe variation in the promotional activity of the different products both across time for the same store and across stores for the same time. This enables me to generate differentiation instruments that interact the promotional activity of rival firms in a given market (i.e. a specific store-time combination) with measures of relative proximity in the characteristics space. The intuition for these instruments is that for a specific product j, the variation in rival firms' promotional activities across different markets generates variation in the consumers' substitution patterns. This variation should be correlated with the price of a firm because of the effects on demand, but not with the unobserved product-specific demand residual.[16] Overall, the different instruments contribute to identifying the endogenous price effect α, the different random coefficients, and their respective standard errors.

[16] Gandhi and Houde (2015) point out that for differentiation instruments being able to identify the underlying parameters, "it is essential that the model implies that local variation in the characteristics of nearby competitors is more relevant than the overall distribution of characteristic differences." Since the underlying product characteristics set by the manufacturers do not change, this is clearly given when exploiting information on promotional activity. Gandhi and Houde further point out the possibility to create differentiation instruments that exploit the differences in demographics when not having variation in the product space. However, since I do not observe changes in the demographics around the different stores over time, this is a less promising option.

4.2 Identification of cost and supply parameters

The supply model in each market t consists of a system of J equations for the different products whose prices are functions of the underlying form of conduct, Λ^t, and the degree of profit internalization, $\tilde{\lambda}^t$. The main identification task is to jointly identify firms' supply-side behavior and the marginal costs by exploiting meaningful moments.

Conditional on a specific form of industry conduct Λ^t in market t, I can back out marginal costs via a first order condition and then regress them on observable product characteristics combined with input prices. This allows me to predict the input cost component $\tilde{mc}$ of the marginal costs using input price data and the estimated parameters and in turn recover the unobserved cost vector ω. I make the implicit assumption that firms cannot substitute between different input goods. The recipes and production processes for a specific product in the RTE cereal industry remain constant over time, such that this assumption is likely to hold in the medium and short term. Together with the change in ownership and consumers' differentiated product preferences, the marginal costs influence the industry pricing. The brand-specific and time variant unobservable marginal cost residual may be correlated with unobservable product characteristics. Therefore it is essential to look for appropriate supply-side instruments. I use differentiation instruments as in the previous section.

It is worth highlighting one particular advantage of differentiation instruments compared to Chamberlain (1987)-style asymptotically optimal instruments as for example in Berry *et al.* (1999) and Reynaert and Verboven (2014). While the asymptotically optimal instruments exploit the non-linearity of the model structure, they require additional assumptions about the underlying industry competition when analyzing both the demand and supply side at least in a first stage. This can be problematic when the supply side is assumed to be unknown, as in my case. In contrast, one can compute the supply-side differentiation instruments directly from the data without having to make extra assumptions on the supply side.

Model identification when estimating joint-profit maximization of merging firms Rewriting equation (9) and solving for the unobserved cost residual $\varpi_{\tilde{\lambda}}(\theta, \gamma, \Lambda, \tilde{\lambda}) \equiv \Delta\omega$, yields:

$$\varpi_{\tilde{\lambda}}(\theta, \gamma, \Lambda, \tilde{\lambda}) = p - rx - \bar{\omega} - \tilde{mc}(\theta, \gamma, \Lambda, \tilde{\lambda}) - \Omega^{-1}(\theta, \Lambda, \tilde{\lambda})s.$$

As an identification restriction for the degree of joint-profit maximization, I use orthogonality conditions between the residual of observed and predicted post-merger prices, which results in the structural error $\varpi_{\tilde{\lambda},j}(\theta, \gamma, \Lambda, \tilde{\lambda})$ for a product j, and a $M_{\tilde{\lambda}}$-dimensional matrix of instruments $Z_{\tilde{\lambda}}$. This leads to the identifying moment condition

$$E[Z'_{\tilde{\lambda}}\varpi_{\tilde{\lambda}}(\theta,\gamma,\Lambda,\tilde{\lambda})] = 0. \tag{12}$$

Model identification when estimating industry conduct When estimating industry conduct, the main difference is that one estimates the industry ownership matrix Λ instead of the internalization parameters $\tilde{\lambda}$. Equating equation (10) with the observed prices p, and solving for the unobserved cost error $\varpi_{\Lambda}(\theta,\gamma,\Lambda;b()) \equiv \Delta\omega$ yields:

$$\varpi_{\Lambda}(\theta,\gamma,\Lambda;b(\cdot)) = p - rx - \bar{\omega} - \tilde{mc}(\theta,\gamma,\Lambda) - \Omega^{-1}(\theta,\Lambda,b(\Lambda))s.$$

Stacking the different instruments into the M_{Λ}-dimensional instrument matrix Z_{Λ} yields the identifying moment condition

$$E[Z'_{\Lambda}\varpi_{\Lambda}(\theta,\gamma,\Lambda;b(\cdot))] = 0. \tag{13}$$

One key assumption is that industry conduct is known among firms. Relaxing this assumption would cause two problems. First, this would make the assumption on symmetric behavior between two different firms harder to sustain. Second, I would have to specify beliefs of the different firms regarding other firms' behavior, which would severely increase the complexity of the analysis.

4.3 Relationship to Corts' (1999) Critique

Previous research has often used a conjectural variation approach to identify industry conduct, see for example Bresnahan (1989). In these models, a firm forms a "conjecture" about the responses of its competitors towards an increase in its own quantity. In this context, a conjecture can be seen as a reduced-form game theoretic best response function in symmetric quantity setting games. Corts (1999) critically discusses the identification of conjectural variation parameters. He shows that a conjectural variation parameter only estimates the marginal responsiveness of the marginal cost function with respect to changes in a demand shifter. As a researcher, one is however interested in the average slope of the marginal cost function instead of the marginal slope. My approach differs significantly from the conjectural variations approach and is not subject to this critique. This is because each firm sets prices for its portfolio of brands instead of quantities. Instead of forming conjectures about other brands' reactions, each firm's underlying objective function includes preferences for profits of other firms, thus allowing for cooperation among different firms. The preference parameters with respect to other firms' profits are essentially the conduct parameters I am interested in. I assume that these conduct parameters, as well as the marginal costs of all brands, are common knowledge in the industry, but not observed by the researcher. Using first order conditions of all brands' objective functions, my identification strategy allows to estimate both marginal cost parameters and the level conduct parameters. These amount to the "as-if conduct parameters" in Corts (1999).

Corts' also criticizes the static game character of conventional conduct estimation models. My approach is not fully exempt from this critique. I partially account for industry dynamics by modeling the merger-induced industry change. Nonetheless, my static approach may not detect certain fluctuating dynamic collusion patterns. One big advantage of my approach, however, is a higher degree of tractability. Modeling repeated games makes identification of conduct even more difficult due to a larger set of potential dynamic equilibrium strategies, which is a task for future research. With my approach, I am also able to identify patterns of full collusion as well as patterns of collusion between only a subset of firms.

5 Data and Estimation

5.1 Data description

I use scanner data from the DFF database. The database includes information from DFF's supermarket stores located in the Chicago Metropolitan area. In particular, it includes weekly information on product prices, quantities sold, and promotions, as well as 1990 census data yielding demographic variables for the different store locations. For my main analysis, I include data from 58 DFF stores and focus on 26 brands from the 6 different nationwide firms present in the industry in a time span from January 1992 until January 1995.[17] All of the products are consistently offered throughout the whole time span analyzed. There is a frequent introduction of new products by existing firms, which is consistent with large advertising campaigns in the beginning of a product's lifecycle. I find no persistent successful entry of new products in terms of a significant market share during the time-span I analyze. Therefore, I do not include these products in my estimations. The database further covers data on in-store promotions which DFF occasionally offers for the different products. There are four different forms of promotions, namely bonus buy sales, coupon sales, general sales, and price reduction sales.

I add input price data from the Thomson Reuters Datastream database and from the website www.indexmundi.com which I later use to estimate cost functions. The data contain prices on commodities needed for the production of cereal such as sugar and the different

[17] Overall, I exclude five weeks of observations because of missing data in the dataset.

grains, data on energy and electricity, and labor cost data. I further couple this data with nutrition facts from www.nutritiondata.self.com and information on the different production and processing techniques for the different cereals.

For each product, I include all package sizes between 10 and 32 OZ to form aggregated quantities and weighted prices. Over time, there is very little change in package sizes across the different products. Leaving out some larger package sizes would lead to losing information on some sales. This would however imply that the excluded products pose no competitive threat to the other products in terms of substitutabilty in the stores, which is a highly doubtful assumption. When incorporating multiple size choices per cereal brand name in my estimation, I would need to estimate many more substitution patterns in an unbalanced panel. This would hurt the identification assumptions of my model with respect to being able to forecast demand for each product also after a merger. Second, the products of the different firms come in different package sizes which would make a comparison without using weighted quantities even harder. One drawback of my strategy is that if the price per quantity is lower for larger than for smaller package sizes, forming a weighted average will not capture the full curvature of demand. It further implies that both competition and substitution between different products is primarily brand and not size specific, which I think is a reasonable assumption in this market. Overall, given my objective of identifying industry supply-side behavior, I think that the advantages of my specification clearly outweigh its disadvantages compared to the alternative specifications.

I do not include DFF private label cereal in my estimation such that its respective market share is attributed to the market share of the outside product, see Nevo (2001) for a similar assumption. Because the private label is only present for the locally operating retailer, it will have different underlying objectives than the nationwide operating manufacturers. This is even more so because of the vertically integrated relationship of the retailer and the private label for which I would have to account for in my model. Starting in the late 1980s, there is an increase in the market share for private label cereals in the US. This can of course affect choices of the branded cereals. To account for such a potential effect, I allow for a time trend in the indirect consumer utility of consuming one of the branded cereals.

I define a single quantity as a 1 OZ serving size for each product. Furthermore, the overall market size is defined as 2.5 times the mean store-specific number of total customers.[18]

Since I am primarily interested in the interactions among the manufacturing firms, ob-

[18] I find the results to be highly robust for different market size specifications. A market size higher than the number of customers in a store can capture both multi-person households with only a single shopping household member and consumers that usually shop at different retailers but might consider shopping at DFF. Because revenue from RTE cereal only amounts to a very small fraction of the total revenue generated in a store that is relatively stable throughout all seasons, the potential endogeneity between the market size and cereal prices should be negligible.

serving the wholesale prices for the different products rather than only the retail prices allows for a more precise characterization of the manufacturing firms' marginal costs and markups. I observe the retailer's average acquisition costs for each product at a given time in the data. This variable reflects the inventory-weighted average of the fraction of the retail price that was paid to the producer. From this variable I compute average wholesale prices for a given period. Note that this measure gives the weighted average of the wholesale prices for the products in the inventory, also see Chevalier *et al.* (2003) for a discussion of this variable.[19] Since I aggregate data on a monthly basis, this should lower concerns of stocking inventory at low wholesale prices for later dates.

5.2 Estimation algorithm

My structural estimation routine can be decomposed into five steps, which I outline below. For brevity, when estimating the degree of joint-profit maximization, I denote the specific internalization parameters that have to be estimated using my estimation routine by $vec(\tilde{\lambda})$. In this case, the pre-merger conduct matrix Λ is assumed to be known by the researcher.

When estimating industry conduct, I denote the conduct parameters to be estimated within the matrix Λ by $vec(\Lambda)$. Therefore, these parameters can change with every iteration of my estimation algorithm. In this case, the underlying post-merger integration parameters $vec(\tilde{\lambda})$ of the merging firms are assumed to be equal to 1 immediately after the merger for the merging firms, which reflects full and immediate maximization of joint profits.

1. **Estimate the demand parameters θ and compute $\frac{\partial s()}{\partial p}$.**

2. **Pick supply parameters $(vec(\Lambda)$,$vec(\tilde{\lambda}))$ given the identification restrictions.** When estimating the degree of joint-profit maximization, $vec(\Lambda)$ is invariant and chosen, while $vec(\tilde{\lambda})$ changes with every iteration of the algorithm. When estimating industry conduct, only $vec(\Lambda)$ changes with every iteration of the algorithm.

3. **Infer marginal costs for given pick of $(vec(\Lambda)$, $vec(\tilde{\lambda}))$, and $\frac{\partial s()}{\partial p}$ from the demand side estimation, and estimate marginal cost function.** Combining the derivatives of market share with respect to prices $\frac{\partial s()}{\partial p}$ from step 1 with the pick of $(vec(\Lambda), vec(\tilde{\lambda}))$ from step 2, I infer marginal costs and estimate a marginal costs equation to obtain observable marginal cost parameters γ, and the unobservable marginal cost component, ω.

[19]DFF uses the following formula for the average acquisition costs (AAC): AAC(t+1) = (Inventory bought in t) Price paid(t) + (Inventory, end of t-l-sales(t)) AAC(t). From an economic perspective, the variable's reflects the weighted profit share sold for each product in a period, absent the retailer's costs. Thus, it is a weighted average in terms of the time of purchase of the products in inventory, and does not reflect a product's current replacement value.

4. **Compute supply-side GMM criterion function.** Generate moments using the recovered cost function residuals $\varpi_{\tilde{\lambda}}$ for estimation of degree of joint-profit maximization, and ϖ_{Λ} for the conduct estimation, respectively, together with the supply-side instruments to compute the GMM criterion function for the specific combination of $\theta, \gamma, (vec(\Lambda), vec(\tilde{\lambda}))$.

5. **Repeat steps 2-4 until GMM criterion is minimized.**

In principle, it is also possible to estimate demand and supply jointly. This can potentially lead to efficiency gains because of using information on off-diagonals of the variance-covariance matrix. This requires to draw both the non-linear demand parameters θ_2 and the supply-side parameters $(vec(\Lambda), vec(\tilde{\lambda}))$ and estimate the implied demand side derivatives of market share with respect to price to estimate the residuals ϖ_{ξ} and $\varpi_{\tilde{\lambda}}$ (or $\varpi_{\tilde{\Lambda}}$, respectively) for every iteration of the estimation routine. The main reason for not estimating the model jointly is for technical robustness. I find that derivative-based optimization techniques are more robust in finding the minimum of the demand-side GMM function, which is consistent with the findings of Knittel and Metaxoglou (2014) using data from the same industry and a different dataset. On the supply-side, however, I find that non-derivative based optimization techniques are more reliable in finding the minimum of the GMM function. As a consequence of the sequential optimization, I have to account for demand-side errors when estimating standard errors of the cost and supply parameters via using two-stage standard error estimation routine. I use a two-stage correction method as outlined in Wooldridge (2010), chapters 12 and 14. In online appendix B, I discuss technical properties of the general estimation techniques in more detail.

5.3 Estimation criterion functions

Demand estimation I use the technique of Nevo (2001) to recover the structural demand parameters θ and the brand fixed effects $\bar{\xi}$. This allows me to estimate all the demand side parameters independently of the supply side. I solve for the mean utility level across all brands at market t, $\delta_{.t}$, to match the model's predicted market shares $s_{jt}(x_{.t}, p_{.t}, \xi_{.t}, \theta)$ from equation (11) with the actual market shares s_{jt} observed in the data. Following equation (11) the objective is to find:

$$\hat{\theta} = \arg\min_{\theta} \varpi_{\xi}(\theta)' Z_{\xi} \bar{A}_{\xi}^{-1} Z_{\xi}' \varpi_{\xi}(\theta); \tag{14}$$

where $\bar{A}_{\xi}^{-1}$ is an estimate of the asymptotically efficient covariance matrix $E[Z_{\xi}' \varpi_{\xi} \varpi_{\xi}' Z_{\xi}]$ calculated based on demand parameters obtained from a first-stage estimation with the identity matrix as weighting matrix.

Post-merger profit internalization estimation The nested estimation algorithm requires to invert marginal costs conditional on the specific form of industry conduct, Λ, and the degree of joint-profit maximization $\tilde{\lambda}$. In the third step of my estimation algorithm, I use the inverted marginal costs, and estimate the marginal cost equation (8) via minimizing the following objective function:

$$\hat{\gamma} = \arg\min_{\gamma} \varpi_{\tilde{\lambda}}(\theta, \gamma, \Lambda, \tilde{\lambda})' Z_{\omega} A_{\omega}^{-1} Z_{\omega}' \varpi_{\tilde{\lambda}}(\theta, \gamma, \Lambda, \tilde{\lambda}), \tag{15}$$

where $A_{\omega}^{-1} = Z_{\omega}' Z_{\omega}$; therefore this amounts to a linear GMM estimator. The unobserved brand specific cost error $\varpi_{\tilde{\lambda},j}$ for product j may be correlated with the price. As discussed before, I use both differentiation instruments and the cost characteristics as instruments.[20]

Having obtained the demand side coefficients θ and updating the cost parameters γ for the given form of of industry conduct Λ, I estimate the profit-internalization parameters $vec(\tilde{\lambda})$ by minimizing the following GMM objective function:

$$vec(\hat{\tilde{\lambda}}) = \arg\min_{vec(\tilde{\lambda})} \varpi_{\tilde{\lambda}}(\theta, \gamma, \Lambda, \tilde{\lambda}) Z_{\tilde{\lambda}} \bar{W}_{\tilde{\lambda}}^{-1} Z_{\tilde{\lambda}}' \varpi_{\tilde{\lambda}}(\theta, \gamma, \Lambda, \tilde{\lambda}), \tag{16}$$

where $\bar{W}_{\tilde{\lambda}}$ is an asymptotically efficient estimate of the covariance matrix $E[Z_{\tilde{\lambda}}' \varpi_{\tilde{\lambda}} \varpi_{\tilde{\lambda}}' Z_{\tilde{\lambda}}]$ computed from a first stage estimation. The moments consist of the empirical residuals $\varpi_{\tilde{\lambda}}(\theta, \gamma, \Lambda, \tilde{\lambda})$ interacted with the instruments $Z_{\tilde{\lambda}}$ as described in Section 4.

Industry conduct estimation When estimating industry conduct, I again use a nested routine that requires recovering marginal costs conditional on the form of industry conduct Λ, and then using a cost function estimation to obtain the residuals which I combine with suitable instruments to generate the industry conduct moments. This is because my object of interest, i.e. the industry conduct matrix Λ, influences the implied marginal costs. From equation (8), the cost parameters can be obtained via minimizing the following GMM

[20] A further potential instrument for the supply side can be changes in consumer income. An increase in consumer income should further translate into a positive demand effect at a given price. If such an income shock does not go along higher cereal manufacturing labor cost, then the resulting increase in demand should be uncorrelated with the unobserved cost component vector ω, also see Miller and Weinberg (2015). However, using regional income data and local consumer price indexes as additional instruments does not lead to accurate results in my case. One probable reason for this are the only small differences in the variation in income changes across the different geographic areas in my dataset.

objective function:

$$\hat{\gamma} = \arg\min_{\gamma} \varpi_{\Lambda}(\theta, \gamma, \Lambda; b(\cdot))' Z_{\omega} \tilde{A}_{\omega}^{-1} Z_{\omega}' \varpi_{\Lambda}(\theta, \gamma, \Lambda; b(\cdot)). \tag{17}$$

Having obtained the demand side coefficients θ and the cost parameters γ for any form of industry conduct, I estimate the flexible conduct parameters $vec(\Lambda)$ by minimizing the industry conduct GMM objective function. The moments consist of the empirical residuals $\varpi_{\Lambda}(\theta, \gamma, \Lambda; b(\cdot))$ interacted with the specific instruments Z_{Λ}, as described in section 4. Then this GMM objective function can be written as:

$$vec(\hat{\Lambda}) = \arg\min_{vec(\Lambda)} \varpi_{\Lambda}(\hat{\theta}, \hat{\gamma}, \Lambda; b(\cdot)) Z_{\Lambda}' \bar{W}_{\Lambda}^{-1} Z_{\Lambda}' \varpi_{\Lambda}(\hat{\theta}, \hat{\gamma}, \Lambda; b(\cdot)), \tag{18}$$

where $\bar{W}_{\Lambda}$ is an asymptotically efficient estimate of the covariance matrix $E[Z_{\Lambda}' \varpi_{\Lambda} \varpi_{\Lambda}' Z_{\Lambda}]$ computed with parameters obtained from a first stage estimation.

6 Results

6.1 Demand Estimates

Table 1 shows the estimation results for the full random coefficients logit demand model. In the main specification I include random coefficients for price, a constant, sogginess of cereal, sugar content, and fiber content. Furthermore, I use information regarding demographics around the stores to interact them with the different explanatory variables. These demographics are the level of education (percentage with a college degree), fraction of households with small children (i.e. less than 10 years old), and log median income around the store.[21]

As differentiation instruments for each product I generate histograms using the first and second tercile of the characteristics differences for all other products with respect to the non-linear product characteristics sugar and fiber content, i.e. set $\upsilon = 2$. I further use the non-linear discrete product characteristic sogginess to create a similar measure of characteristic differences. I interact these measures with the number of the corresponding other products' sales periods of combined price reduction, coupon, and promotion sales, and separately interact them for bonusbuy sales. I generate a second set of differentiation instruments only

[21]Note that I use the information on sale periods only for creating the differentiation instruments, but not as a variable in the demand equation. The main reason for this is that when incorporating a "dummy" variable (or multiple ones) for sale periods, I find that this variable takes away a lot of price variation, since prices are particularly low during sale periods. Including such a variable, both when accounting and not accounting for an extra sale price effect, leads to less realistic distributions of the underlying marginal costs, in particular due to highly negative marginal costs outliers. Because the underlying distribution of marginal costs is crucial in my case for the subsequent supply-side estimations, I find focusing only on the price effect of the sale periods to be the best option.

for the subsets of rival firms' products. Overall this leads to 24 differentiation instruments.

Table 1: Demand Estimates Full Random Coefficients Logit Model

Variable	Mean Coef.	Random Coef. σ	Interaction Education	Interaction Small Child	Interaction Income
Constant	−2.07	0.77		−0.69	0.46
	(0.06)	(0.09)		(0.39)	(0.90)
Price	−19.96	2.31	171.84	4.59	−8.28
	(0.83)	(0.90)	(16.50)	(1.39)	(3.35)
Sogginess	8.36	0.79		0.24	1.14
	(0.19)	(0.95)		(0.19)	(0.53)
Sugar	1.65	3.16		−1.22	2.56
	(0.14)	(0.46)		(0.46)	(1.01)
Fiber	−3.48	0.08	−214.38		23.14
	(0.43)	(122.10)	(45.90)		(5.61)
Time trend	-.01				
	(.00)				

Note: Number of Observation: 55796. Standard errors in parentheses. The store-specific demographic interactions are based on US 1990 Census data. The estimation include product-specific fixed-effects. Instruments include gasoline prices interacted with distance from manufacturers and product fixed effects, predicted wholesale prices interacted brand fixed effects, a post-merger dummy, and differentiation instruments based on the observable product characteristcs interacted with rival firms' and other products number of sale periods in the same store.

The estimation results indicate a positive relationship between income and higher price sensitivity, but a lower price sensitivity associated with a higher education. Price sensitivity is also positively related with more small children in a neighborhood. Demand for sugar is positively related with income, but negatively related with small children, which can be interpreted as families of small children being more sensitive to health considerations. One surprising outcome is that higher education is negatively related to the demand for fiber and positively related to the demand for sugar. However, this finding is similar to the one of Meza and Sudhir (2010) based on the same dataset for an earlier time period.

To test for empirical validity of the instruments, I apply several robustness checks. Focusing on the endogeneity concerns regarding the price sensitivity, I estimate different variants of a multinomial logit model. Table 15 in the online appendix shows the demand side estimation results for several specifications of the multinomial logit model, using the different subsets of the instruments used in the full random coefficients logit model. One can clearly see that all the different instruments contribute to estimating a higher price responsiveness than the baseline specification without instruments.

I check whether the differentiation instruments are indeed able to identify consumer heterogeneity with respect to the observable product characteristics. To do so, I apply a Wald specification test as proposed by Gandhi and Houde (2015). This test clearly rejects the null hypothesis of Independence of Irrelevant Alternatives for all differentiation instruments and for any subset consisting of the differentiation instruments for just a single observable product characteristic, see Table 7 in the online appendix for the results. The differentiation instruments seem to be important to accurately identify the random coefficients. This can be seen when comparing the main demand estimation with a specification of a full random coefficients logit model that excludes the differentiation instruments, as shown in Table 14 in the online appendix. In this case, the remaining instruments are not able to accurately identify the consumer heterogeneity with respect to the consumer demographics.

Demand Elasticities Integrating the market shares over the whole distribution of individuals yields the aggregated market shares from the model. Similarly, the cross-price elasticity between goods j and k at market t, η_{jkt}, can be computed as

$$\eta_{jkt} = \begin{cases} \frac{-p_{jt}}{s_{jt}} \int \alpha_i s_{ijt}(1 - s_{ijt}) dP_D(D) dP_v(v) & j = k \\ \frac{p_{kt}}{s_{jt}} \int \alpha_i s_{ijt} s_{ikt} dP_D(D) dP_v(v) & j \neq k \end{cases}$$

Table 8 and Table 9 in the online appendix show the median elasticities over all markets for my baseline random coefficients logit specification. The own-price elasticities are highly negative for all firms. There is furthermore significant variation in the different brands' substitution patterns. The median cross-price elasticities are all positive, which is consistent with products being imperfect substitutes. The results indicate that the cross-price elasticities are particularly high between products that belong either both to Kellogg's or both to General Mills, which are the two biggest firms in the industry, and between the most prominent products of these two firms. For the two merging firms, the results indicate high cross-price effects between Nabisco Shredded Wheat and both Post Grape Nuts and Post Honey Comb. Overall, the substitution patterns are relatively close to previous industry demand estimates using different instruments, see for example Nevo (2001), using the product prices of other metropolitan areas as instruments.

6.2 Cost Estimates

Table 10 in the online appendix shows the cost function estimates for different forms of underlying industry conduct, Λ. The specifications include multi-product Nash competition, different symmetric forms of partial cooperation, i.e. $\Lambda = .1$ and $\Lambda = .5$ for all cross-conduct parameters, and the estimated conduct of my full model specification as discussed in the next subsection. On the cost side, I account for both product and time fixed effects, as well as variable measures for transportation costs and input prices. These measures are monthly gasoline prices interacted with the distance to the factories of the different manufacturers, and input prices for corn, wheat, sugar, rice, and oat, interacted with the relative input contents for each product. The median marginal costs implied by the model lie between \$.111 per serving for multi-product Nash pricing and \$.077 for partial collusion with a weight of .5 for each brand of any rival firm. For the full conduct model in the next section, I find that the median marginal costs are 9 percent lower than the median marginal costs for multi-product Nash pricing. The estimations show a positive impact of gasoline prices on marginal costs. Regarding the different input prices, I find a positive impact of both the corn and oat price on marginal costs when also accounting for product fixed effects. The sign of the wheat price varies across the different specifications and is in most cases statistically not significant. For the rice price, which is relevant for only two products, and for the sugar price I find negative effects on marginal costs. Besides potential hedging of input prices, another reason for the negative impact can be demand driven from the model: if, ceteris paribus, a time period of relatively low implied wholesale or retail markups coincides with a high sugar price, then this could amount to a negative relationship between sugar price and implied costs. When excluding the product fixed effects, all of the input characteristics have a positive and significant impact on costs, as expected from the theory. Still, for my main specification I prefer to include product fixed effects. Doing so leads to more conservative estimates with respect to estimating the supply-side parameters of interest, particular when estimating cooperative behavior between firms within the industry, and also leads to much smaller standard errors. In particular, this leads to lower estimates of cooperative behavior between firms. My main results are also largely robust when excluding variable input prices and only accounting for time and product fixed effects, and variable gasoline prices.

6.3 Profit-internalization estimates

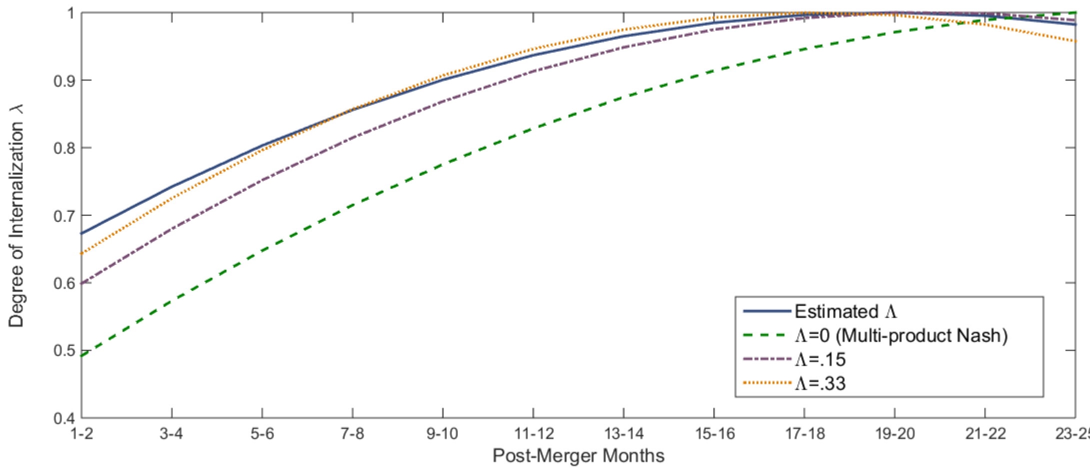
I estimate the degree of post-merger profit internalization of the merging firms by imposing a linear-quadratic structure for the internalization development. Thus, the degree of joint-profit maximization in market t at post-merger period τ can be written as

$$\tilde{\lambda}^t(\tau) = \tilde{\lambda}_0 + \tau\tilde{\lambda}_1 + \tau^2\tilde{\lambda}_2. \tag{19}$$

Figure 1 graphically shows the development of profit internalization parameters for the two merging firms over time for four different underlying industry conduct specifications.[22] In particular, among the specifications are multi-product Nash pricing, the estimates of the full

[22]Recall that the degree of joint-profit-maximization of merging firms is the same for all markets in a given post-merger period, and that a market is defined as a period-store pair.

Figure 1: Degree of joint-profit maximization $\tilde{\lambda}$ over time



Note: Lines reflect degree of merging firms' joint-profit maximization, $\tilde{\lambda}$, over time for different forms industry competition, Λ. Degree of joint-profit maximization is estimated for two-month intervals, using linear-quadratic relationship with respect to the post-merger time period τ: $\tilde{\lambda}(\tau) = \tilde{\lambda}_0 + \tau\tilde{\lambda}_1 + \tau^2\tilde{\lambda}_2$.

industry conduct model, and two different forms of partial symmetric internalization between firms in the industry. Table 11 in the online appendix shows the corresponding estimates and standard errors. One can see a clear pattern independent of the underlying form of industry conduct. There is a partial internalization of pricing externalities immediately after the merger, which increases over time to close to full internalization between 18 and 24 months after the merger. The results indicate an initially higher internalization of joint-profits under partial cooperation between the different firms in the industry than under multi-product Nash pricing. Recall that my estimation algorithm makes use of the difference between observed and the predicted model prices to form supply-side moments that also depend on the underlying industry conduct. For my estimates, the same degree of post-merger profit internalization under partial industry cooperation thus implies a smaller price increase for the merging firms' products than under multi-product Nash pricing.[23] Overall, however, the joint-profit internalization estimates only vary relatively little between the different industry

[23]There is no clear theoretical relationship between the magnitude of a post-merger price change and the underlying industry conduct because of opposing effects, see for example also Jaffe and Weyl (2013) for a discussion. Under partial industry cooperation, a price increase by the merging firms' products induces a larger price increase of the rival products as a response than under multi-product Nash pricing, leading to a lower loss in demand for the merging firms. However, when merging firms already partially internalize each others' profits pre-merger, this leads to a smaller maximum increase in the internalization level between the merging firms (for example from .15 pre-merger to 1 post-merger) than under Nash pricing (from 0 to 1). A third ambiguous effect in my case comes from marginal costs being both unobserved and contingent on the underlying form of industry conduct, Λ. A change in Λ can lead to different magnitudes of the associated changes in marginal costs for the different products in the industry. This in turn can affect the first order pricing conditions differently and can lead to both an increase and a decrease in the potential to increase prices under a more cooperative industry structure.

specifications, all indicating an increasing internalization pattern that leads to close to full joint-profit maximization.

The results are qualitatively robust with respect to several specifications changes. When only accounting for product and time fixed effects (and potentially gasoline prices), the internalization development is qualitatively similar to my main specification. The same holds when introducing 5 percent marginal cost synergies, with a slightly higher degree of internalization immediately post-merger. Note however, that due to the fact that the production factories remain unchanged in the first two years after the merger, such synergies in marginal costs do not seem likely.[24]

6.4 Conduct estimates

Table 2: Conduct estimates full model

	RA	KE	PO	NA	QU
General Mills	0.001	0.468	0.000	0.007	0.115
	(0.000)	(0.093)	(0.000)	(0.002)	(0.027)
Ralston		0.999	0.353	0.002	0.441
		(0.097)	(0.111)	(0.001)	(0.103)
Kellogg			0.284	0.984	0.000
			(0.101)	(0.172)	(0.000)
Post				0.009	0.983
				(0.004)	(0.417)
Nabisco					0.680
					(0.089)

	Full Model	Multi-pr. Nash	Full Model 5% Synerg.
Median PCM	0.430	0.322	0.436
Mean PCM	0.501	0.387	0.506
St. dev. PCM	0.283	0.256	0.279

Note: Table shows the estimated conduct parameters for the full model under the assumption of bilateral firm symmetry. It further gives information on the price-cost margins for the estimated specification and alternative specifications. Standard errors in parentheses are computed using two-step estimation correction to account for demand errors. Number of obs.: 55796.

Table 2 shows the conduct estimation results for my full model which uses the assumption of bilateral firm symmetry as introduced in section 3.3.2. Overall, I have to estimate 15 parameters within the ownership matrix. The parameter estimates show heterogeneity in the competitive behavior between the different firm pairs. The estimates suggest that the two leaders in the industry, Kellogg's and General Mills, partly internalize their pricing

[24] I also estimated a model variant in which I accounted for heterogeneous degrees of joint-profit internalization for the different merging firms. When not allowing for a dynamic time effect, full internalization of joint profits is the best predictor for both firms. When allowing for time-varying degrees of joint-profit maximization as in my baseline model, however, this does not lead to precise estimates. Because of the strong robustness of the results under symmetric internalization, I highly prefer this specification.

externalities rather than competing more aggressively with each other via multi-product Nash pricing. These two firms have an estimated internalization parameter of .47. With their large number of products, due to the strategic complementarity of prices, this also has a positive effect on the markups of the other firms' products. Except for partial internalizing Quaker's profits (.12), General Mills otherwise interacts fully competitively with the other companies in the industry. This is different for Kellogg's which has particularly high cross-conduct parameter values for the interactions with Ralston (.99) and Nabisco (.98). Except for its interactions with Quaker (.98), Post otherwise has relatively moderate conduct parameters, with a maximum of .35 for the interaction with Ralston. The estimated pre-merger conduct between Post and Nabisco is .01, indicating no internalization of pricing externalities before the merger, thus giving scope for changes in pricing strategies after the merger.[25]

Table 3 shows the conduct estimation results when estimating a single conduct parameter for different specifications regarding the cost structures and retail behavior. For the baseline specification the estimated parameter value is .28, indicating partial cooperation in the industry. The value is slightly lower when not accounting for variable input prices (.24), gasoline prices (.27) or when accounting for 5 percent synergies in marginal costs (.27). Overall, this suggests that the results are robust with respect to these changes in the cost specification.

The estimated conduct parameter is higher when ignoring the wholesale prices and thus only focusing on retail prices when estimating industry conduct (.47), and when not accounting for product fixed effects (.52). As argued before, because accounting for product fixed effects seems to be both more plausible and more conservative, i.e. leads to results that are closer to multi-product Nash pricing, I use these specification for my main estimations.

Table 3: Conduct estimates single parameter

Conduct Parameter (λ)	0.279	0.239	0.226	0.268	0.271	0.471	0.516
	(0.037)	(0.074)	(0.068)	(0.045)	(0.058)	(0.160)	(0.185)
Median PCM	0.419	0.404	0.399	0.417	0.356	0.493	0.500
Mean PCM	0.488	0.471	0.466	0.484	0.435	0.568	0.573
Std. dev. PCM	0.271	0.267	0.266	0.268	0.320	0.297	0.299
Var. Ingredient Cost	Yes	No	No	Yes	Yes	Yes	Yes
Gasoline Costs	Yes	Yes	No	No	Yes	Yes	Yes
Product Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	No
5 percent synergies	No	No	No	Yes	No	Yes	Yes
Separate Retail	Yes	Yes	Yes	Yes	Yes	No	Yes

Note: Table shows estimates and implied price-cost margins for different single conduct parameter specifications. Standard errors in parentheses are computed using two-step estimation correction to account for demand errors. Number of obs.: 55796.

[25] Relating to my profit-internalization estimates, I find that the conduct estimation results are qualitatively robust when first assuming only a partial internalization of pricing externalities post merger.

Implied price-cost margins One crucial question when estimating industry conduct is to what extent firm behavior contributes to industry markups. As a benchmark for competitive firm behavior, I use the assumption that firms compete via multi-product Nash pricing. Under this assumption, each firm maximizes the profits of its own brand portfolio, and all of the markups can be attributed to product differentiation, rather than to cooperative effects.

Table 2 shows median price-cost margins for the full model, for multi-product Nash pricing, and for the full model estimation under an assumption of 5 percent synergies after the merger. For my full model, I find a median 21.1 percent higher markup than under multi-product Nash pricing, while for the baseline single parameter model, the same figure amounts to 18.3 percent. This suggests that a significant part of the industry markups are attributable to cooperative firm behavior, while the major part is still due to product differentiation.

Table 12 in the online appendix further shows the median product-specific price-cost margins both for the full conduct model and under the assumption of multi-product Nash pricing. Under multi-product Nash-pricing, the results suggest that Kellogg's signature brands Corn Flakes, Frosted Flakes, Raising Bran, and Special K clearly have the highest margins in the industry, ranging from 46 to 88 percent of median markups. This is likely due to very high sales for these products, together with the fact that these products are more often on sale than other products. Furthermore, Kellogg's Corn Flakes on average has by far the lowest retail and wholesale prices in the industry. The markups of all Post, Nabisco, General Mills, and Quaker are all comparable with median markups mostly being between 25 and 35 percent under multi-product Nash pricing, with only the markups of Ralston being lower. The markups naturally increase using the full conduct model, mostly ranging between 35 and 50 percent. In terms of percentage median markup increases under the full model compared to multi-product Nash pricing, the biggest gainers are some of General Mills' products such as Wheaties and Cheerios, and Ralston products. For the merging firms' products, between 21 and 42 percent of the median markups are attributed to cooperative industry behavior, the highest percentage being attributed to the target firm Nabisco's product Shredded Wheat.

External validity of the results I next present additional information on the industry development for the time period shortly after the one I consider for my estimations. In particular, I argue that this development is consistent with both my findings of markups being significantly higher than under multi-product Nash pricing, and with a large part of the between-firm behavior that I estimate.

On March 7, 1995, the US Congressmen Samuel Gejdenson and Charles Schumer started a public campaign for lower cereal prices. Cotterill and Franklin (1999) provide a detailed description of the campaign and the subsequent industry changes. They state that this campaign had comprehensive coverage across the major nationwide newspapers and across several major TV news broadcasts. The public campaign was re-initiated in the news in March 1996 for the one year anniversary.

On April 15, 1996, Post and Nabisco decreased nationwide prices of their whole cereal line by 20 percent, at the same time announcing that cereal prices were too high. The other firms in the market responded in different fashions. Kellogg's as the market leader decreased the prices for two thirds of its product line by 19 percent about two months later on June 10th. General Mills responded 9 days later with a 11 percent price cut for 42 percent of its cereal line. Quaker responded another week later with a 15 percent price cut for 87 percent of its volume. Industry-wide, this amounted to an average price cut of 9.66 percent.[26] I interpret the price cut as a change in the underlying industry conduct. The magnitude of the price decreases are in line with firms having a significant markup over a competitive outcome before the price cut. One can naturally ask about whether multi-product Nash pricing is a suitable competitive benchmark. Cotterill and Franklin (1999) mention a 35 percent markup and thus a 26.2 percent implied wholesale margin for the RTE cereal industry in 1997, i.e. one year after the price cut. Except for three of Kellogg's signature brands (Corn Flakes, Raisin Bran, and Frosted Flakes), this number is arguably still relatively close to my median wholesale margin estimates under the assumption of multi-product Nash pricing, as can be seen in Table 12 in the online appendix. Overall, the price movements are thus consistent with a significant markup over multi-product Nash pricing before the price cut, as my full model suggests. Looking again at the estimates of the full conduct model in Table 2, it becomes clear that both Kellogg's and Quaker, the firms that respond the most to Post and Nabisco's initial price decrease, have high cross-conduct parameters with the merging firms. General Mills, which has much lower cross-conduct parameters with Nabisco, responds with a lower price decrease. Ralston as the firm with the lowest market share after the Post-Nabisco merger and low cross-conduct parameters with the merging firms was the only firm that kept a high pricing strategy, which resulted in a loss of market share from 4.1 to 2.6 percent, and eventually lead to a sale of its branded cereal line to General Mills in January 1997.[27]

[26] At the time at which the branded manufacturers decreased their prices, private label firms significantly increased the prices of their products, eventually fully closing the private-branded label pricing gap. Therefore, even if an increased popularity of the initially lower-priced private label products over time might be another explanation for a price cut by most branded manufacturers, private label pricing behavior did not become more aggressive over time.

[27] Because of the decrease in prices, which I interpret as a change in the underlying industry conduct, I cannot use this second merger for a further robustness check of my industry conduct estimation.

7 Conclusion

This paper estimates the degree of joint-profit maximization between merging firms and the form of industry conduct in the RTE cereal industry. The merger-induced ownership change serves as an important variation to identify within-firm and industry behavior. Furthermore, the construction of differentiation instruments crucially contributes to identifying both preference heterogeneity on the demand side and the supply-side parameters of interest. A particular advantage of these instruments for my estimations is that they can be computed independently of the underlying supply-side specification.

The availability of pre- and post-merger industry data allows me to estimate the degree of joint-profit maximization instead of assuming it. The empirical results shed light on the question of cooperation within a firm after a merger. The empirical findings from the structural model indicate a partial pricing adjustment by the merging firms immediately after the merger, and an increasing coordination over time. This suggests a close to full joint maximization of profits after a two-year transition period post-merger.

Estimating the coordination of joint pricing decisions is only a first step with respect to studying the implications of organizational integration using a structural model. Because the approach used in this paper allows for estimating intra-firm aspects while accounting for oligopolistic industry competition, one possible next step is to focus on the determinants of changes in corporate pricing decisions, which also relates to the literature in business strategy.

Incentives to cooperate across different divisions of a firm also likely depend on how profits of joint projects are distributed among the different merging divisions. Using data with more detailed information on organizational structures in this case could be used to make inferences about the determinants of divisional (and firm) productivity and on the optimal incentive structures. This could lead to more empirical research on the border between industrial organization and organizational economics, see for example Legros and Newman (2014) for a summary of the primarily theoretical literature.

The merged firm's pricing potential also crucially depends on the underlying form of industry conduct. Exploiting supply-side shifts using both pre- and post-merger industry data allows for estimating a heterogeneous supply structure across different firms. I do not have to rely on aggregate outside data or on non-nested selection methods to determine industry conduct. My results suggest that markups in the industry are above those predicted under multi-product Nash pricing. Specifically, I find that between 18.3 and 21.1 percent of the median manufacturer markups can be attributed to cooperative industry behavior. In this context, a next step is to focus on dynamic interactions between firms and thus to relate the empirical model to the large theoretical literature on collusion in repeated games.

The estimation strategy requires sufficient variation in the retail and input prices. In principle it can be applied to many different industries. Besides additional information about competition and merging firms' behavior within an industry, this would also give information about differences in behavior across industries. From an organizational perspective, this can yield insights about the effects of different managerial firm structures on post-merger behavior, and about differences in the potential to maximize joint profits. From an industry perspective, this can also provide more general information about the relationship between competition and market power across different industries.

References

ALONSO, R., DESSEIN, W. and MATOUSCHEK, N. (2008). When does coordination require centralization? *The American Economic Review*, **98** (1), 145–179.

APPEL, M., LABARRE, R. and RADULOVIC, D. (2004). On accelerated random search. *SIAM Journal on Optimization*, **14** (3), 708–731.

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

—, LEVINSOHN, J. and PAKES, A. (1995). Automobile prices in market equilibrium. *Econometrica*, **63** (4), 841–890.

—, — and — (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review*, **89** (3), 400–430.

BERRY, S. T. and HAILE, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, **82** (5), 1749–1797.

BLOOM, N., GENAKOS, C., SADUN, R. and VAN REENEN, J. (2012). Management practices across firms and countries. *The Academy of Management Perspectives*, **26** (1), 12–33.

BRESNAHAN, T. (1982). The oligopoly solution concept is identified. *Economics Letters*, **10** (1-2), 87–92.

— (1987). Competition and collusion in the american automobile industry: The 1955 price war. *The Journal of Industrial Economics*, **35** (4), 457–482.

— (1989). Empirical studies of industries with market power. *Handbook of industrial organization*, **2**, 1011–1057.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, **34** (3), 305–334.

Chandler, A. D. (1962). Strategy and structure: Chapters in the history of the american enterprise. *Massachusetts Institute of Technology Cambridge*.

Chevalier, J. A., Kashyap, A. K. and Rossi, P. E. (2003). Why don't prices rise during periods of peak demand? evidence from scanner data. *American Economic Review*, **93** (1), 15–37.

Chintagunta, P. K., Dubé, J.-P. and Singh, V. (2003). Balancing profitability and customer welfare in a supermarket chain. *Quantitative Marketing and Economics*, **1** (1), 111–147.

Ciliberto, F. and Williams, J. W. (2014). Does multimarket contact facilitate tacit collusion? inference on conduct parameters in the airline industry. *The RAND Journal of Economics*, **45** (4), 764–791.

Corts, K. (1995). *The ready-to-eat breakfast cereal industry in 1994 (A)*. Harvard Business School.

— (1999). Conduct parameters and the measurement of market power. *Journal of Econometrics*, **88** (2), 227–250.

Cotterill, R. and Franklin, A. (1999). An estimation of consumer benefits from the public campaign to lower cereal prices. *Agribusiness*, **15** (2), 273–287.

Crawford, G., Lee, R., Whinston, M. and Yurukoglu, A. (2016). The welfare effects of vertical integration in multichannel televistion. *mimeo*.

Dessein, W., Garicano, L. and Gertner, R. (2011). Organizing for synergies. *American Economic Journal: Microeconomics*, **2** (4), 77–114.

Fauli-Oller, R. and Giralt, M. (1995). Competition and cooperation within a multidivisional firm. *The Journal of Industrial Economics*, **43** (1), 77–99.

Fershtman, C. and Judd, K. L. (1987). Equilibrium incentives in oligopoly. *The American Economic Review*, pp. 927–940.

Gandhi, A. and Houde, J. (2015). Measuring substitution patterns in differentiated products industries. *mimeo*.

— and — (2016). Measuring firms' conduct in differentiated products industries. *mimeo*.

Gasmi, F., Laffont, J. J. and Vuong, Q. (1992). Econometric analysisof collusive behaviorin a soft-drink market. *Journal of Economics & Management Strategy*, **1** (2), 277–311.

Goldberg, P. K. and Hellerstein, R. (2013). A structural approach to identifying the sources of local currency price stability. *The Review of Economic Studies*, **80** (1), 175–210.

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

Jaffe, S. and Weyl, E. G. (2013). The first-order approach to merger analysis. *American Economic Journal: Microeconomics*, **5** (4), 188–218.

Knittel, C. R. and Metaxoglou, K. (2014). Estimation of random-coefficient demand models: Two empiricists' perspective. *Review of Economics and Statistics*, **96** (1), 34–59.

Lafontaine, F. and Slade, M. (2012). Inter-firm contracts: Evidence. In R. Gibbons and J. Roberts (eds.), *The Handbook of Organizational Economics*, Princeton University Press.

Lau, L. (1982). On identifying the degree of competitiveness from industry price and output data. *Economics Letters*, **10** (1-2), 93–99.

Legros, P. and Newman, A. F. (2014). Contracts, ownership, and industrial organization: Past and future. *Journal of Law, Economics, and Organization*, **30** (suppl 1), i82–i117.

Machlup, F. (1967). Theories of the firm: marginalist, behavioral, managerial. *The American economic review*, **57** (1), 1–33.

Meza, S. and Sudhir, K. (2010). Do private labels increase retailer bargaining power? *Quantitative Marketing and Economics*, **8** (3), 333–363.

Michel, C. (2013). Essays in industrial organization and consumer policy. *University of Mannheim Doctoral Dissertation*.

Miller, N. H. and Weinberg, M. (2015). Mergers facilitate tacit collusion: An empirical investigation of the miller/coors joint venture. *mimeo*.

Nevo, A. (1998). Identification of the oligopoly solution concept in a differentiated-products industry. *Economics Letters*, **59** (3), 391–395.

— (2000). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics*, **31** (3), 395–421.

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

Reynaert, M. and Verboven, F. (2014). Improving the performance of random coefficients demand models: the role of optimal instruments. *Journal of Econometrics*, **179** (1), 83–98.

Rivers, D. and Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, **39** (3), 347–366.

Rubinfeld, D. (2000). Market definition with differentiated products: the post / nabisco cereal merger. *Antitrust Law Journal*, **68**, 163–185.

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

Schwartz, M. and Thompson, E. A. (1986). Divisionalization and entry deterrence. *The Quarterly Journal of Economics*, pp. 307–321.

Shrivastava, P. (1986). Postmerger integration. *Journal of Business Strategy*, **7** (1), 65–76.

Thomas, C. (2011). Too many products: Decentralized decision making in multinational firms. *American Economic Journal: Microeconomics*, **3** (1), 280–306.

Villas-Boas, S. B. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. *The Review of Economic Studies*, **74** (2), 625–652.

Wickelgren, A. L. (2005). Managerial incentives and the price effects of mergers. *The Journal of Industrial Economics*, **53** (3), 327–353.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*.

Online Appendix - For Online Publication Only

"Identification and Estimation of Intra-Firm and Industry Competition via Ownership Change" by Christian Michel

A Reduced-form estimations

Retail-price development To further obtain insights on the development of industry prices before and after the merger, I estimate several reduced form regressions. Denote by p_{ist} the price of a product i in store s at time t. Denote by $\mathbb{1}_y^{ALL}$, $\mathbb{1}_y^{PO}$, and $\mathbb{1}_y^{NA}$ industry and merging firms' specific dummy variables, respectively, which are equal to 1 in year y post-merger and zero otherwise. Denote by κ and ϕ brand and store-specific fixed effects, respectively. $\mathbb{1}^{promo}$ is a promotion dummy variable, μ a constant, and ϵ an idiosyncratic error-term. I use the following reduced form pricing equation for product i at time t in store s:

$$log(p_{ist}) = \mu + \beta_1 \mathbb{1}_{y1}^{ALL} + \beta_2 \mathbb{1}_{y2}^{ALL} + \beta_3 \mathbb{1}_{y1}^{PO} + \beta_4 \mathbb{1}_{y2}^{PO} + \beta_5 \mathbb{1}_{y1}^{NA} + \beta_6 \mathbb{1}_{y2}^{NA} + \beta_7 \mathbb{1}_t^{promo} + \kappa_i + \phi_s + \epsilon_{ist}.$$

Table 4: Reduced-form retail price estimations

	(1) Baseline b/se	(2) Time-trend b/se	(3) Time-FE b/se
Post-mergerYear1	-0.008***	-0.025***	-0.012***
Post-mergerYear2	0.021***	-0.016***	0.056***
POSTMergerYear1	0.045***	0.045***	0.045***
POSTMergerYear2	0.091***	0.091***	0.091***
NabiscoMergerYear1	0.074***	0.074***	0.075***
NabiscoMergerYear2	0.043***	0.043***	0.043***
Promotion	-0.120***	-0.119***	-0.128***
Time-trend		0.002***	
Observations	55796	55796	55796
R-square	0.78	0.79	0.80

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All model specification include store and brand fixed effects. The dependent variable is the log price of a brand in a specific store at a specific month.

Table 4 sums up the results for different specifications of this equation. After the merger, on average there is an increase in retail prices. This increase is especially apparent in the second year after the merger. Merging firms' prices also particularly increase more than the market trend in the second year. To account for a potential systematic decrease in demand in the industry, I include a time trend in my main demand estimation.

Wholesale price development Table 6 shows the product-specific wholesale price development. The wholesale prices of Post's products on average increase over time, especially those of Raisin Bran and Honey Comb. In contrast, Nabisco's wholesale prices increase in the first year, but decrease in the second.

Table 5: Reduced-form wholesale price estimations

	(1) Baseline	(2) Time-trend	(3) Time-FE
Post-mergerYear1	0.032***	0.037***	0.015***
Post-mergerYear2	-0.036***	-0.025***	-0.043***
POSTMergerYear1	0.009**	0.009**	0.009**
POSTMergerYear2	0.066***	0.066***	0.067***
NABISCOMergerYear1	0.012*	0.012*	0.013**
NABISCOMergerYear2	-0.003	-0.003	-0.002
Promotion	-0.073***	-0.073***	-0.083***
Time-trend		-0.001***	
Observations	55772	55772	55772
R-square	0.82	0.82	0.83
Time FE	No	Yes	No

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All model specification include store and brand fixed effects. The dependent variable is the log wholesaleprice of a brand in a specific store at a specific month.

The wholesale prices for most of Kellogg's and Ralston's products increase, while there is a clear decline for General Mills and fluctuating wholesale prices for Quaker products. I exploit the availability of wholesale prices, wp, to conduct a similar reduced-form analysis as for retail prices. The according equation is

$$log(wp_{ist}) = \mu + \beta_1 \mathbb{1}_{y1}^{ALL} + \beta_2 \mathbb{1}_{y2}^{ALL} + \beta_3 \mathbb{1}_{y1}^{PO} + \beta_4 \mathbb{1}_{y2}^{PO} + \beta_5 \mathbb{1}_{y1}^{NA} + \beta_6 \mathbb{1}_{y2}^{NA} + \beta_7 \mathbb{1}_{t}^{promo} + \kappa_i + \phi_s + \epsilon_{ist}.$$

Table 5 sums up the results of the associated estimation. I find on average an increase in the industry-wide wholesale prices in the first year, and a decrease in the second year after the merger. The wholesale prices of Post's products increase in both years, while those of Nabisco do not significantly depart from the industry movements. For those firms for which the wholesale prices increase, the increase is lower than the increase in retail prices. This suggests that a large fraction of the gains from price increases are captured by the retailer.

On average the retailer's revenue shares, defined as the difference between retail and wholesale price divided by the retail price, are highest for Ralston products and lowest for Quaker products, which are both firms with only small market shares. The retail shares for the products of the two market leader's Kellogg's and General Mills are on average not lower than those of the other manufacturers' products. This suggests that firm size alone does not always result in a higher bargaining power for manufacturers in this industry.

Regarding promotion behavior, I do not observe big changes between pre- and post-merger periods. On average approximately every third month a specific product is on sale for at least one week in that month. For the merging firms, Post slightly increases its promotional activity from 36 to 38 percent of product-months with at least one sale, while the one of Nabisco decreases from 28 to 23 percent post-merger.

B Conduct specifications, alternative sources of identification, and computational details

B.1 Necessary rank conditions for different conduct specifications

This appendix provides identification conditions and examples for different specifications when estimating the continuous conduct parameters directly. This is opposed to the "menu approach", in which one selects among different non-nested models without estimating conduct parameters, see for example Nevo (1998).

Bilateral symmetry between firms As explained in section 4, one specification I consider to reduce the number of parameters to be estimated is that two firms compete against each other in the same way for all brands of their respective portfolios. There are J supply-side equations, i.e. one per brand. Thus, necessary for the rank conditions to be fulfilled is that $\frac{N(N-1)}{2} \leq J$ holds. This implies that there are sufficiently many brands in the brand portfolio compared to the overall number of firms. This is summarized in the first proposition.

Proposition 1 (Necessary conditions for identification under bilateral symmetry between firms)**.** *Suppose that for distinct firms f, g, $\lambda_{ij} = \lambda_{ik} = \lambda_{ji} = \lambda_{ki}$ $\forall i \in \mathbb{F}_f, \forall\ j,k \in \mathbb{F}_g$. Then industry conduct is identified only if the number of firms is sufficiently small compared to the number of products, i.e. if $\frac{N(N-1)}{2} \leq J$.*

Proof. Regarding the supply side, there are J estimable equations, one equation per brand post-merger. Because each firm has one conduct parameter for each competitor, this leads to an overall number of $N(N-1)$ parameters. The bilateral symmetry assumption reduces this number to $\frac{N(N-1)}{2}$. This leads to J equations with $\frac{N(N-1)}{2}$ parameters. The model is only identified if there are at least as many equations as parameters, i.e. if $\frac{N(N-1)}{2} \leq J$. $\square$

Same responsiveness between all firms As discussed before, the most restrictive assumption involves all firms treating all brands of rival firms the same way.

Proposition 2 (Necessary conditions for identification under same responsiveness between all firms)**.** *Suppose for distinct firms f, g, h , $\lambda_{ij} = \lambda_{ji} = \lambda_{ik} = \lambda_{jk} = \lambda_{kj}$ $\forall i \in \mathbb{F}_f, \forall\ j \in \mathbb{F}_g, \forall\ k \in \mathbb{F}_h$. Then the rank condition for industry conduct is always met.*

Proof. Using the same reasoning as in the proof for Proposition 2, there are J equations and one parameter to estimate, so that the result trivially holds. □

B.2 Rank conditions examples

In this subsection I present examples to highlight the effects of the different conduct specifications discussed above. The main question is under which circumstances the rank conditions for marginal costs and industry conduct are fulfilled depending on the total number of firms and brands in the market.

3 firms, brands 1 and 2 belong to same firm Consider an industry that consists of 4 brands, where brands 1 and 2 belong to the same firm. For simplicity, assume in this example that marginal costs are constant for each firm. Furthermore, denote by p_i, mc_i, s_i the price, marginal costs and market share of firm i, respectively. λ_{ij} describes the degree to which brand i takes into account the profits of brand j when making its decision. In the example, the maximization problem of brand 1 thus yields

$$\max_{p_1}(p_1 - mc_1)s_1(p) + (p_2 - mc_2)s_2(p) + \lambda_{13}(p_3 - mc_3)s_3(p) + \lambda_{14}(p_4 - mc_4)s_4(p)$$

The first-order condition for brand 1 with respect to its price then yields

$$(p_1 - mc_1)\frac{\partial s_1}{\partial p_1} + s_1 + (p_2 - mc_2)\frac{\partial s_2}{\partial p_1} + \lambda_{13}(p_3 - mc_3)\frac{\partial s_3}{\partial p_1} + \lambda_{14}(p_4 - mc_4)\frac{\partial s_4}{\partial p_1} = 0$$

There is a change in the ownership matrix pre- and post-merger if firms 2 and 3 merge. When making the additional assumption that each firm maximizes the profits of all of its brands, and merging firms fully internalize their profits, the associated pre- and post-merger conduct matrices can be written as

$$\Lambda = \begin{pmatrix} 1 & 1 & \lambda_{13} & \lambda_{14} \\ 1 & 1 & \lambda_{23} & \lambda_{24} \\ \lambda_{31} & \lambda_{32} & 1 & \lambda_{34} \\ \lambda_{41} & \lambda_{42} & \lambda_{43} & 1 \end{pmatrix}; \qquad \Lambda^{post} = b(\Lambda) = \begin{pmatrix} 1 & 1 & \lambda_{13} & \lambda_{14} \\ 1 & 1 & \lambda_{23} & \lambda_{24} \\ \lambda_{31} & \lambda_{32} & 1 & 1 \\ \lambda_{41} & \lambda_{42} & 1 & 1 \end{pmatrix}$$

From firm 1's first order condition, conditional on the form of industry conduct, firms will adapt their prices after an ownership change. In the above example, without symmetry, there are 10 parameters to estimate, with only 4 equations, such that the rank conditions are never met for identification. I introduce different assumption on firm supply to reduce the number of parameters to be estimated.

Bilateral symmetry between firms Instead of bilateral brand symmetry, a stricter assumption is that for all brands of two distinct firms, each brand will take the other firms' brands into account in the same fashion when making its pricing decision. Pre-merger and post-merger conduct can be written as

$$\Lambda = \begin{pmatrix} 1 & 1 & \lambda^a & \lambda^b \\ 1 & 1 & \lambda^a & \lambda^b \\ \lambda^a & \lambda^a & 1 & \lambda^c \\ \lambda^b & \lambda^b & \lambda^c & 1 \end{pmatrix}; \qquad \Lambda^{post} = b(\Lambda) = \begin{pmatrix} 1 & 1 & \lambda^a & \lambda^b \\ 1 & 1 & \lambda^a & \lambda^b \\ \lambda^a & \lambda^a & 1 & 1 \\ \lambda^b & \lambda^b & 1 & 1 \end{pmatrix}$$

This leads to a number of 3 parameters to estimate, with 4 available equations, such that the system is identified in absence of multi-collinearity.

Symmetry among all cross-firm brands When assuming that all brands take the brands of all other firms into account in the same way, this results in the following pre- and post-merger conduct:

$$\Lambda = \begin{pmatrix} 1 & 1 & \lambda^a & \lambda^a \\ 1 & 1 & \lambda^a & \lambda^a \\ \lambda^a & \lambda^a & 1 & \lambda^a \\ \lambda^a & \lambda^a & \lambda^a & 1 \end{pmatrix}; \qquad \Lambda^{post} = b(\Lambda) = \begin{pmatrix} 1 & 1 & \lambda^a & \lambda^a \\ 1 & 1 & \lambda^a & \lambda^a \\ \lambda^a & \lambda^a & 1 & 1 \\ \lambda^a & \lambda^a & 1 & 1 \end{pmatrix}$$

There is only one conduct parameter to estimate and in 4 equations.

B.3 Alternative identification sources and assumptions

Identifying industry conduct via product entry or exit Besides using a merger as an identification strategy for estimating industry conduct, one can also think of using other structural changes. Concerning product entry, there is the problem of comparing competition with and without the entrant. While one can still make the assumption that entry does not change how existing brands compete with each other, one has to define how a new product will interact with the existing products.

Unlike product entry, using product exit as an identification strategy is still feasible. However, one has to ask why a product will exit. One reason can be that it is just not profitable, which makes it likely that its impact on the market is relatively low. Therefore, a reduction of the brand space would not result in a big shift for firms strategies. Another possibility would be that a brand is profitable on its own, but it would be more profitable for a multi-brand firm to exit the product out of the market. This would result in an endogeneity problem when estimating conduct using product exit.

Different profit internalization assumptions to estimate industry conduct To infer the supply-side parameters of interest when estimating industry conduct, it is necessary to make specific assumptions about how the merging firms internalize their profits after the merger. I briefly introduce and discuss some of these assumptions.

Assumption 1 (Conduct between merging and non-merging firms)**.** *Let f, g be two distinct merging firms, and h a non-merging firm. Let λ_{ik}^{pre} and λ_{ik}^{post} denote the pre- and post-merger conduct parameters between firm i and k, respectively. Then, $\forall i \in \mathbb{F}_f$, $\forall j \in \mathbb{F}_g$, $\forall k \in \mathbb{F}_h$, one of the following cases holds regarding the conduct between a merging and a non-merging firm:*

a. $\lambda_{ik}^{post} = 1$; $\lambda_{jk}^{post} = 1$ *(full internalization);*

b. $\lambda_{ik}^{post} = \lambda_{ik}^{pre}$; $\lambda_{jk}^{post} = \lambda_{jk}^{pre}$ *(no change in conduct);*

c. $\lambda_{ik}^{post} = \lambda_{ik}^{pre}$; $\lambda_{jk}^{post} = \lambda_{ik}^{pre}$ *(acquiring firm standard).*

d. $\lambda_{ik}^{post} = \lambda_{jk}^{pre}$; $\lambda_{jk}^{post} = \lambda_{jk}^{pre}$ *(target firm standard);*

Assumption 1a represents full internalization of merging firms profits, which we assume in the paper when making inference on industry conduct. Under Assumption 1b, the merger does not change how competitors consider the two merging firms after the merger. In this case, the merging firms will fully internalize the profits after the merger. Under Assumption 1c, the fully merged entity is considered and behaves as the acquirer did pre-merger. Assumption 1d implies the reverse, i.e. that the merged entity behaves as the target. I do not have to pre-specify the values of the conduct parameters, but just the way in which the parameters change. Other change patterns can also be accounted for, as long as the change in conduct between merging and non-merging firms is known post-merger.

B.4 Choice of optimization techniques and tolerances

Demand side In my estimation routine, I simulate 200 individuals per store using Halton draws to estimate the different random coefficients. Overall, I estimate 5 random coefficients and 11 demographic interaction terms, leading to a total of 16 non-linear demand parameters. The importance of appropriate choices of optimizers and tolerances levels has been well documented for such highly nonlinear systems, see for example Knittel and Metaxoglou (2014) and the online appendix of Goldberg and Hellerstein (2013). On the demand side, I use the gradient based SOLVOPT optimizer while computing gradients analytically in each step. Similar to Knittel and Metaxoglou (2014) for a comparable setting with many different random coefficients and brands, I find this optimizer to be very powerful and robust in finding the minimum of the GMM objective function in comparison to other gradient- and non-gradient based optimizers. Inside the contraction mapping, I use a tolerance level of 10^{-9}. I use a termination tolerance level of 10^{-5} for the GMM objective function value. This is the tightest termination value for which I could achieve convergence. By using multiple starting values, I verify that the obtained minimum of the GMM function is a global minimum. The highest of the 16 gradients is of magnitude 10^{-5}, which I argue to be reasonably close to zero.

In terms of instrument choice, the differentiation instruments seem to be able to identify the underlying heterogeneity in consumer preferences. This can be seen when comparing the results of my full demand model using the full set of instruments in Table 1 with the demand estimation results excluding the differentiation instruments in Table 14 in appendix C. In the latter case, none of the non-linear coefficients is statistically significant, while 12 out of the 16 non-linear parameters are statistically significant in my preferred specification.

I next focus on whether the results are reasonable from an economic viewpoint. My estimation can disentangle the effects of education and median income. I find that while a higher fraction of consumers with a college degree decreases the price sensitivity, the opposite holds for a general income effect. Furthermore, I find that for a higher fraction of small children, there is on average a statistically significant decrease in the preference for sugar, which could relate to a higher health awareness of young families. While a lower preference for fiber and a higher preference for sugar related to a higher education level might initially seem counter-intuitive, these results are consistent with estimates of Meza and Sudhir (2010) using an earlier period of the same dataset.

Supply side On the supply side, I find non-gradient based optimizers to be more robust in finding the minimum of the GMM objective function. Specifically, because of its robustness, I use a finite-descent accelerated random search (ARS) algorithm, as proposed by Appel *et al.* (2004). For the estimation routine, I use a contraction factor $c = 2$ and a precision factor $\rho = 10^{-8}$, and allow for 16 different starting values to account for potential local equilibria. Corresponding to my theory model, I constrain all the internalization and conduct parameters to lie between 0 and 1. I find that gradient-based optimizers are less reliable and stable compared to non-gradient based optimizers, which is likely partly due to the non-analytical solutions of the supply-side gradients. Furthermore, I find that when estimating a high number of supply-side parameters, accelerated random search outperforms some of the other non-derivative approaches such as the simplex-based fminsearch optimizer.

Because of the sequential character of the estimation routine, I have to account for the demand estimation error when estimating standard errors on the supply side. I account for these effects by using a two-step error correction, which takes into account the demand error via adapting the weighting matrix with a term that is contingent on the structural demand error $\Delta\xi$, and on the gradients of the gmm objective demand function with respect to both the supply- and demand-side estimates. An extensive discussion of the required steps can be found in Wooldridge (2010), chapters 12 and 14.

I next focus on the reasonability of the estimation results. As argued before in the main text in section 6.4, the conduct estimation results imply a markup over multi-product Nash pricing. The on average 10 percent drop in the industry retail prices starting in April 1996 paired with still relatively high wholesale margins in 1997 are consistent with such the estimated behavior. Moreover, after Post-Nabisco start with a steep drop in industry prices, the firms with high cross-conduct parameters with Post-Nabisco respond with particular big drops in prices themselves.

C Additional tables and figures

Table 6: Product-specific price development

	Retail			Wholes.		
Brand Name	Year 0	Year 1	Year 2	Year 0	Year 1	Year 2
NAB Shredded Wheat	2.70	2.86	2.94	2.22	2.28	2.15
PO Raisin Bran	2.33	2.45	2.78	1.95	1.99	2.02
PO Grape Nuts	2.13	2.18	2.30	1.69	1.74	1.67
PO Honey Comb	3.39	3.56	3.87	2.77	2.89	2.97
GM Raisin Nut Bran	2.86	2.92	3.05	2.37	2.39	2.28
GM Apple Cinnamon Cheerios	3.21	3.11	3.03	2.54	2.56	2.21
GM Wheaties	2.77	2.71	2.65	2.16	2.21	1.99
GM Cheerios	3.22	3.06	3.11	2.59	2.63	2.40
GM Honey Nut Cheerios	3.09	2.95	2.94	2.51	2.53	2.23
GM Lucky Charms	3.53	3.50	3.47	2.81	2.83	2.58
GM Total Corn Flakes	3.89	3.93	3.80	3.21	3.22	2.82
GM Trix	4.00	3.77	3.91	3.29	3.18	2.92
KE Fruit Loops	3.48	3.31	3.57	2.75	2.83	2.66
KE Special K	3.49	3.64	3.68	2.84	2.91	2.70
KE Frosted Flakes	2.50	2.51	2.66	2.02	2.06	2.01
KE Corn Pops	3.44	3.46	3.42	2.75	2.80	2.63
KE Raisin Bran	2.36	2.25	2.57	1.89	1.91	1.93
KE Corn Flakes	1.75	1.81	2.07	1.38	1.49	1.61
KE Honey Smacks	2.94	3.07	3.18	2.34	2.48	2.36
KE Crispix	3.34	3.52	3.69	2.72	2.80	2.74
KE Rice Krispies	2.68	2.86	3.01	2.10	2.38	2.30
RAL Chex	3.37	3.46	3.97	2.62	2.71	2.79
RAL Wheat Chex	2.53	2.60	2.98	1.88	2.06	2.09
RAL Rice Chex	3.37	3.45	3.98	2.65	2.72	2.83
QU Quaker Oats 100%	2.12	2.15	1.99	1.72	1.68	1.50
QU Cap'n Crunch	2.76	2.70	2.88	2.11	2.23	2.06

Note: Columns $2-4$ show the product-specific mean deflated posted retailprices across all stores for the year before the merger, the first year after the merger, and the months $13-25$ after the merger for a hypothetical 15 OZ box of cereal. Columns $5-7$ show the analogous deflated mean wholesaleprices for the same time periods.

Table 7: IIA specification test for differentiation instruments

	All Differentiation Instruments	Sugar $*$ OtherSales	Fiber $*$ OtherSales	Sogginess $*$ OtherSales
χ^2	374.69	206.99	230.33	199.164
P-value	.00	.00	.00	.00
Degr. of freedom	24	8	8	8

Note: Wald test for joint significance of differentiation instruments in logit equation based on Gandhi and Houde (2015). Estimation uses as excluded instruments predicted wholesale prices interacted with product fixed effects, gasoline prices interacted with product fixed effects and distance to producers, and a post-merger dummy.

Table 8: Median elasticities random coefficients logit model part 1

	NAB ShW	PO RBr	PO GNu	PO HCb	GM RNB	GM ACC	GM Whe	GM Che	GM HNC	GM LCh	GM TCF	GM Tri	KE FrL
NAB Shred Wheat	-4.35	0.02	0.11	0.04	0.06	0.04	0.04	0.43	0.16	0.08	0.11	0.05	0.07
PO Raisin Bran	0.03	-3.17	0.03	0.02	0.02	0.02	0.15	0.07	0.06	0.04	0.02	0.03	0.04
PO Grape Nuts	0.14	0.02	-3.03	0.04	0.06	0.05	0.03	0.34	0.16	0.09	0.09	0.06	0.08
PO Honey Comb	0.08	0.02	0.07	-5.30	0.07	0.11	0.02	0.22	0.24	0.20	0.09	0.19	0.23
GM RaisinNutBran	0.11	0.02	0.09	0.06	-4.05	0.07	0.03	0.26	0.20	0.13	0.09	0.10	0.14
GM ApplCin Cheer	0.07	0.02	0.07	0.09	0.06	-4.39	0.02	0.20	0.23	0.20	0.08	0.17	0.22
GM Wheaties	0.06	0.15	0.03	0.02	0.02	0.02	-4.37	0.13	0.06	0.03	0.04	0.02	0.03
GM Cheerios	0.18	0.02	0.12	0.04	0.06	0.05	0.04	-4.68	0.18	0.10	0.12	0.06	0.09
GM HonNut Cheer	0.10	0.02	0.08	0.07	0.06	0.08	0.02	0.24	-4.03	0.15	0.09	0.12	0.16
GM Luck Charms	0.08	0.02	0.07	0.10	0.07	0.11	0.02	0.22	0.24	-5.00	0.09	0.17	0.23
GM Tot CoFlakes	0.13	0.02	0.09	0.05	0.05	0.05	0.03	0.31	0.17	0.11	-5.44	0.07	0.10
GM Trix	0.07	0.02	0.07	0.12	0.07	0.13	0.02	0.22	0.26	0.24	0.09	-5.97	0.28
KE Froot Loops	0.07	0.02	0.07	0.11	0.07	0.13	0.02	0.20	0.25	0.23	0.09	0.22	-5.06
KE Special K	0.06	0.15	0.03	0.02	0.02	0.02	0.25	0.13	0.06	0.04	0.05	0.03	0.04
KE Frost Flakes	0.03	0.16	0.03	0.02	0.02	0.03	0.18	0.08	0.08	0.05	0.03	0.04	0.06
KE Corn Pops	0.06	0.02	0.06	0.12	0.07	0.15	0.02	0.18	0.26	0.26	0.08	0.24	0.31
KE Raisin Bran	0.03	0.12	0.03	0.02	0.02	0.02	0.15	0.07	0.06	0.04	0.03	0.03	0.04
KE Corn Flakes	0.06	0.16	0.04	0.02	0.02	0.02	0.26	0.15	0.06	0.03	0.04	0.02	0.03
KE Honey Smacks	0.05	0.02	0.06	0.12	0.07	0.19	0.02	0.16	0.28	0.29	0.07	0.30	0.39
KE Crispix	0.13	0.02	0.09	0.05	0.05	0.05	0.03	0.30	0.17	0.10	0.10	0.07	0.10
KE Rice Krispies	0.14	0.02	0.10	0.04	0.05	0.05	0.03	0.34	0.17	0.10	0.10	0.07	0.09
RAL Chex	0.15	0.02	0.10	0.05	0.06	0.05	0.03	0.36	0.18	0.11	0.11	0.08	0.10
RAL Wheat Chex	0.15	0.02	0.11	0.04	0.06	0.05	0.03	0.37	0.17	0.10	0.10	0.06	0.09
RA Rice Chex	0.14	0.02	0.09	0.04	0.05	0.05	0.03	0.34	0.16	0.10	0.10	0.07	0.09
QU Quaker Oats	0.10	0.02	0.08	0.04	0.05	0.06	0.02	0.24	0.17	0.11	0.08	0.07	0.11
QU Capn Crunch	0.06	0.02	0.06	0.08	0.06	0.10	0.02	0.17	0.21	0.18	0.07	0.15	0.21

Note: Cell entries i (indexing row),j (indexing column), give the percent change in market share of brand i with a one percent change in price of j. Each entry represents the median of the elasticities from the 2146 markets. NAB=Nabisco, PO=Post, KE=Kellogg, GM=General Mills, QU=Quaker Oats, RA=Ralston.

Table 9: Median elasticities random coefficient logit model part 2

	KE SpK	KE FFl	KE CPo	KE RBr	KE CFl	KE HSm	KE Cri	KE RKr	RA Che	RA WCh	RA RCh	QU QO	QU CCr
NAB Shred Wheat	0.07	0.06	0.05	0.04	0.10	0.01	0.08	0.31	0.04	0.04	0.06	0.08	0.06
PO Raisin Bran	0.29	0.52	0.03	0.24	0.37	0.01	0.02	0.06	0.01	0.01	0.01	0.02	0.03
PO Grape Nuts	0.06	0.07	0.06	0.04	0.08	0.02	0.07	0.27	0.03	0.03	0.04	0.08	0.07
PO Honey Comb	0.05	0.12	0.21	0.05	0.05	0.06	0.06	0.22	0.03	0.02	0.04	0.08	0.15
GM RaisinNutBran	0.05	0.09	0.11	0.05	0.06	0.03	0.06	0.22	0.03	0.03	0.04	0.07	0.10
GM ApplCin Cheer	0.04	0.12	0.21	0.06	0.05	0.07	0.06	0.19	0.03	0.02	0.03	0.07	0.16
GM Wheaties	0.45	0.56	0.02	0.30	0.63	0.01	0.02	0.09	0.01	0.01	0.02	0.03	0.03
GM Cheerios	0.07	0.07	0.06	0.05	0.10	0.02	0.08	0.33	0.05	0.04	0.06	0.08	0.07
GM HonNut Cheer	0.05	0.10	0.13	0.05	0.05	0.04	0.06	0.22	0.03	0.02	0.04	0.07	0.12
GM Luck Charms	0.05	0.12	0.21	0.06	0.05	0.06	0.06	0.21	0.03	0.02	0.04	0.08	0.15
GM Tot CoFlakes	0.07	0.08	0.08	0.04	0.07	0.02	0.07	0.25	0.04	0.03	0.05	0.07	0.07
GM Trix	0.05	0.13	0.26	0.05	0.05	0.09	0.06	0.21	0.03	0.02	0.04	0.08	0.19
KE Froot Loops	0.05	0.13	0.26	0.06	0.05	0.09	0.06	0.19	0.03	0.02	0.04	0.08	0.19
KE Special K	-6.74	0.67	0.03	0.32	0.74	0.01	0.03	0.11	0.02	0.01	0.02	0.03	0.03
KE Frost Flakes	0.40	-3.89	0.05	0.37	0.56	0.01	0.02	0.07	0.01	0.01	0.01	0.03	0.04
KE Corn Pops	0.04	0.14	-5.44	0.05	0.04	0.11	0.06	0.19	0.03	0.02	0.03	0.07	0.21
KE Raisin Bran	0.30	0.59	0.03	-3.06	0.38	0.01	0.02	0.06	0.01	0.01	0.01	0.02	0.03
KE Corn Flakes	0.60	0.81	0.02	0.35	-3.25	0.01	0.03	0.12	0.01	0.01	0.02	0.03	0.03
KE Honey Smacks	0.04	0.14	0.39	0.06	0.04	-5.33	0.05	0.15	0.02	0.02	0.03	0.08	0.26
KE Crispix	0.06	0.07	0.07	0.04	0.07	0.02	-4.88	0.26	0.04	0.03	0.05	0.07	0.07
KE Rice Krispies	0.06	0.07	0.07	0.04	0.08	0.02	0.08	-3.73	0.04	0.03	0.05	0.07	0.07
RAL Chex	0.07	0.08	0.07	0.04	0.08	0.02	0.08	0.28	-5.32	0.03	0.05	0.07	0.07
RAL Wheat Chex	0.06	0.07	0.06	0.05	0.09	0.02	0.07	0.28	0.04	-4.04	0.05	0.08	0.07
RA Rice Chex	0.07	0.07	0.06	0.04	0.07	0.02	0.07	0.26	0.04	0.03	-4.96	0.06	0.07
QU Quaker Oats	0.04	0.08	0.08	0.04	0.06	0.02	0.05	0.20	0.03	0.02	0.03	-2.62	0.08
QU Capn Crunch	0.04	0.11	0.19	0.05	0.04	0.07	0.05	0.17	0.02	0.02	0.03	0.07	-3.44

Note: Cell entries i (indexing row),j (indexing column), give the percent change in market share of brand i with a one percent change in price of j. Each entry represents the median of the elasticities from the 2146 markets. NAB=Nabisco, PO=Post, KE=Kellogg, GM=General Mills, QU=Quaker Oats, RA=Ralston.

Table 10: Cost function estimates

Industry Competition	Nash	$(\Lambda = 0)$	$\Lambda = 0.1$		$\Lambda = 0.5$		Full Model	$(\Lambda = \hat{\Lambda})$
Constant	−0.014	(0.023)	−0.002	(0.023)	0.024	(0.024)	0.108	(0.023)
Promotion Period	−0.015	(0.000)	−0.015	(0.000)	−0.014	(0.000)	−0.015	(0.000)
Gasprice*Distance	0.141	(0.021)	0.124	(0.021)	0.119	(0.022)	0.058	(0.020)
rel. Wheatprice	0.016	(0.025)	0.006	(0.025)	−0.023	(0.026)	−0.057	(0.024)
rel. Cornprice	0.254	(0.029)	0.230	(0.029)	0.193	(0.030)	0.1681	(0.028)
rel. Sugarprice	−0.140	(0.040)	−0.140	(0.040)	−0.150	(0.042)	−0.20	(0.040)
rel. Oatprice	0.076	(0.004)	0.075	(0.004)	0.078	(0.004)	0.067	(0.004)
rel. Riceprice	−0.043	(0.012)	−0.037	(0.012)	−0.023	(0.012)	−0.036	(0.011)
Median MC	0.111		0.105		0.077		0.088	
Mean MC	0.103		0.098		0.069		0.081	
Std MC	0.054		0.054		0.057		0.056	

Note: Columns 2-5 show cost estimates for different forms of industry competition Λ, with respective standard errors in parentheses. In particular, the specifications are the full conduct model, multi-product Nash pricing, and different forms of symmetric partial cooperation. All estimations include product and time fixed effects. Measures for input prices interacted with ingredient weights computed as 10 times input price (in $ per g) times the amount of the product input (in g per 100g of the specific cereal). Number of obs.: 55796.

Table 11: Estimates degree of joint-profit maximization of merging-firms

	Full Model $(\Lambda = \hat{\Lambda})$	Nash $(\Lambda = 0)$	$\Lambda = 0.15$	$\Lambda = 0.33$	$\Lambda = 0.5$
Constant	0.596 (0.002)	0.403 (0.001)	0.509 (0.001)	0.578 (0.001)	0.434 (0.001)
Lin. trend	0.082 (0.001)	0.092 (0.001)	0.095 (0.007)	0.092 (0.000)	0.094 (0.002)
Quad. trend	−0.004 (0.012)	−0.004 (0.013)	−0.005 (0.079)	−0.005 (0.000)	−0.004 (0.022)
M1-M2	0.673 (0.116)	0.492 (0.075)	0.599 (0.077)	0.664 (0.164)	0.524 (0.168)
M3-M4	0.742 (0.157)	0.573 (0.091)	0.680 (0.096)	0.741 (0.227)	0.607 (0.241)
M5-M6	0.803 (0.331)	0.648 (0.078)	0.752 (0.126)	0.808 (0.089)	0.681 (0.160)
M7-M8	0.856 (0.086)	0.715 (0.185)	0.815 (0.175)	0.865 (0.270)	0.748 (0.233)
M9-M10	0.900 (0.116)	0.775 (0.229)	0.868 (0.139)	0.912 (0.262)	0.807 (0.124)
M11-M12	0.937 (0.204)	0.829 (0.120)	0.913 (0.149)	0.949 (0.205)	0.858 (0.191)
M13-M14	0.965 (0.160)	0.875 (0.164)	0.948 (0.163)	0.976 (0.172)	0.901 (0.122)
M15-M16	0.985 (0.089)	0.914 (0.141)	0.975 (0.120)	0.993 (0.175)	0.936 (0.128)
M17-M18	0.997 (0.119)	0.946 (0.150)	0.992 (0.064)	1.000 (0.208)	0.964 (0.216)
M19-M20	1.000 (0.081)	0.971 (0.109)	1.000 (0.135)	0.997 (0.188)	0.984 (0.221)
M21-M22	0.995 (0.124)	0.989 (0.190)	0.999 (0.192)	0.985 (0.144)	0.996 (0.199)
M23-M25	0.982 (0.123)	1.000 (0.317)	0.989 (0.096)	0.962 (0.272)	1.000 (0.074)

Note: Standard errors in parentheses are computed using two-step estimation correction to account for demand errors. Columns 2-5 show estimates for the degree of joint-profit maximization over time $\tilde{\lambda}$ imposing symmetry between the merging firms, for different forms of industry competition Λ. Degree of joint-profit maximization is estimated using two-month intervals with a linear-quadratic relationship with respect to the post-merger time period τ: $\tilde{\lambda}(\tau) = \tilde{\lambda}_0 + \tau\tilde{\lambda}_1 + \tau^2\tilde{\lambda}_2$. Number of obs.: 55796.

Table 12: Median wholesale margin differences full model compared to Nash pricing

	med. % PCM over Nash	med. PCM Full Model	med. PCM Multi-prod. Nash	St. Dev. PCM
NAB Shredded Wheat	0.42	0.49	0.33	0.09
PO Raisin Bran	0.35	0.48	0.35	0.11
PO Grape Nuts	0.21	0.49	0.41	0.07
PO Honey Comb	0.22	0.33	0.27	0.05
GM Raisin Nut Bran	0.27	0.43	0.33	0.08
GM Apple Cinnamon Cheerios	0.30	0.42	0.32	0.09
GM Wheaties	0.64	0.52	0.31	0.18
GM Cheerios	0.26	0.39	0.31	0.09
GM Honey Nut Cheerios	0.28	0.43	0.33	0.09
GM Lucky Charms	0.30	0.37	0.29	0.08
GM Total Corn Flakes	0.27	0.32	0.25	0.06
GM Trix	0.32	0.34	0.26	0.08
KE Fruit Loops	0.28	0.42	0.32	0.13
KE Special K	0.17	0.55	0.46	0.22
KE Frosted Flakes	0.16	0.79	0.67	0.34
KE Corn Pops	0.27	0.43	0.33	0.13
KE Raisin Bran	0.16	0.82	0.70	0.36
KE Corn Flakes	0.16	1.05	0.88	0.47
KE Honey Smacks	0.25	0.48	0.38	0.62
KE Crispix	0.29	0.40	0.31	0.07
KE Rice Krispies	0.29	0.49	0.36	0.13
RAL Chex	0.59	0.34	0.22	0.07
RAL Wheat Chex	0.57	0.44	0.29	0.09
RAL Rice Chex	0.59	0.34	0.22	0.07
QU Quaker Oats 100%	0.17	0.44	0.38	0.10
QU Cap'n Crunch	0.16	0.34	0.29	0.10

Table 13: Sample statistics

Description	Mean	Median	Std	Min	Max
Gasoline price ($\$/gal$)	1.145	1.141	0.085	1.016	1.518
Wheat price ($\$/kg$)	.142	.139	.024	.110	.227
Corn price ($\$/kg$)	.107	.103	.020	.084	.173
Rice price ($\$/kg$)	.261	.264	.038	.169	.331
Oat price ($\$/kg$)	.110	.102	.026	.080	.176
Sugar price ($\$/kg$)	.199	.199	.098	.180	.233
Fiber ($g/(100g\ cereal)$)	5.42	3.85	3.93	.80	13.40
Sugar ($g/(100g\ cereal)$)	25.10	26.40	15.96	1	56
Sogginess (=1 if soggy in cereal)	.23		.42		
Education (% with college degree)	.218	.216	.105		
Young Children (% households with child $<$ 10 years)	.135	.137	.025		
Income (log median income)	10.62	10.58	.291		

Table 14: Random coefficients logit estimates without differentiation instruments

Variable	Mean Coef.	Random Coef. σ	Interaction Education	Interaction Small Child	Interaction Income
Constant	5.13	-1.62		-1.46	-3.55
	(0.62)	(3.73)		(1.27)	(4.81)
Price	-22.06	-4.52	-22.28	1.60	1.80
	(6.13)	(4.15)	(43.63)	(5.02)	(9.86)
Sogginess	15.83	0.04		3.21	4.15
	(0.83)	(52.94)		(3.91)	(11.76)
Sugar	-8.72	4.76		-5.67	0.90
	(0.87)	(3.36)		(3.43)	(4.62)
Fiber	5.25	0.91	-59.48		13.83
	(0.79)	(482.97)	(392.99)		(28.58)

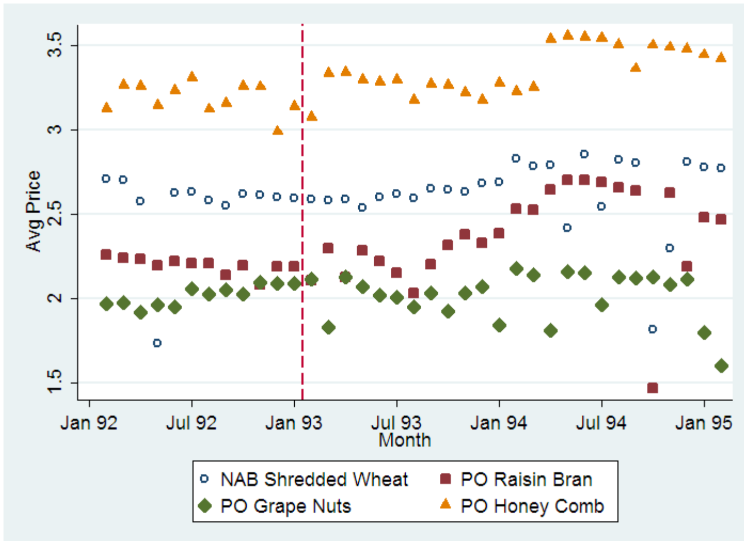
Note: Number of Observation: 55796. Standard errors in parentheses. The store-specific demographic interactions are based on US 1990 Census data. The estimation include product-specific fixed-effects. Instruments include gasoline prices interacted with distance from manufacturers and product fixed effects, predicted wholesale prices interacted brand fixed effects, a post-merger dummy.

Table 15: Demand estimates multinomial logit model

	OLS	IV All	IV Costs+ Wholesalep.	IV Wholesalep.+ DiffIV	IV Costs + DiffIV
Price (α)	-11.16	-19.92	-20.39	-20.01	-17.52
	(.10)	(.45)	(1.78)	(.64)	(.70)
Education	3.2	-2.07	-72.56	$-.04$	4.94
	(.09)	(4.00)	(48.07)	(.05)	(3.76)
Small Children	.08	.50	-1.44	.60	.47
	(.00)	(.13)	(1.71)	(.11)	(.08)
Income	.45	.65	7.35	.80	.19
	(.01)	(.39)	(5.41)	(.43)	(.34)
Time-trend	.00	.01	.01	.01	.01
	(.00)	(.00)	(.00)	(.00)	(.00)
Product fixed effects	Yes	Yes	Yes	Yes	Yes

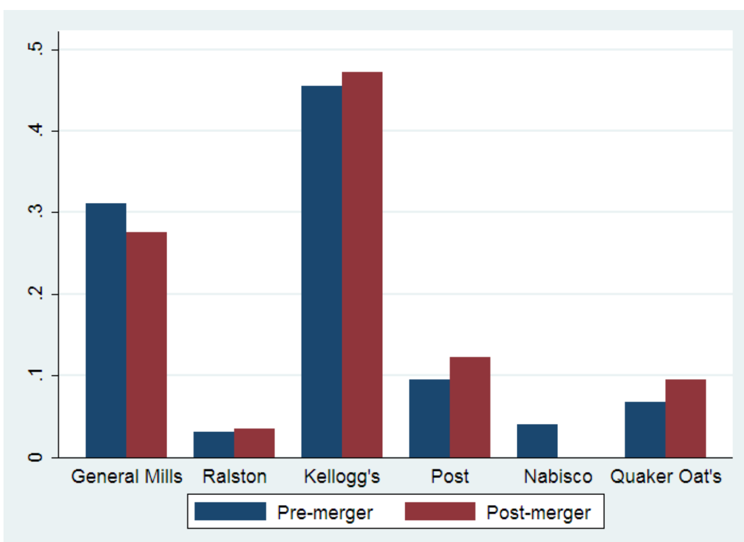
Note: Number of Observations: 55796.

Figure 2: Retail price development of merging brands



Note: Each data point indicates quantity-weighted average deflated retail price for a merging brand across all stores in a month for a hypothetical 15 OZ box of cereal.

Figure 3: Pre- and post-merger market shares



Note: Aggregate market shares for all products and entire time span across all stores used for demand estimation.